CHAPTER 9

Time-Varying Fields and Maxwell's Equations

The basic relationships of the electrostatic field and the steady magnetic field were obtained in the previous eight chapters, and we are now ready to discuss time-varying fields. The discussion will be short, for vector analysis and vector calculus should now be more familiar tools; some of the relationships are unchanged, and most of the relationships are changed only slightly.

Two new concepts will be introduced: the electric field produced by a changing magnetic field and the magnetic field produced by a changing electric field. The first of these concepts resulted from experimental research by Michael Faraday and the second from the theoretical efforts of James Clerk Maxwell.

Maxwell actually was inspired by Faraday's experimental work and by the mental picture provided through the "lines of force" that Faraday introduced in developing his theory of electricity and magnetism. He was 40 years younger than Faraday, but they knew each other during the five years Maxwell spent in London as a young professor, a few years after Faraday had retired. Maxwell's theory was developed subsequent to his holding this university position while he was working alone at his home in Scotland. It occupied him for five years between the ages of 35 and 40.

The four basic equations of electromagnetic theory presented in this chapter bear his name. ■

9.1 FARADAY'S LAW

After Oersted[1] demonstrated in 1820 that an electric current affected a compass needle, Faraday professed his belief that if a current could produce a magnetic field, then a magnetic field should be able to produce a current. The concept of the "field"

Animations

---

[1] Hans Christian Oersted was professor of physics at the University of Copenhagen in Denmark.

277

was not available at that time, and Faraday's goal was to show that a current could be produced by "magnetism."

He worked on this problem intermittently over a period of 10 years, until he was finally successful in 1831.[2] He wound two separate windings on an iron toroid and placed a galvanometer in one circuit and a battery in the other. Upon closing the battery circuit, he noted a momentary deflection of the galvanometer; a similar deflection in the opposite direction occurred when the battery was disconnected. This, of course, was the first experiment he made involving a *changing* magnetic field, and he followed it with a demonstration that either a *moving* magnetic field or a moving coil could also produce a galvanometer deflection.

In terms of fields, we now say that a time-varying magnetic field produces an *electromotive force* (emf) that may establish a current in a suitable closed circuit. An electromotive force is merely a voltage that arises from conductors moving in a magnetic field or from changing magnetic fields, and we shall define it in this section. Faraday's law is customarily stated as

$$\text{emf} = -\frac{d\Phi}{dt} \text{ V} \tag{1}$$

Equation (1) implies a closed path, although not necessarily a closed conducting path; the closed path, for example, might include a capacitor, or it might be a purely imaginary line in space. The magnetic flux is that flux which passes through any and every surface whose perimeter is the closed path, and $d\Phi/dt$ is the time rate of change of this flux.

A nonzero value of $d\Phi/dt$ may result from any of the following situations:

1. A time-changing flux linking a stationary closed path
2. Relative motion between a steady flux and a closed path
3. A combination of the two

The minus sign is an indication that the emf is in such a direction as to produce a current whose flux, if added to the original flux, would reduce the magnitude of the emf. This statement that the induced voltage acts to produce an opposing flux is known as *Lenz's law*.[3]

If the closed path is that taken by an N-turn filamentary conductor, it is often sufficiently accurate to consider the turns as coincident and let

$$\text{emf} = -N\frac{d\Phi}{dt} \tag{2}$$

where Φ is now interpreted as the flux passing through any one of N coincident paths.

---

[2] Joseph Henry produced similar results at Albany Academy in New York at about the same time.

[3] Henri Frederic Emile Lenz was born in Germany but worked in Russia. He published his law in 1834.

We need to define emf as used in (1) or (2). The emf is obviously a scalar, and (perhaps not so obviously) a dimensional check shows that it is measured in volts. We define the emf as

$$\text{emf} = \oint \mathbf{E} \cdot d\mathbf{L} \tag{3}$$

and note that it is the voltage about a specific *closed path*. If any part of the path is changed, generally the emf changes. The departure from static results is clearly shown by (3), for an electric field intensity resulting from a static charge distribution must lead to zero potential difference about a closed path. In electrostatics, the line integral leads to a potential difference; with time-varying fields, the result is an emf or a voltage.

Replacing Φ in (1) with the surface integral of $\mathbf{B}$, we have

$$\text{emf} = \oint \mathbf{E} \cdot d\mathbf{L} = -\frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{S} \tag{4}$$

where the fingers of our right hand indicate the direction of the closed path, and our thumb indicates the direction of $d\mathbf{S}$. A flux density $\mathbf{B}$ in the direction of $d\mathbf{S}$ and increasing with time thus produces an average value of $\mathbf{E}$ which is *opposite* to the positive direction about the closed path. The right-handed relationship between the surface integral and the closed line integral in (4) should always be kept in mind during flux integrations and emf determinations.

We will divide our investigation into two parts by first finding the contribution to the total emf made by a changing field within a stationary path (transformer emf), and then we will consider a moving path within a constant (motional, or generator, emf).

We first consider a stationary path. The magnetic flux is the only time-varying quantity on the right side of (4), and a partial derivative may be taken under the integral sign,

$$\text{emf} = \oint \mathbf{E} \cdot d\mathbf{L} = -\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} \tag{5}$$

Before we apply this simple result to an example, let us obtain the point form of this integral equation. Applying Stokes' theorem to the closed line integral, we have

$$\int_S (\nabla \times \mathbf{E}) \cdot d\mathbf{S} = -\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S}$$

where the surface integrals may be taken over identical surfaces. The surfaces are perfectly general and may be chosen as differentials,

$$(\nabla \times \mathbf{E}) \cdot d\mathbf{S} = -\frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S}$$

and

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{6}$$

This is one of Maxwell's four equations as written in differential, or point, form, the form in which they are most generally used. Equation (5) is the integral form of this equation and is equivalent to Faraday's law as applied to a fixed path. If **B** is not a function of time, (5) and (6) evidently reduce to the electrostatic equations

$$\oint \mathbf{E} \cdot d\mathbf{L} = 0 \quad \text{(electrostatics)}$$

and

$$\nabla \times \mathbf{E} = 0 \quad \text{(electrostatics)}$$

As an example of the interpretation of (5) and (6), let us assume a simple magnetic field which increases exponentially with time within the cylindrical region $\rho < b$,

$$\mathbf{B} = B_0 e^{kt} \mathbf{a}_z \tag{7}$$

where $B_0 = $ constant. Choosing the circular path $\rho = a$, $a < b$, in the $z = 0$ plane, along which E_ϕ must be constant by symmetry, we then have from (5)

$$\text{emf} = 2\pi a E_\phi = -k B_0 e^{kt} \pi a^2$$

The emf around this closed path is $-k B_0 e^{kt} \pi a^2$. It is proportional to a^2 because the magnetic flux density is uniform and the flux passing through the surface at any instant is proportional to the area.

If we now replace a with ρ, $\rho < b$, the electric field intensity at any point is

$$\mathbf{E} = -\tfrac{1}{2} k B_0 e^{kt} \rho \mathbf{a}_\phi \tag{8}$$

Let us now attempt to obtain the same answer from (6), which becomes

$$(\nabla \times \mathbf{E})_z = -k B_0 e^{kt} = \frac{1}{\rho} \frac{\partial(\rho E_\phi)}{\partial \rho}$$

Multiplying by ρ and integrating from 0 to ρ (treating t as a constant, since the derivative is a partial derivative),

$$-\tfrac{1}{2} k B_0 e^{kt} \rho^2 = \rho E_\phi$$

or

$$\mathbf{E} = -\tfrac{1}{2} k B_0 e^{kt} \rho \mathbf{a}_\phi$$

once again.

If B_0 is considered positive, a filamentary conductor of resistance R would have a current flowing in the negative $\mathbf{a}_\phi$ direction, and this current would establish a flux within the circular loop in the negative $\mathbf{a}_z$ direction. Because E_ϕ increases exponentially with time, the current and flux do also, and thus they tend to reduce the time rate of increase of the applied flux and the resultant emf in accordance with Lenz's law.

Before leaving this example, it is well to point out that the given field **B** does not satisfy all of Maxwell's equations. Such fields are often assumed (*always* in ac-circuit problems) and cause no difficulty when they are interpreted properly. They
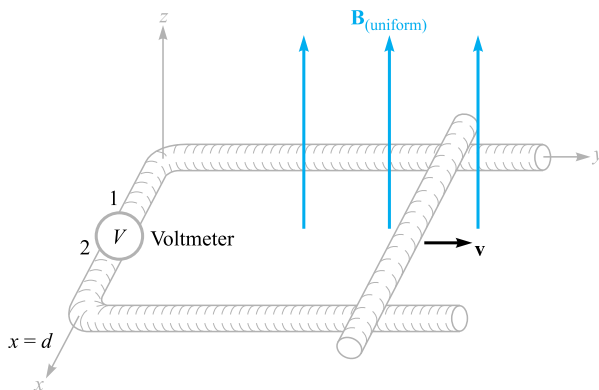
Figure 9.1 An example illustrating the application of Faraday's law to the case of a constant magnetic flux density B and a moving path. The shorting bar moves to the right with a velocity **v**, and the circuit is completed through the two rails and an extremely small high-resistance voltmeter. The voltmeter reading is $V_{12} = -Bvd$.

occasionally cause surprise, however. This particular field is discussed further in Problem 9.19 at the end of the chapter.

Now let us consider the case of a time-constant flux and a moving closed path. Before we derive any special results from Faraday's law (1), let us use the basic law to analyze the specific problem outlined in Figure 9.1. The closed circuit consists of two parallel conductors which are connected at one end by a high-resistance voltmeter of negligible dimensions and at the other end by a sliding bar moving at a velocity **v**. The magnetic flux density **B** is constant (in space and time) and is normal to the plane containing the closed path.

Let the position of the shorting bar be given by y; the flux passing through the surface within the closed path at any time t is then

$$\Phi = Byd$$

From (1), we obtain

$$\text{emf} = -\frac{d\Phi}{dt} = -B\frac{dy}{dt}d = -Bvd \tag{9}$$

The emf is defined as $\oint \mathbf{E} \cdot d\mathbf{L}$ and we have a conducting path, so we may actually determine **E** at every point along the closed path. We found in electrostatics that the tangential component of **E** is zero at the surface of a conductor, and we shall show in Section 9.4 that the tangential component is zero at the surface of a *perfect* conductor ($\sigma = \infty$) for all time-varying conditions. This is equivalent to saying that a perfect conductor is a "short circuit." The entire closed path in Figure 9.1 may be considered a perfect conductor, with the exception of the voltmeter. The actual computation of $\oint \mathbf{E} \cdot d\mathbf{L}$ then must involve no contribution along the entire moving bar, both rails, and the voltmeter leads. Because we are integrating in a counterclockwise direction

(keeping the interior of the positive side of the surface on our left as usual), the contribution $E \, \Delta L$ across the voltmeter must be $-Bvd$, showing that the electric field intensity in the instrument is directed from terminal 2 to terminal 1. For an upscale reading, the positive terminal of the voltmeter should therefore be terminal 2.

The direction of the resultant small current flow may be confirmed by noting that the enclosed flux is reduced by a clockwise current in accordance with Lenz's law. The voltmeter terminal 2 is again seen to be the positive terminal.

Let us now consider this example using the concept of *motional emf*. The force on a charge Q moving at a velocity $\mathbf{v}$ in a magnetic field $\mathbf{B}$ is

$$\boxed{\mathbf{F} = Q\mathbf{v} \times \mathbf{B}}$$

or

$$\frac{\mathbf{F}}{Q} = \mathbf{v} \times \mathbf{B} \tag{10}$$

The sliding conducting bar is composed of positive and negative charges, and each experiences this force. The force per unit charge, as given by (10), is called the *motional* electric field intensity $\mathbf{E}_m$,

$$\boxed{\mathbf{E}_m = \mathbf{v} \times \mathbf{B}} \tag{11}$$

If the moving conductor were lifted off the rails, this electric field intensity would force electrons to one end of the bar (the far end) until the *static field* due to these charges just balanced the field induced by the motion of the bar. The resultant tangential electric field intensity would then be zero along the length of the bar.

The motional emf produced by the moving conductor is then

$$\text{emf} = \oint \mathbf{E}_m \cdot d\mathbf{L} = \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{L} \tag{12}$$

where the last integral may have a nonzero value only along that portion of the path which is in motion, or along which $\mathbf{v}$ has some nonzero value. Evaluating the right side of (12), we obtain

$$\oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{L} = \int_d^0 vB \, dx = -Bvd$$
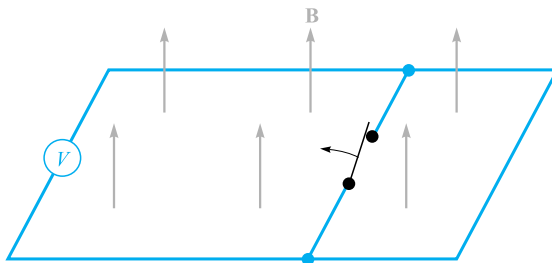
as before. This is the total emf, since $\mathbf{B}$ is not a function of time.

In the case of a conductor moving in a uniform constant magnetic field, we may therefore ascribe a motional electric field intensity $\mathbf{E}_m = \mathbf{v} \times \mathbf{B}$ to every portion of the moving conductor and evaluate the resultant emf by

$$\text{emf} = \oint \mathbf{E} \cdot d\mathbf{L} = \oint \mathbf{E}_m \cdot d\mathbf{L} = \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{L} \tag{13}$$

If the magnetic flux density is also changing with time, then we must include both contributions, the transformer emf (5) and the motional emf (12),

$$\text{emf} = \oint \mathbf{E} \cdot d\mathbf{L} = -\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} + \oint (\mathbf{v} \times \mathbf{B}) \cdot d\mathbf{L} \tag{14}$$

Figure 9.2 An apparent increase in flux linkages does not lead to an induced voltage when one part of a circuit is simply substituted for another by opening the switch. No indication will be observed on the voltmeter.

This expression is equivalent to the simple statement

$$\text{emf} = -\frac{d\Phi}{dt} \tag{1}$$

and either can be used to determine these induced voltages.

Although (1) appears simple, there are a few contrived examples in which its proper application is quite difficult. These usually involve sliding contacts or switches; they always involve the substitution of one part of a circuit by a new part.[4] As an example, consider the simple circuit of Figure 9.2, which contains several perfectly conducting wires, an ideal voltmeter, a uniform constant field **B**, and a switch. When the switch is opened, there is obviously more flux enclosed in the voltmeter circuit; however, it continues to read zero. The change in flux has not been produced by either a time-changing **B** [first term of (14)] or a conductor moving through a magnetic field [second part of (14)]. Instead, a new circuit has been substituted for the old. Thus it is necessary to use care in evaluating the change in flux linkages.

The separation of the emf into the two parts indicated by (14), one due to the time rate of change of **B** and the other to the motion of the circuit, is somewhat arbitrary in that it depends on the relative velocity of the *observer* and the system. A field that is changing with both time and space may look constant to an observer moving with the field. This line of reasoning is developed more fully in applying the special theory of relativity to electromagnetic theory.[5]

D9.1. Within a certain region, $\epsilon = 10^{-11}$ F/m and $\mu = 10^{-5}$ H/m. If $B_x = 2 \times 10^{-4} \cos 10^5 t \sin 10^{-3} y$ T: (*a*) use $\nabla \times \mathbf{H} = \epsilon \dfrac{\partial \mathbf{E}}{\partial t}$ to find **E**; (*b*) find the total magnetic flux passing through the surface $x = 0$, $0 < y < 40$ m, $0 < z < 2$ m,

---

[4] See Bewley, in References at the end of the chapter, particularly pp. 12–19.

[5] This is discussed in several of the references listed in the References at the end of the chapter. See Panofsky and Phillips, pp. 142–51; Owen, pp. 231–45; and Harman in several places.

at $t = 1 \ \mu s$; (c) find the value of the closed line integral of **E** around the peri-
meter of the given surface.

Ans. $-20\,000 \sin 10^5 t \cos 10^{-3} y \mathbf{a}_z$ V/m; 0.318 mWb; -3.19 V

D9.2. With reference to the sliding bar shown in Figure 9.1, let $d = 7$ cm,
$\mathbf{B} = 0.3\mathbf{a}_z$ T, and $\mathbf{v} = 0.1\mathbf{a}_y e^{20y}$ m/s. Let $y = 0$ at $t = 0$. Find: (a) $v(t = 0)$;
(b) $y(t = 0.1)$; (c) $v(t = 0.1)$; (d) V_{12} at $t = 0.1$.

Ans. 0.1 m/s; 1.12 cm; 0.125 m/s; -2.63 mV

9.2 DISPLACEMENT CURRENT

Faraday's experimental law has been used to obtain one of Maxwell's equations in
differential form,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{15}$$

which shows us that a time-changing magnetic field produces an electric field. Re-
membering the definition of curl, we see that this electric field has the special property
of circulation; its line integral about a general closed path is not zero. Now let us turn
our attention to the time-changing electric field.

We should first look at the point form of Ampère's circuital law as it applies to
steady magnetic fields,

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{16}$$

and show its inadequacy for time-varying conditions by taking the divergence of each
side,

$$\nabla \cdot \nabla \times \mathbf{H} \equiv 0 = \nabla \cdot \mathbf{J}$$

The divergence of the curl is identically zero, so $\nabla \cdot \mathbf{J}$ is also zero. However, the
equation of continuity,

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho_v}{\partial t}$$

then shows us that (16) can be true only if $\partial \rho_v / \partial t = 0$. This is an unrealistic limitation,
and (16) must be amended before we can accept it for time-varying fields. Suppose
we add an unknown term **G** to (16),

$$\nabla \times \mathbf{H} = \mathbf{J} + \mathbf{G}$$

Again taking the divergence, we have

$$0 = \nabla \cdot \mathbf{J} + \nabla \cdot \mathbf{G}$$

Thus

$$\nabla \cdot \mathbf{G} = \frac{\partial \rho_v}{\partial t}$$

Replacing ρ_v with $\nabla \cdot \mathbf{D}$,

$$\nabla \cdot \mathbf{G} = \frac{\partial}{\partial t}(\nabla \cdot \mathbf{D}) = \nabla \cdot \frac{\partial \mathbf{D}}{\partial t}$$

from which we obtain the simplest solution for $\mathbf{G}$,

$$\mathbf{G} = \frac{\partial \mathbf{D}}{\partial t}$$

Ampère's circuital law in point form therefore becomes

$$\boxed{\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}} \tag{17}$$

Equation (17) has not been derived. It is merely a form we have obtained that does not disagree with the continuity equation. It is also consistent with all our other results, and we accept it as we did each experimental law and the equations derived from it. We are building a theory, and we have every right to our equations *until they are proved wrong*. This has not yet been done.

We now have a second one of Maxwell's equations and shall investigate its significance. The additional term $\partial \mathbf{D}/\partial t$ has the dimensions of current density, amperes per square meter. Because it results from a time-varying electric flux density (or displacement density), Maxwell termed it a *displacement current density*. We sometimes denote it by $\mathbf{J}_d$:

$$\nabla \times \mathbf{H} = \mathbf{J} + \mathbf{J}_d$$
$$\mathbf{J}_d = \frac{\partial \mathbf{D}}{\partial t}$$

This is the third type of current density we have met. Conduction current density,

$$\mathbf{J} = \sigma \mathbf{E}$$

is the motion of charge (usually electrons) in a region of zero net charge density, and convection current density,

$$\mathbf{J} = \rho_v \mathbf{v}$$

is the motion of volume charge density. Both are represented by $\mathbf{J}$ in (17). Bound current density is, of course, included in $\mathbf{H}$. In a nonconducting medium in which no volume charge density is present, $\mathbf{J} = 0$, and then

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} \quad (\text{if } \mathbf{J} = 0) \tag{18}$$

Notice the symmetry between (18) and (15):

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{15}$$

Again, the analogy between the intensity vectors $\mathbf{E}$ and $\mathbf{H}$ and the flux density vectors $\mathbf{D}$ and $\mathbf{B}$ is apparent. We cannot place too much faith in this analogy, however, for it fails when we investigate forces on particles. The force on a charge is related to $\mathbf{E}$
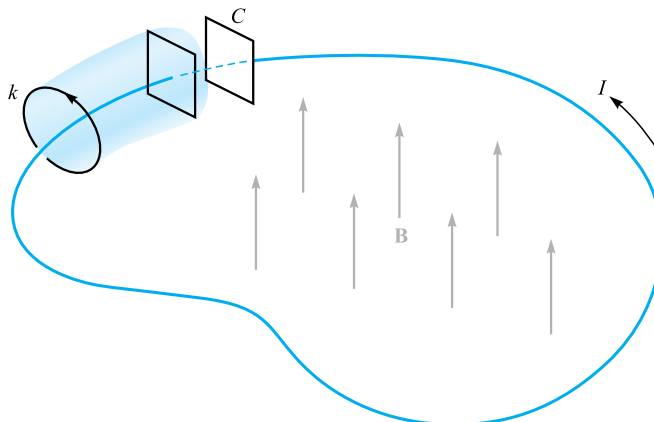
Figure 9.3 A filamentary conductor forms a loop connecting the two plates of a parallel-plate capacitor. A time-varying magnetic field inside the closed path produces an emf of $V_0 \cos \omega t$ around the closed path. The conduction current I is equal to the displacement current between the capacitor plates.

and to **B**, and some good arguments may be presented showing an analogy between **E** and **B** and between **D** and **H**. We omit them, however, and merely say that the concept of displacement current was probably suggested to Maxwell by the symmetry first mentioned in this paragraph.[6]

The total displacement current crossing any given surface is expressed by the surface integral,

$$I_d = \int_S \mathbf{J}_d \cdot d\mathbf{S} = \int_S \frac{\partial \mathbf{D}}{\partial t} \cdot d\mathbf{S}$$

and we may obtain the time-varying version of Ampère's circuital law by integrating (17) over the surface S,

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

and applying Stokes' theorem,

$$\oint \mathbf{H} \cdot d\mathbf{L} = I + I_d = I + \int_S \frac{\partial \mathbf{D}}{\partial t} \cdot d\mathbf{S} \qquad (19)$$

What is the nature of displacement current density? Let us study the simple circuit of Figure 9.3, which contains a filamentary loop and a parallel-plate capacitor. Within

---

[6] The analogy that relates **B** to **D** and **H** to **E** is strongly advocated by Fano, Chu, and Adler (see References for Chapter 6); the case for comparing **B** to **E** and **D** to **H** is presented in Halliday and Resnick (see References for this chapter).

the loop, a magnetic field varying sinusoidally with time is applied to produce an emf about the closed path (the filament plus the dashed portion between the capacitor plates), which we shall take as

$$\text{emf} = V_0 \cos \omega t$$

Using elementary circuit theory and assuming that the loop has negligible resistance and inductance, we may obtain the current in the loop as

$$I = -\omega C V_0 \sin \omega t$$
$$= -\omega \frac{\epsilon S}{d} V_0 \sin \omega t$$

where the quantities ϵ, S, and d pertain to the capacitor. Let us apply Ampère's circuital law about the smaller closed circular path k and neglect displacement current for the moment:

$$\oint_k \mathbf{H} \cdot d\mathbf{L} = I_k$$

The path and the value of $\mathbf{H}$ along the path are both definite quantities (although difficult to determine), and $\oint_k \mathbf{H} \cdot d\mathbf{L}$ is a definite quantity. The current I_k is that current through every surface whose perimeter is the path k. If we choose a simple surface punctured by the filament, such as the plane circular surface defined by the circular path k, the current is evidently the conduction current. Suppose now we consider the closed path k as the mouth of a paper bag whose bottom passes between the capacitor plates. The bag is not pierced by the filament, and the conductor current is zero. Now we need to consider displacement current, for within the capacitor

$$D = \epsilon E = \epsilon \left(\frac{V_0}{d} \cos \omega t \right)$$

and therefore

$$I_d = \frac{\partial D}{\partial t} S = -\omega \frac{\epsilon S}{d} V_0 \sin \omega t$$

This is the same value as that of the conduction current in the filamentary loop. Therefore the application of Ampère's circuital law, including displacement current to the path k, leads to a definite value for the line integral of $\mathbf{H}$. This value must be equal to the total current crossing the chosen surface. For some surfaces the current is almost entirely conduction current, but for those surfaces passing between the capacitor plates, the conduction current is zero, and it is the displacement current which is now equal to the closed line integral of $\mathbf{H}$.

Physically, we should note that a capacitor stores charge and that the electric field between the capacitor plates is much greater than the small leakage fields outside. We therefore introduce little error when we neglect displacement current on all those surfaces which do not pass between the plates.

Displacement current is associated with time-varying electric fields and therefore exists in all imperfect conductors carrying a time-varying conduction current. The last

part of the following drill problem indicates the reason why this additional current was never discovered experimentally.

D9.3. Find the amplitude of the displacement current density: (*a*) adjacent to an automobile antenna where the magnetic field intensity of an FM signal is $H_x = 0.15 \cos[3.12(3 \times 10^8 t - y)]$ A/m; (*b*) in the air space at a point within a large power distribution transformer where $\mathbf{B} = 0.8 \cos[1.257 \times 10^{-6}(3 \times 10^8 t - x)]\mathbf{a}_y$ T; (*c*) within a large, oil-filled power capacitor where $\epsilon_r = 5$ and $\mathbf{E} = 0.9 \cos[1.257 \times 10^{-6}(3 \times 10^8 t - z\sqrt{5})]\mathbf{a}_x$ MV/m; (*d*) in a metallic conductor at 60 Hz, if $\epsilon = \epsilon_0$, $\mu = \mu_0$, $\sigma = 5.8 \times 10^7$ S/m, and $\mathbf{J} = \sin(377t - 117.1z)\mathbf{a}_x$ MA/m².

Ans. 0.468 A/m²; 0.800 A/m²; 0.0150 A/m²; 57.6 pA/m²

9.3 MAXWELL'S EQUATIONS IN POINT FORM

We have already obtained two of Maxwell's equations for time-varying fields,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{20}$$

and

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{21}$$

The remaining two equations are unchanged from their non-time-varying form:

$$\nabla \cdot \mathbf{D} = \rho_v \tag{22}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{23}$$

Equation (22) essentially states that charge density is a source (or sink) of electric flux lines. Note that we can no longer say that *all* electric flux begins and terminates on charge, because the point form of Faraday's law (20) shows that **E**, and hence **D**, may have circulation if a changing magnetic field is present. Thus the lines of electric flux may form closed loops. However, the converse is still true, and every coulomb of charge must have one coulomb of electric flux diverging from it.

Equation (23) again acknowledges the fact that "magnetic charges," or poles, are not known to exist. Magnetic flux is always found in closed loops and never diverges from a point source.

These four equations form the basis of all electromagnetic theory. They are partial differential equations and relate the electric and magnetic fields to each other and to

their sources, charge and current density. The auxiliary equations relating **D** and **E**,

$$\mathbf{D} = \epsilon \mathbf{E} \tag{24}$$

relating **B** and **H**,

$$\mathbf{B} = \mu \mathbf{H} \tag{25}$$

defining conduction current density,

$$\mathbf{J} = \sigma \mathbf{E} \tag{26}$$

and defining convection current density in terms of the volume charge density ρ_v,

$$\mathbf{J} = \rho_v \mathbf{v} \tag{27}$$

are also required to define and relate the quantities appearing in Maxwell's equations.

The potentials V and **A** have not been included because they are not strictly necessary, although they are extremely useful. They will be discussed at the end of this chapter.

If we do not have "nice" materials to work with, then we should replace (24) and (25) with the relationships involving the polarization and magnetization fields,

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \tag{28}$$

$$\mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) \tag{29}$$

For linear materials we may relate **P** to **E**

$$\mathbf{P} = \chi_e \epsilon_0 \mathbf{E} \tag{30}$$

and **M** to **H**

$$\mathbf{M} = \chi_m \mathbf{H} \tag{31}$$

Finally, because of its fundamental importance we should include the Lorentz force equation, written in point form as the force per unit volume,

$$\mathbf{f} = \rho_v (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{32}$$

The following chapters are devoted to the application of Maxwell's equations to several simple problems.

D9.4. Let $\mu = 10^{-5}$ H/m, $\epsilon = 4 \times 10^{-9}$ F/m, $\sigma = 0$, and $\rho_v = 0$. Find k (including units) so that each of the following pairs of fields satisfies Maxwell's equations: (a) $\mathbf{D} = 6\mathbf{a}_x - 2y\mathbf{a}_y + 2z\mathbf{a}_z$ nC/m^2, $\mathbf{H} = kx\mathbf{a}_x + 10y\mathbf{a}_y - 25z\mathbf{a}_z$ A/m; (b) $\mathbf{E} = (20y - kt)\mathbf{a}_x$ V/m, $\mathbf{H} = (y + 2 \times 10^6 t)\mathbf{a}_z$ A/m.

Ans. 15 A/m^2; -2.5×10^8 V/(m·s)

9.4 MAXWELL'S EQUATIONS IN INTEGRAL FORM

The integral forms of Maxwell's equations are usually easier to recognize in terms of the experimental laws from which they have been obtained by a generalization process. Experiments must treat physical macroscopic quantities, and their results therefore are expressed in terms of integral relationships. A differential equation always represents a theory. Let us now collect the integral forms of Maxwell's equations from Section 9.3.

Integrating (20) over a surface and applying Stokes' theorem, we obtain Faraday's law,

$$\oint \mathbf{E} \cdot d\mathbf{L} = -\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} \tag{33}$$

and the same process applied to (21) yields Ampère's circuital law,

$$\oint \mathbf{H} \cdot d\mathbf{L} = I + \int_S \frac{\partial \mathbf{D}}{\partial t} \cdot d\mathbf{S} \tag{34}$$

Gauss's laws for the electric and magnetic fields are obtained by integrating (22) and (23) throughout a volume and using the divergence theorem:

$$\oint_S \mathbf{D} \cdot d\mathbf{S} = \int_{\text{vol}} \rho_v dv \tag{35}$$

$$\oint_S \mathbf{B} \cdot d\mathbf{S} = 0 \tag{36}$$

These four integral equations enable us to find the boundary conditions on **B**, **D**, **H**, and **E**, which are necessary to evaluate the constants obtained in solving Maxwell's equations in partial differential form. These boundary conditions are in general unchanged from their forms for static or steady fields, and the same methods may be used to obtain them. Between any two real physical media (where **K** must be zero on the boundary surface), (33) enables us to relate the tangential **E**-field components,

$$E_{t1} = E_{t2} \tag{37}$$

and from (34),

$$H_{t1} = H_{t2} \tag{38}$$

The surface integrals produce the boundary conditions on the normal components,

$$D_{N1} - D_{N2} = \rho_S \tag{39}$$

and

$$B_{N1} = B_{N2} \tag{40}$$

It is often desirable to idealize a physical problem by assuming a perfect conductor for which σ is infinite but $\mathbf{J}$ is finite. From Ohm's law, then, in a perfect conductor,

$$\mathbf{E} = 0$$

and it follows from the point form of Faraday's law that

$$\mathbf{H} = 0$$

for time-varying fields. The point form of Ampère's circuital law then shows that the finite value of $\mathbf{J}$ is

$$\mathbf{J} = 0$$

and current must be carried on the conductor surface as a surface current $\mathbf{K}$. Thus, if region 2 is a perfect conductor, (37) to (40) become, respectively,

$$E_{t1} = 0 \tag{41}$$

$$H_{t1} = K \quad (\mathbf{H}_{t1} = \mathbf{K} \times \mathbf{a}_N) \tag{42}$$

$$D_{N1} = \rho_s \tag{43}$$

$$B_{N1} = 0 \tag{44}$$

where $\mathbf{a}_N$ is an outward normal at the conductor surface.

Note that surface charge density is considered a physical possibility for either dielectrics, perfect conductors, or imperfect conductors, but that surface *current* density is assumed only in conjunction with perfect conductors.

The preceding boundary conditions are a very necessary part of Maxwell's equations. All real physical problems have boundaries and require the solution of Maxwell's equations in two or more regions and the matching of these solutions at the boundaries. In the case of perfect conductors, the solution of the equations within the conductor is trivial (all time-varying fields are zero), but the application of the boundary conditions (41) to (44) may be very difficult.

Certain fundamental properties of wave propagation are evident when Maxwell's equations are solved for an *unbounded* region. This problem is treated in Chapter 11. It represents the simplest application of Maxwell's equations because it is the only problem which does not require the application of any boundary conditions.

D9.5. The unit vector $0.64\mathbf{a}_x + 0.6\mathbf{a}_y - 0.48\mathbf{a}_z$ is directed from region 2 ($\epsilon_r = 2, \mu_r = 3, \sigma_2 = 0$) toward region 1 ($\epsilon_{r1} = 4, \mu_{r1} = 2, \sigma_1 = 0$). If $\mathbf{B}_1 = (\mathbf{a}_x - 2\mathbf{a}_y + 3\mathbf{a}_z) \sin 300t$ T at point P in region 1 adjacent to the boundary, find the amplitude at P of: (*a*) $\mathbf{B}_{N1}$; (*b*) $\mathbf{B}_{t1}$; (*c*) $\mathbf{B}_{N2}$; (*d*) $\mathbf{B}_2$.

Ans. 2.00 T; 3.16 T; 2.00 T; 5.15 T

D9.6. The surface $y = 0$ is a perfectly conducting plane, whereas the region $y > 0$ has $\epsilon_r = 5, \mu_r = 3$, and $\sigma = 0$. Let $\mathbf{E} = 20 \cos(2 \times 10^8 t - 2.58z)\mathbf{a}_y$ V/m for $y > 0$, and find at $t = 6$ ns; (*a*) ρ_S at $P(2, 0, 0.3)$; (*b*) $\mathbf{H}$ at P; (*c*) $\mathbf{K}$ at P.

Ans. 0.81 nC/m^2; $-62.3\mathbf{a}_x$ mA/m; $-62.3\mathbf{a}_z$ mA/m

9.5 THE RETARDED POTENTIALS

The time-varying potentials, usually called *retarded* potentials for a reason that we will see shortly, find their greatest application in radiation problems (to be addressed in Chapter 14) in which the distribution of the source is known approximately. We should remember that the scalar electric potential V may be expressed in terms of a static charge distribution,

$$V = \int_{\text{vol}} \frac{\rho_v dv}{4\pi \epsilon R} \quad \text{(static)} \tag{45}$$

and the vector magnetic potential may be found from a current distribution which is constant with time,

$$\mathbf{A} = \int_{\text{vol}} \frac{\mu \mathbf{J}\, dv}{4\pi R} \quad \text{(dc)} \tag{46}$$

The differential equations satisfied by V,

$$\nabla^2 V = -\frac{\rho_v}{\epsilon} \quad \text{(static)} \tag{47}$$

and $\mathbf{A}$,

$$\nabla^2 \mathbf{A} = -\mu \mathbf{J} \quad \text{(dc)} \tag{48}$$

may be regarded as the point forms of the integral equations (45) and (46), respectively.

Having found V and $\mathbf{A}$, the fundamental fields are then simply obtained by using the gradient,

$$\mathbf{E} = -\nabla V \quad \text{(static)} \tag{49}$$

or the curl,

$$\mathbf{B} = \nabla \times \mathbf{A} \quad \text{(dc)} \tag{50}$$

We now wish to define suitable time-varying potentials which are consistent with the preceding expressions when only static charges and direct currents are involved.

Equation (50) apparently is still consistent with Maxwell's equations. These equations state that $\nabla \cdot \mathbf{B} = 0$, and the divergence of (50) leads to the divergence of

the curl that is identically zero. Let us therefore tentatively accept (50) as satisfactory for time-varying fields and turn our attention to (49).

The inadequacy of (49) is obvious because application of the curl operation to each side and recognition of the curl of the gradient as being identically zero confront us with $\nabla \times \mathbf{E} = 0$. However, the point form of Faraday's law states that $\nabla \times \mathbf{E}$ is not generally zero, so let us try to effect an improvement by adding an unknown term to (49),

$$\mathbf{E} = -\nabla V + \mathbf{N}$$

taking the curl,

$$\nabla \times \mathbf{E} = 0 + \nabla \times \mathbf{N}$$

using the point form of Faraday's law,

$$\nabla \times \mathbf{N} = -\frac{\partial \mathbf{B}}{\partial t}$$

and using (50), giving us

$$\nabla \times \mathbf{N} = -\frac{\partial}{\partial t}(\nabla \times \mathbf{A})$$

or

$$\nabla \times \mathbf{N} = -\nabla \times \frac{\partial \mathbf{A}}{\partial t}$$

The simplest solution of this equation is

$$\mathbf{N} = -\frac{\partial \mathbf{A}}{\partial t}$$

and this leads to

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t} \tag{51}$$

We still must check (50) and (51) by substituting them into the remaining two of Maxwell's equations:

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}$$

$$\nabla \cdot \mathbf{D} = \rho_v$$

Doing this, we obtain the more complicated expressions

$$\frac{1}{\mu} \nabla \times \nabla \times \mathbf{A} = \mathbf{J} + \epsilon \left(-\nabla \frac{\partial V}{\partial t} - \frac{\partial^2 \mathbf{A}}{\partial t^2} \right)$$

and

$$\epsilon \left(-\nabla \cdot \nabla V - \frac{\partial}{\partial t} \nabla \cdot \mathbf{A} \right) = \rho_v$$

or

$$\nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mu \mathbf{J} - \mu\epsilon \left(\nabla \frac{\partial V}{\partial t} + \frac{\partial^2 \mathbf{A}}{\partial t^2} \right) \tag{52}$$

and

$$\nabla^2 V + \frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = -\frac{\rho_v}{\epsilon} \tag{53}$$

There is no apparent inconsistency in (52) and (53). Under static or dc conditions $\nabla \cdot \mathbf{A} = 0$, and (52) and (53) reduce to (48) and (47), respectively. We will therefore assume that the time-varying potentials may be defined in such a way that $\mathbf{B}$ and $\mathbf{E}$ may be obtained from them through (50) and (51). These latter two equations do not serve, however, to define $\mathbf{A}$ and V *completely*. They represent necessary, but not sufficient, conditions. Our initial assumption was merely that $B = \nabla \times \mathbf{A}$, and a vector cannot be defined by giving its curl alone. Suppose, for example, that we have a very simple vector potential field in which A_y and A_z are zero. Expansion of (50) leads to

$$B_x = 0$$
$$B_y = \frac{\partial A_x}{\partial z}$$
$$B_z = -\frac{\partial A_x}{\partial y}$$

and we see that no information is available about the manner in which A_x varies with x. This information could be found if we also knew the value of the divergence of $\mathbf{A}$, for in our example

$$\nabla \cdot \mathbf{A} = \frac{\partial A_x}{\partial x}$$

Finally, we should note that our information about $\mathbf{A}$ is given only as partial derivatives and that a space-constant term might be added. In all physical problems in which the region of the solution extends to infinity, this constant term must be zero, for there can be no fields at infinity.

Generalizing from this simple example, we may say that a vector field is defined completely when both its curl and divergence are given and when its value is known at any one point (including infinity). We are therefore at liberty to specify the divergence of $\mathbf{A}$, and we do so with an eye on (52) and (53), seeking the simplest expressions. We define

$$\nabla \cdot \mathbf{A} = -\mu\epsilon \frac{\partial V}{\partial t} \tag{54}$$

and (52) and (53) become

$$\nabla^2 \mathbf{A} = -\mu\mathbf{J} + \mu\epsilon \frac{\partial^2 \mathbf{A}}{\partial t^2} \tag{55}$$

and

$$\nabla^2 V = -\frac{\rho_v}{\epsilon} + \mu\epsilon \frac{\partial^2 V}{\partial t^2} \tag{56}$$

These equations are related to the wave equation, which will be discussed in Chapters 10 and 11. They show considerable symmetry, and we should be highly

pleased with our definitions of V and $\mathbf{A}$,

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{50}$$

$$\nabla \cdot \mathbf{A} = -\mu\epsilon \frac{\partial V}{\partial t} \tag{54}$$

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t} \tag{51}$$

The integral equivalents of (45) and (46) for the time-varying potentials follow from the definitions (50), (51), and (54), but we shall merely present the final results and indicate their general nature. In Chapter 11, we will find that any electromagnetic disturbance will travel at a velocity

$$v = \frac{1}{\sqrt{\mu\epsilon}}$$

through any homogeneous medium described by μ and ϵ. In the case of free space, this velocity turns out to be the velocity of light, approximately 3×10^8 m/s. It is logical, then, to suspect that the potential at any point is due not to the value of the charge density at some distant point at the same instant, but to its value at some previous time, because the effect propagates at a finite velocity. Thus (45) becomes

$$V = \int_{\text{vol}} \frac{[\rho_v]}{4\pi\epsilon R} dv \tag{57}$$

where $[\rho_v]$ indicates that every t appearing in the expression for ρ_v has been replaced by a *retarded* time,

$$t' = t - \frac{R}{v}$$

Thus, if the charge density throughout space were given by

$$\rho_v = e^{-r} \cos \omega t$$

then

$$[\rho_v] = e^{-r} \cos\left[\omega\left(t - \frac{R}{v}\right)\right]$$

where R is the distance between the differential element of charge being considered and the point at which the potential is to be determined.

The retarded vector magnetic potential is given by

$$\mathbf{A} = \int_{\text{vol}} \frac{\mu[\mathbf{J}]}{4\pi R} dv \tag{58}$$

The use of a retarded time has resulted in the time-varying potentials being given the name of retarded potentials. In Chapter 14 we will apply (58) to the simple situation of a differential current element in which I is a sinusoidal function of time. Other simple applications of (58) are considered in several problems at the end of this chapter.

We may summarize the use of the potentials by stating that a knowledge of the distribution of ρ_v and **J** throughout space theoretically enables us to determine V and **A** from (57) and (58). The electric and magnetic fields are then obtained by applying (50) and (51). If the charge and current distributions are unknown, or reasonable approximations cannot be made for them, these potentials usually offer no easier path toward the solution than does the direct application of Maxwell's equations.

D9.7. A point charge of $4 \cos 10^8 \pi t \; \mu$C is located at $P_+(0, 0, 1.5)$, whereas $-4 \cos 10^8 \pi t \; \mu$C is at $P_-(0, 0, -1.5)$, both in free space. Find V at $P(r = 450, \theta, \phi = 0)$ at $t = 15$ ns for $\theta =:$ (a) $0°$; (b) $90°$; (c) $45°$.

Ans. 159.8 V; 0; 143 V

REFERENCES

1. Bewley, L. V. *Flux Linkages and Electromagnetic Induction*. New York: Macmillan, 1952. This little book discusses many of the paradoxical examples involving induced (?) voltages.

2. Faraday, M. *Experimental Researches in Electricity*. London: B. Quaritch, 1839, 1855. Very interesting reading of early scientific research. A more recent and available source is *Great Books of the Western World,* vol. 45, Encyclopaedia Britannica, Inc., Chicago, 1952.

3. Halliday, D., R. Resnick, and J. Walker. *Fundamentals of Physics*. 5th ed. New York: John Wiley & Sons, 1997. This text is widely used in the first university-level course in physics.

4. Harman, W. W. *Fundamentals of Electronic Motion*. New York: McGraw-Hill, 1953. Relativistic effects are discussed in a clear and interesting manner.

5. Nussbaum, A. *Electromagnetic Theory for Engineers and Scientists*. Englewood Cliffs, N.J.: Prentice-Hall, 1965. See the rocket-generator example beginning on p. 211.

6. Owen, G. E. *Electromagnetic Theory*. Boston: Allyn and Bacon, 1963. Faraday's law is discussed in terms of the frame of reference in Chapter 8.

7. Panofsky, W. K. H., and M. Phillips. *Classical Electricity and Magnetism*. 2d ed. Reading, Mass.: Addison-Wesley, 1962. Relativity is treated at a moderately advanced level in Chapter 15.

CHAPTER 9   PROBLEMS

Quizzes

9.1 In Figure 9.4, let $B = 0.2 \cos 120 \pi t$ T, and assume that the conductor joining the two ends of the resistor is perfect. It may be assumed that the magnetic field produced by $I(t)$ is negligible. Find (a) $V_{ab}(t)$; (b) $I(t)$.

9.2 In the example described by Figure 9.1, replace the constant magnetic flux density by the time-varying quantity **B** $= B_0 \sin \omega t \; \mathbf{a}_z$. Assume that **U** is constant and that the displacement y of the bar is zero at $t = 0$. Find the emf at any time, t.
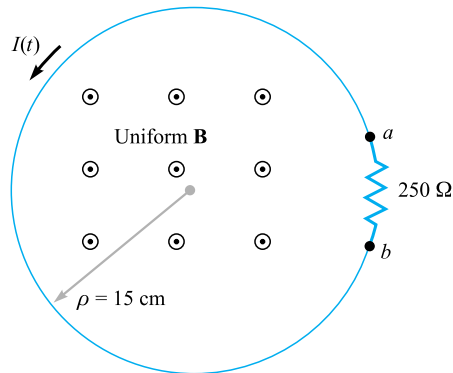
Figure 9.4 See Problem 9.1.

9.3 Given $\mathbf{H} = 300\mathbf{a}_z \cos(3 \times 10^8 t - y)$ A/m in free space, find the emf developed in the general $\mathbf{a}_\phi$ direction about the closed path having corners at (a) $(0, 0, 0)$, $(1, 0, 0)$, $(1, 1, 0)$, and $(0, 1, 0)$; (b) $(0, 0, 0)$ $(2\pi, 0, 0)$, $(2\pi, 2\pi, 0)$, and $(0, 2\pi, 0)$.

9.4 A rectangular loop of wire containing a high-resistance voltmeter has corners initially at $(a/2, b/2, 0)$, $(-a/2, b/2, 0)$, $(-a/2, -b/2, 0)$, and $(a/2, -b/2, 0)$. The loop begins to rotate about the x axis at constant angular velocity ω, with the first-named corner moving in the $\mathbf{a}_z$ direction at $t = 0$. Assume a uniform magnetic flux density $\mathbf{B} = B_0 \mathbf{a}_z$. Determine the induced emf in the rotating loop and specify the direction of the current.

9.5 The location of the sliding bar in Figure 9.5 is given by $x = 5t + 2t^3$, and the separation of the two rails is 20 cm. Let $\mathbf{B} = 0.8x^2 \mathbf{a}_z$ T. Find the voltmeter reading at (a) $t = 0.4$ s; (b) $x = 0.6$ m.

9.6 Let the wire loop of Problem 9.4 be stationary in its $t = 0$ position and find the induced emf that results from a magnetic flux density given by $\mathbf{B}(y, t) = B_0 \cos(\omega t - \beta y) \mathbf{a}_z$, where ω and β are constants.
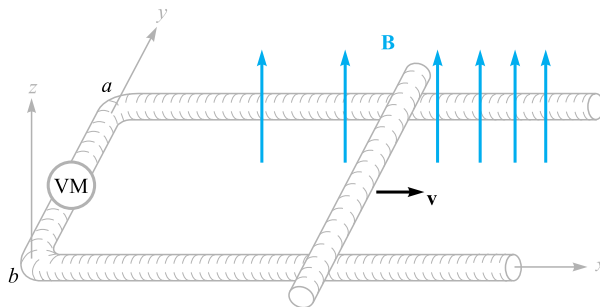
Figure 9.5 See Problem 9.5.
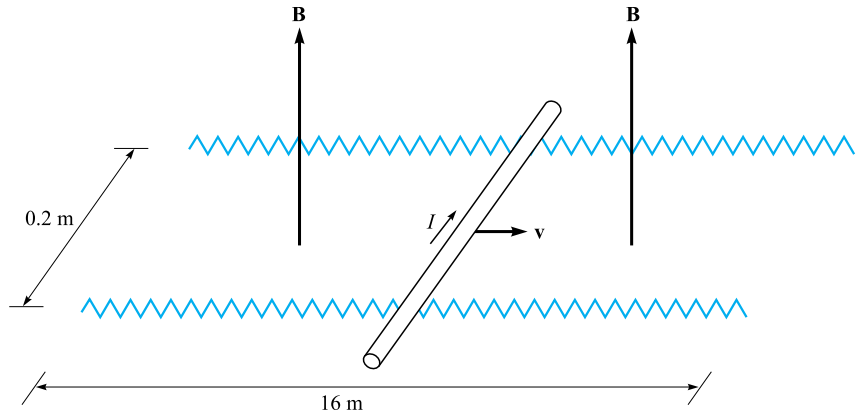
Figure 9.6 See Problem 9.7.

9.7 The rails in Figure 9.6 each have a resistance of 2.2 Ω/m. The bar moves to the right at a constant speed of 9 m/s in a uniform magnetic field of 0.8 T. Find $I(t)$, $0 < t < 1$ s, if the bar is at $x = 2$ m at $t = 0$ and (a) a 0.3 Ω resistor is present across the left end with the right end open-circuited; (b) a 0.3 Ω resistor is present across each end.

9.8 A perfectly conducting filament is formed into a circular ring of radius a. At one point, a resistance R is inserted into the circuit, and at another a battery of voltage V_0 is inserted. Assume that the loop current itself produces negligible magnetic field. (a) Apply Faraday's law, Eq. (4), evaluating each side of the equation carefully and independently to show the equality; (b) repeat part a, assuming the battery is removed, the ring is closed again, and a linearly increasing **B** field is applied in a direction normal to the loop surface.

9.9 A square filamentary loop of wire is 25 cm on a side and has a resistance of 125 Ω per meter length. The loop lies in the $z = 0$ plane with its corners at $(0, 0, 0)$, $(0.25, 0, 0)$, $(0.25, 0.25, 0)$, and $(0, 0.25, 0)$ at $t = 0$. The loop is moving with a velocity $v_y = 50$ m/s in the field $B_z = 8 \cos(1.5 \times 10^8 t - 0.5x)\,\mu$T. Develop a function of time that expresses the ohmic power being delivered to the loop.

9.10 (a) Show that the ratio of the amplitudes of the conduction current density and the displacement current density is $\sigma/\omega\epsilon$ for the applied field $E = E_m \cos \omega t$. Assume $\mu = \mu_0$. (b) What is the amplitude ratio if the applied field is $E = E_m e^{-t/\tau}$, where τ is real?

9.11 Let the internal dimensions of a coaxial capacitor be $a = 1.2$ cm, $b = 4$ cm, and $l = 40$ cm. The homogeneous material inside the capacitor has the parameters $\epsilon = 10^{-11}$ F/m, $\mu = 10^{-5}$ H/m, and $\sigma = 10^{-5}$ S/m. If the electric field intensity is $\mathbf{E} = (10^6/\rho)\cos 10^5 t\, \mathbf{a}_\rho$ V/m, find (a) **J**; (b) the

total conduction current I_c through the capacitor; (c) the total displacement current I_d through the capacitor; (d) the ratio of the amplitude of I_d to that of I_c, the quality factor of the capacitor.

9.12 Find the displacement current density associated with the magnetic field
$\mathbf{H} = A_1 \sin(4x) \cos(\omega t - \beta z) \mathbf{a}_x + A_2 \cos(4x) \sin(\omega t - \beta z) \mathbf{a}_z$.

9.13 Consider the region defined by $|x|$, $|y|$, and $|z| < 1$. Let $\epsilon_r = 5$, $\mu_r = 4$, and $\sigma = 0$. If $J_d = 20 \cos(1.5 \times 10^8 t - bx) \mathbf{a}_y \, \mu A/m^2$ (a) find $\mathbf{D}$ and $\mathbf{E}$; (b) use the point form of Faraday's law and an integration with respect to time to find $\mathbf{B}$ and $\mathbf{H}$; (c) use $\nabla \times \mathbf{H} = \mathbf{J}_d + \mathbf{J}$ to find $\mathbf{J}_d$. (d) What is the numerical value of b?

9.14 A voltage source $V_0 \sin \omega t$ is connected between two concentric conducting spheres, $r = a$ and $r = b$, $b > a$, where the region between them is a material for which $\epsilon = \epsilon_r \epsilon_0$, $\mu = \mu_0$, and $\sigma = 0$. Find the total displacement current through the dielectric and compare it with the source current as determined from the capacitance (Section 6.3) and circuit-analysis methods.

9.15 Let $\mu = 3 \times 10^{-5}$ H/m, $\epsilon = 1.2 \times 10^{-10}$ F/m, and $\sigma = 0$ everywhere. If $\mathbf{H} = 2 \cos(10^{10}t - \beta x) \mathbf{a}_z$ A/m, use Maxwell's equations to obtain expressions for $\mathbf{B}$, $\mathbf{D}$, $\mathbf{E}$, and β.

9.16 Derive the continuity equation from Maxwell's equations.

9.17 The electric field intensity in the region $0 < x < 5$, $0 < y < \pi/12$, $0 < z < 0.06$ m in free space is given by $\mathbf{E} = C \sin 12y \sin az \cos 2 \times 10^{10} t \mathbf{a}_x$ V/m. Beginning with the $\nabla \times \mathbf{E}$ relationship, use Maxwell's equations to find a numerical value for a, if it is known that a is greater than zero.

9.18 The parallel-plate transmission line shown in Figure 9.7 has dimensions $b = 4$ cm and $d = 8$ mm, while the medium between the plates is characterized by $\mu_r = 1$, $\epsilon_r = 20$, and $\sigma = 0$. Neglect fields outside the dielectric. Given the field $\mathbf{H} = 5 \cos(10^9 t - \beta z) \mathbf{a}_y$ A/m, use Maxwell's

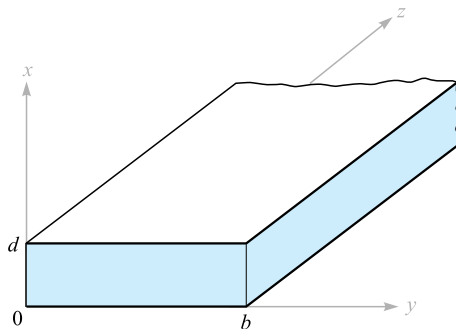**Figure 9.7** See Problem 9.18.

equations to help find (*a*) β, if $\beta > 0$; (*b*) the displacement current density at $z = 0$; (*c*) the total displacement current crossing the surface $x = 0.5d$, $0 < y < b$, $0 < z < 0.1$ m in the $\mathbf{a}_x$ direction.

9.19 In Section 9.1, Faraday's law was used to show that the field $\mathbf{E} = -\frac{1}{2}kB_0e^{kt}\rho\mathbf{a}_\phi$ results from the changing magnetic field $\mathbf{B} = B_0e^{kt}\mathbf{a}_z$. (*a*) Show that these fields do not satisfy Maxwell's other curl equation. (*b*) If we let $B_0 = 1$ T and $k = 10^6 \ s^{-1}$, we are establishing a fairly large magnetic flux density in 1 μs. Use the $\nabla \times \mathbf{H}$ equation to show that the rate at which B_z should (but does not) change with ρ is only about 5×10^{-6} T per meter in free space at $t = 0$.

9.20 Given Maxwell's equations in point form, assume that all fields vary as e^{st} and write the equations without explicitly involving time.

9.21 (*a*) Show that under static field conditions, Eq. (55) reduces to Ampère's circuital law. (*b*) Verify that Eq. (51) becomes Faraday's law when we take the curl.

9.22 In a sourceless medium in which $\mathbf{J} = 0$ and $\rho_v = 0$, assume a rectangular coordinate system in which $\mathbf{E}$ and $\mathbf{H}$ are functions only of z and t. The medium has permittivity ϵ and permeability μ. (*a*) If $\mathbf{E} = E_x\mathbf{a}_x$ and $\mathbf{H} = H_y\mathbf{a}_y$, begin with Maxwell's equations and determine the second-order partial differential equation that E_x must satisfy. (*b*) Show that $E_x = E_0\cos(\omega t - \beta z)$ is a solution of that equation for a particular value of β. (*c*) Find β as a function of given parameters.

9.23 In region 1, $z < 0$, $\epsilon_1 = 2 \times 10^{-11}$ F/m, $\mu_1 = 2 \times 10^{-6}$ H/m, and $\sigma_1 = 4 \times 10^{-3}$ S/m; in region 2, $z > 0$, $\epsilon_2 = \epsilon_1/2$, $\mu_2 = 2\mu_1$, and $\sigma_2 = \sigma_1/4$. It is known that $\mathbf{E}_1 = (30\mathbf{a}_x + 20\mathbf{a}_y + 10\mathbf{a}_z)\cos 10^9t$ V/m at $P(0, 0, 0^-)$. (*a*) Find $\mathbf{E}_{N1}$, $\mathbf{E}_{t1}$, $\mathbf{D}_{N1}$, and $\mathbf{D}_{t1}$ at P_1. (*b*) Find $\mathbf{J}_{N1}$ and $\mathbf{J}_{t1}$ at P_1. (*c*) Find $\mathbf{E}_{t2}$, $\mathbf{D}_{t2}$, and $\mathbf{J}_{t2}$ at $P_2(0, 0, 0^+)$. (*d*) (Harder) Use the continuity equation to help show that $J_{N1} - J_{N2} = \partial D_{N2}/\partial t - \partial D_{N1}/\partial t$, and then determine $\mathbf{D}_{N2}$, $\mathbf{J}_{N2}$, and $\mathbf{E}_{N2}$.

9.24 A vector potential is given as $\mathbf{A} = A_0\cos(\omega t - kz)\mathbf{a}_y$. (*a*) Assuming as many components as possible are zero, find $\mathbf{H}$, $\mathbf{E}$, and V. (*b*) Specify k in terms of A_0, ω, and the constants of the lossless medium, ϵ and μ.

9.25 In a region where $\mu_r = \epsilon_r = 1$ and $\sigma = 0$, the retarded potentials are given by $V = x(z - ct)$ V and $\mathbf{A} = x\left(\dfrac{z}{c} - t\right)\mathbf{a}_z$ Wb/m, where $c = 1\sqrt{\mu_0\epsilon_0}$. (*a*) Show that $\nabla \cdot \mathbf{A} = -\mu\epsilon\dfrac{\partial V}{\partial t}$. (*b*) Find $\mathbf{B}$, $\mathbf{H}$, $\mathbf{E}$, and $\mathbf{D}$. (*c*) Show that these results satisfy Maxwell's equations if $\mathbf{J}$ and ρ_v are zero.

9.26 Write Maxwell's equations in point form in terms of $\mathbf{E}$ and $\mathbf{H}$ as they apply to a sourceless medium, where $\mathbf{J}$ and ρ_v are both zero. Replace ϵ by μ, μ by ϵ, $\mathbf{E}$ by $\mathbf{H}$, and $\mathbf{H}$ by $-\mathbf{E}$, and show that the equations are unchanged. This is a more general expression of the *duality principle* in circuit theory.

Transmission Lines

Transmission lines are used to transmit electric energy and signals from one point to another, specifically from a source to a load. Examples include the connection between a transmitter and an antenna, connections between computers in a network, or connections between a hydroelectric generating plant and a substation several hundred miles away. Other familiar examples include the interconnects between components of a stereo system and the connection between a cable service provider and your television set. Examples that are less familiar include the connections between devices on a circuit board that are designed to operate at high frequencies.

What all of these examples have in common is that the devices to be connected are separated by distances on the order of a wavelength or much larger, whereas in basic circuit analysis methods, connections between elements are assumed to have negligible length. The latter condition enabled us, for example, to take for granted that the voltage across a resistor on one side of a circuit was exactly in phase with the voltage source on the other side, or, more generally, that the time measured at the source location is precisely the same time as measured at all other points in the circuit. When distances are sufficiently large between source and receiver, time delay effects become appreciable, leading to delay-induced phase differences. In short, we deal with *wave phenomena* on transmission lines in the same manner that we deal with point-to-point energy propagation in free space or in dielectrics.

The basic elements in a circuit, such as resistors, capacitors, inductors, and the connections between them, are considered *lumped* elements if the time delay in traversing the elements is negligible. On the other hand, if the elements or interconnections are large enough, it may be necessary to consider them as *distributed* elements. This means that their resistive, capacitive, and inductive characteristics must be evaluated on a per-unit-distance basis. Transmission lines have this property in general, and thus they become circuit elements in themselves, possessing impedances that contribute to the circuit problem. The basic rule is that one must consider elements as distributed if the propagation delay across the element dimension is on the order of the shortest time interval of interest. In the time-harmonic case,

this condition would lead to a measurable phase difference between each end of the device in question.

In this chapter, we investigate wave phenomena in transmission lines. Our objectives include (1) to understand how to treat transmission lines as circuit elements possessing complex impedances that are functions of line length and frequency, (2) to understand wave propagation on lines, including cases in which losses may occur, (3) to learn methods of combining different transmission lines to accomplish a desired objective, and (4) to understand transient phenomena on lines. ■

10.1 PHYSICAL DESCRIPTION OF TRANSMISSION LINE PROPAGATION

To obtain a feel for the manner in which waves propagate on transmission lines, the following demonstration may be helpful. Consider a *lossless* line, as shown in Figure 10.1. By lossless, we mean that all power that is launched into the line at the input end eventually arrives at the output end. A battery having voltage V_0 is connected to the input by closing switch S_1 at time $t = 0$. When the switch is closed, the effect is to launch voltage, $V^+ = V_0$. This voltage does not instantaneously appear everywhere on the line, but rather begins to travel from the battery toward the load resistor, R, at a certain velocity. The *wavefront,* represented by the vertical dashed line in Figure 10.1, represents the instantaneous boundary between the section of the line that has been charged to V_0 and the remaining section that is yet to be charged. It also represents the boundary between the section of the line that carries the charging current, I^+, and the remaining section that carries no current. Both current and voltage are discontinuous across the wavefront.

As the line charges, the wavefront moves from left to right at velocity ν, which is to be determined. On reaching the far end, all or a fraction of the wave voltage and current will reflect, depending on what the line is attached to. For example, if the resistor at the far end is left disconnected (switch S_2 is open), then all of the wavefront voltage will be reflected. If the resistor is connected, then some fraction of the incident voltage will reflect. The details of this will be treated in Section 10.9. Of interest at the moment are the factors that determine the wave velocity. The key
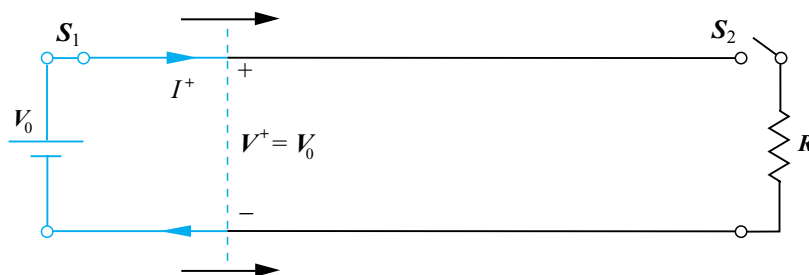
Figure 10.1 Basic transmission line circuit, showing voltage and current waves initiated by closing switch S_1.
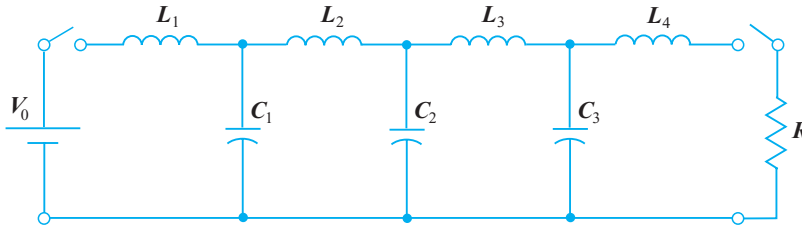
Figure 10.2 Lumped-element model of a transmission line. All inductance values are equal, as are all capacitance values.

to understanding and quantifying this is to note that the conducting transmission line will possess capacitance and inductance that are expressed on a per-unit-length basis. We have already derived expressions for these and evaluated them in Chapters 6 and 8 for certain transmission line geometries. Knowing these line characteristics, we can construct a model for the transmission line using lumped capacitors and inductors, as shown in Figure 10.2. The ladder network thus formed is referred to as a *pulse-forming network,* for reasons that will soon become clear.[1]

Consider now what happens when connecting the same switched voltage source to the network. Referring to Figure 10.2, on closing the switch at the battery location, current begins to increase in L_1, allowing C_1 to charge. As C_1 approaches full charge, current in L_2 begins to increase, allowing C_2 to charge next. This progressive charging process continues down the network, until all three capacitors are fully charged. In the network, a "wavefront" location can be identified as the point between two adjacent capacitors that exhibit the most difference between their charge levels. As the charging process continues, the wavefront moves from left to right. Its speed depends on how fast each inductor can reach its full-current state and, simultaneously, by how fast each capacitor is able to charge to full voltage. The wave is faster if the values of L_i and C_i are lower. We therefore expect the wave velocity to be inversely proportional to a function involving the product of inductance and capacitance. In the lossless transmission line, it turns out (as will be shown) that the wave velocity is given by $v = 1/\sqrt{LC}$, where L and C are specified per unit length.

Similar behavior is seen in the line and network when either is *initially charged*. In this case, the battery remains connected, and a resistor can be connected (by a switch) across the output end, as shown in Figure 10.2. In the case of the ladder network, the capacitor nearest the shunted end (C_3) will discharge through the resistor first, followed by the next-nearest capacitor, and so on. When the network is completely discharged, a voltage pulse has been formed across the resistor, and so we see why this ladder configuration is called a pulse-forming network. Essentially identical behavior is seen in a charged transmission line when connecting a resistor between conductors at the output end. The switched voltage exercises, as used in these discussions, are examples of transient problems on transmission lines. Transients will be treated in detail in Section 10.14. In the beginning, line responses to sinusoidal signals are emphasized.

---

[1] Designs and applications of pulse-forming networks are discussed in Reference 1.

Finally, we surmise that the existence of voltage and current across and within the transmission line conductors implies the existence of electric and magnetic fields in the space around the conductors. Consequently, we have two possible approaches to the analysis of transmission lines: (1) We can solve Maxwell's equations subject to the line configuration to obtain the fields, and with these find general expressions for the wave power, velocity, and other parameters of interest. (2) Or we can (for now) avoid the fields and solve for the voltage and current using an appropriate circuit model. It is the latter approach that we use in this chapter; the contribution of field theory is solely in the prior (and assumed) evaluation of the inductance and capacitance parameters. We will find, however, that circuit models become inconvenient or useless when losses in transmission lines are to be fully characterized, or when analyzing more complicated wave behavior (i.e., moding) which may occur as frequencies get high. The loss issues will be taken up in Section 10.5. Moding phenomena will be considered in Chapter 13.

10.2   THE TRANSMISSION LINE EQUATIONS

Our first goal is to obtain the differential equations, known as the *wave equations,* which the voltage or current must satisfy on a uniform transmission line. To do this, we construct a circuit model for an incremental length of line, write two circuit equations, and use these to obtain the wave equations.

Our circuit model contains the *primary constants* of the transmission line. These include the inductance, L, and capacitance, C, as well as the shunt conductance, G, and series resistance, R—all of which have values that are specified *per unit length*. The shunt conductance is used to model leakage current through the dielectric that may occur throughout the line length; the assumption is that the dielectric may possess conductivity, σ_d, in addition to a dielectric constant, ϵ_r, where the latter affects the capacitance. The series resistance is associated with any finite conductivity, σ_c, in the conductors. Either one of the latter parameters, R and G, will be responsible for power loss in transmission. In general, both are functions of frequency. Knowing the frequency and the dimensions, we can determine the values of R, G, L, and C by using formulas developed in earlier chapters.

We assume propagation in the $\mathbf{a}_z$ direction. Our model consists of a line section of length Δz containing resistance $R\Delta z$, inductance $L\Delta z$, conductance $G\Delta z$, and capacitance $C\Delta z$, as indicated in Figure 10.3. Because the section of the line looks the same from either end, we divide the series elements in half to produce a symmetrical network. We could equally well have placed half the conductance and half the capacitance at each end.

Our objective is to determine the manner and extent to which the output voltage and current are changed from their input values in the limit as the length approaches a very small value. We will consequently obtain a pair of differential equations that describe the rates of change of voltage and current with respect to z. In Figure 10.3, the input and output voltages and currents differ respectively by quantities ΔV and ΔI, which are to be determined. The two equations are obtained by successive applications of Kirchoff's voltage law (KVL) and Kirchoff's current law (KCL).
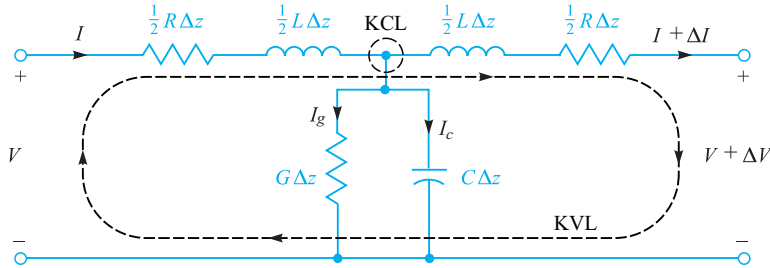
Figure 10.3 Lumped-element model of a short transmission line section with losses. The length of the section is Δz. Analysis involves applying Kirchoff's voltage and current laws (KVL and KCL) to the indicated loop and node, respectively.

First, KVL is applied to the loop that encompasses the entire section length, as shown in Figure 10.3:

$$V = \frac{1}{2}RI\Delta z + \frac{1}{2}L\frac{\partial I}{\partial t}\Delta z + \frac{1}{2}L\left(\frac{\partial I}{\partial t} + \frac{\partial \Delta I}{\partial t}\right)\Delta z$$
$$+ \frac{1}{2}R(I + \Delta I)\Delta z + (V + \Delta V) \tag{1}$$

We can solve Eq. (1) for the ratio, $\Delta V/\Delta z$, obtaining:

$$\frac{\Delta V}{\Delta z} = -\left(RI + L\frac{\partial I}{\partial t} + \frac{1}{2}L\frac{\partial \Delta I}{\partial t} + \frac{1}{2}R\Delta I\right) \tag{2}$$

Next, we write:

$$\Delta I = \frac{\partial I}{\partial z}\Delta z \qquad \text{and} \qquad \Delta V = \frac{\partial V}{\partial z}\Delta z \tag{3}$$

which are then substituted into (2) to result in

$$\frac{\partial V}{\partial z} = -\left(1 + \frac{\Delta z}{2}\frac{\partial}{\partial z}\right)\left(RI + L\frac{\partial I}{\partial t}\right) \tag{4}$$

Now, in the limit as Δz approaches zero (or a value small enough to be negligible), (4) simplifies to the final form:

$$\boxed{\frac{\partial V}{\partial z} = -\left(RI + L\frac{\partial I}{\partial t}\right)} \tag{5}$$

Equation (5) is the first of the two equations that we are looking for. To find the second equation, we apply KCL to the upper central node in the circuit of Figure 10.3, noting from the symmetry that the voltage at the node will be $V + \Delta V/2$:

$$I = I_g + I_c + (I + \Delta I) = G\Delta z\left(V + \frac{\Delta V}{2}\right)$$
$$+ C\Delta z\frac{\partial}{\partial t}\left(V + \frac{\Delta V}{2}\right) + (I + \Delta I) \tag{6}$$

Then, using (3) and simplifying, we obtain

$$\frac{\partial I}{\partial z} = -\left(1 + \frac{\Delta z}{2}\frac{\partial}{\partial z}\right)\left(GV + C\frac{\partial V}{\partial t}\right) \tag{7}$$

Again, we obtain the final form by allowing Δz to be reduced to a negligible magnitude. The result is

$$\frac{\partial I}{\partial z} = -\left(GV + C\frac{\partial V}{\partial t}\right) \tag{8}$$

The coupled differential equations, (5) and (8), describe the evolution of current and voltage in any transmission line. Historically, they have been referred to as the *telegraphist's equations*. Their solution leads to the wave equation for the transmission line, which we now undertake. We begin by differentiating Eq. (5) with respect to z and Eq. (8) with respect to t, obtaining:

$$\frac{\partial^2 V}{\partial z^2} = -R\frac{\partial I}{\partial z} - L\frac{\partial^2 I}{\partial t \partial z} \tag{9}$$

and

$$\frac{\partial I}{\partial z \partial t} = -G\frac{\partial V}{\partial t} - C\frac{\partial^2 V}{\partial t^2} \tag{10}$$

Next, Eqs. (8) and (10) are substituted into (9). After rearranging terms, the result is:

$$\frac{\partial^2 V}{\partial z^2} = LC\frac{\partial^2 V}{\partial t^2} + (LG + RC)\frac{\partial V}{\partial t} + RGV \tag{11}$$

An analogous procedure involves differentiating Eq. (5) with respect to t and Eq. (8) with respect to z. Then, Eq. (5) and its derivative are substituted into the derivative of (8) to obtain an equation for the current that is in identical form to that of (11):

$$\frac{\partial^2 I}{\partial z^2} = LC\frac{\partial^2 I}{\partial t^2} + (LG + RC)\frac{\partial I}{\partial t} + RGI \tag{12}$$

Equations (11) and (12) are the *general wave equations* for the transmission line. Their solutions under various conditions form a major part of our study.

10.3 LOSSLESS PROPAGATION

Lossless propagation means that power is not dissipated or otherwise deviated as the wave travels down the transmission line; all power at the input end eventually reaches the output end. More realistically, any mechanisms that would cause losses to occur have negligible effect. In our model, lossless propagation occurs when $R = G = 0$.

Under this condition, only the first term on the right-hand side of either Eq. (11) or Eq. (12) survives. Eq. (11), for example, becomes

$$\frac{\partial^2 V}{\partial z^2} = LC \frac{\partial^2 V}{\partial t^2} \tag{13}$$

In considering the voltage function that will satisfy (13), it is most expedient to simply state the solution, and then show that it is correct. The solution of (13) is of the form:

$$V(z, t) = f_1\left(t - \frac{z}{v}\right) + f_2\left(t + \frac{z}{v}\right) = V^+ + V^- \tag{14}$$

where v, the *wave velocity,* is a constant. The expressions $(t \pm z/v)$ are the arguments of functions f_1 and f_2. The identities of the functions themselves are not critical to the solution of (13). Therefore, f_1 and f_2 can be *any* function.

The arguments of f_1 and f_2 indicate, respectively, travel of the functions in the forward and backward z directions. We assign the symbols V^+ and V^- to identify the forward and backward voltage wave components. To understand the behavior, consider for example the value of f_1 (whatever this might be) at the zero value of its argument, occurring when $z = t = 0$. Now, as time increases to positive values (as it must), and if we are to keep track of $f_1(0)$, then the value of z must also increase to keep the argument $(t - z/v)$ equal to zero. The function f_1 therefore moves (or propagates) in the positive z direction. Using similar reasoning, the function f_2 will propagate in the *negative z* direction, as z in the argument $(t + z/v)$ must *decrease* to offset the increase in t. Therefore we associate the argument $(t - z/v)$ with *forward z* propagation, and the argument $(t + z/v)$ with *backward z* travel. This behavior occurs irrespective of what f_1 and f_2 are. As is evident in the argument forms, the propagation velocity is v in both cases.

We next verify that functions having the argument forms expressed in (14) are solutions to (13). First, we take partial derivatives of f_1, for example with respect to z and t. Using the chain rule, the z partial derivative is

$$\frac{\partial f_1}{\partial z} = \frac{\partial f_1}{\partial (t - z/v)} \frac{\partial (t - z/v)}{\partial z} = -\frac{1}{v} f_1' \tag{15}$$

where it is apparent that the primed function, f_1', denotes the derivative of f_1 with respect to its argument. The partial derivative with respect to time is

$$\frac{\partial f_1}{\partial t} = \frac{\partial f_1}{\partial (t - z/v)} \frac{\partial (t - z/v)}{\partial t} = f_1' \tag{16}$$

Next, the second partial derivatives with respect to z and t can be taken using similar reasoning:

$$\frac{\partial^2 f_1}{\partial z^2} = \frac{1}{v^2} f_1'' \qquad \text{and} \qquad \frac{\partial^2 f_1}{\partial t^2} = f_1'' \tag{17}$$

where f_1'' is the second derivative of f_1 with respect to its argument. The results in (17) can now be substituted into (13), obtaining

$$\frac{1}{v^2} f_1'' = LC f_1'' \qquad (18)$$

We now identify the wave velocity for lossless propagation, which is the condition for equality in (18):

$$v = \frac{1}{\sqrt{LC}} \qquad (19)$$

Performing the same procedure using f_2 (and its argument) leads to the same expression for v.

The form of v as expressed in Eq. (19) confirms our original expectation that the wave velocity would be in some inverse proportion to L and C. The same result will be true for current, as Eq. (12) under lossless conditions would lead to a solution of the form identical to that of (14), with velocity given by (19). What is not known yet, however, is the relation *between* voltage and current.

We have already found that voltage and current are related through the telegraphist's equations, (5) and (8). These, under lossless conditions ($R = G = 0$), become

$$\frac{\partial V}{\partial z} = -L \frac{\partial I}{\partial t} \qquad (20)$$

$$\frac{\partial I}{\partial z} = -C \frac{\partial V}{\partial t} \qquad (21)$$

Using the voltage function, we can substitute (14) into (20) and use the methods demonstrated in (15) to write

$$\frac{\partial I}{\partial t} = -\frac{1}{L} \frac{\partial V}{\partial z} = \frac{1}{Lv}(f_1' - f_2') \qquad (22)$$

We next integrate (22) over time, obtaining the current in terms of its forward and backward propagating components:

$$I(z, t) = \frac{1}{Lv} \left[f_1 \left(t - \frac{z}{v} \right) - f_2 \left(t + \frac{z}{v} \right) \right] = I^+ + I^- \qquad (23)$$

In performing this integration, all integration constants are set to zero. The reason for this, as demonstrated by (20) and (21), is that a time-varying voltage must lead to a time-varying current, with the reverse also true. The factor $1/Lv$ appearing in (23) multiplies voltage to obtain current, and so we identify the product Lv as the *characteristic impedance, Z_0,* of the lossless line. Z_0 is defined as the ratio of
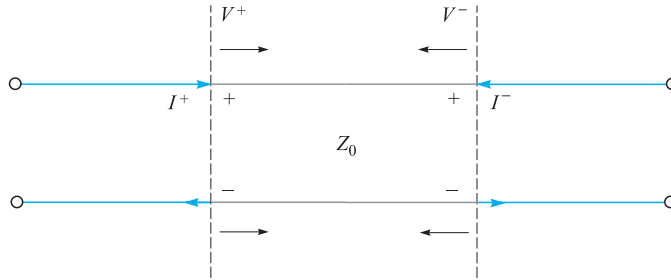
Figure 10.4 Current directions in waves having positive voltage polarity.

the voltage to the current in a single propagating wave. Using (19), we write the characteristic impedance as

$$Z_0 = Lv = \sqrt{\frac{L}{C}} \qquad (24)$$

By inspecting (14) and (23), we now note that

$$V^+ = Z_0 I^+ \qquad (25a)$$

and

$$V^- = -Z_0 I^- \qquad (25b)$$

The significance of the preceding relations can be seen in Figure 10.4. The figure shows forward- and backward-propagating voltage waves, V^+ and V^-, both of which have positive polarity. The currents that are associated with these voltages will flow in opposite directions. We define *positive current* as having a *clockwise* flow in the line, and *negative current* as having a *counterclockwise* flow. The minus sign in (25b) thus assures that negative current will be associated with a backward-propagating wave that has positive polarity. This is a general convention, applying to lines with losses also. Propagation with losses is studied by solving (11) under the assumption that either R or G (or both) are not zero. We will do this in Section 10.7 under the special case of sinusoidal voltages and currents. Sinusoids in lossless transmission lines are considered in Section 10.4.

10.4 LOSSLESS PROPAGATION OF SINUSOIDAL VOLTAGES

An understanding of sinusoidal waves on transmission lines is important because any signal that is transmitted in practice can be decomposed into a discrete or continuous summation of sinusoids. This is the basis of *frequency domain* analysis of signals on

lines. In such studies, the effect of the transmission line on any signal can be deter-mined by noting the effects on the frequency components. This means that one can effectively propagate the spectrum of a given signal, using frequency-dependent line parameters, and then reassemble the frequency components into the resultant signal in time domain. Our objective in this section is to obtain an understanding of sinusoidal propagation and the implications on signal behavior for the lossless line case.

We begin by assigning sinusoidal functions to the voltage functions in Eq. (14). Specifically, we consider a specific frequency, $f = \omega/2\pi$, and write $f_1 = f_2 = V_0 \cos(\omega t + \phi)$. By convention, the cosine function is chosen; the sine is obtainable, as we know, by setting $\phi = -\pi/2$. We next replace t with $(t \pm z/v_p)$, obtaining

$$\mathcal{V}(z,t) = |V_0| \cos[\omega(t \pm z/v_p) + \phi] = |V_0| \cos[\omega t \pm \beta z + \phi] \qquad (26)$$

where we have assigned a new notation to the velocity, which is now called the *phase velocity, v_p*. This is applicable to a pure sinusoid (having a single frequency) and will be found to depend on frequency in some cases. Choosing, for the moment, $\phi = 0$, we obtain the two possibilities of forward or backward z travel by choosing the minus or plus sign in (26). The two cases are:

$$\boxed{\mathcal{V}_f(z,t) = |V_0| \cos(\omega t - \beta z)} \qquad \text{(forward z propagation)} \qquad (27a)$$

and

$$\boxed{\mathcal{V}_b(z,t) = |V_0| \cos(\omega t + \beta z)} \qquad \text{(backward z propagation)} \qquad (27b)$$

where the magnitude factor, $|V_0|$, is the value of $\mathcal{V}$ at $z = 0, t = 0$. We define the *phase constant β*, obtained from (26), as

$$\boxed{\beta \equiv \frac{\omega}{v_p}} \qquad (28)$$

We refer to the solutions expressed in (27a) and (27b) as the *real instantaneous* forms of the transmission-line voltage. They are the mathematical representations of what one would experimentally measure. The terms ωt and βz, appearing in these equations, have units of angle and are usually expressed in radians. We know that ω is the radian time frequency, measuring phase shift *per unit time,* and it has units of rad/s. In a similar way, we see that β will be interpreted as a *spatial* frequency, which in the present case measures the phase shift *per unit distance* along the z direction. Its units are rad/m. If we were to fix the time at $t = 0$, Eqs. (27a) and (27b) would become

$$\mathcal{V}_f(z,0) = \mathcal{V}_b(z,0) = |V_0| \cos(\beta z) \qquad (29)$$

which we identify as a simple periodic function that repeats every incremental dis-tance λ, known as the *wavelength*. The requirement is that $\beta\lambda = 2\pi$, and so

$$\boxed{\lambda = \frac{2\pi}{\beta} = \frac{v_p}{f}} \qquad (30)$$

We next consider a point (such as a wave crest) on the cosine function of Eq. (27a), the occurrence of which requires the argument of the cosine to be an integer multiple of 2π. Considering the mth crest of the wave, the condition at $t = 0$ becomes

$$\beta z = 2m\pi$$

To keep track of this point on the wave, we require that the entire cosine argument be the same multiple of 2π for all time. From (27a) the condition becomes

$$\omega t - \beta z = \omega(t - z/v_p) = 2m\pi \tag{31}$$

Again, with increasing time, the position z must also increase in order to satisfy (31). Consequently the wave crest (and the entire wave) travels in the positive z direction at velocity v_p. Eq. (27b), having cosine argument ($\omega t + \beta z$), describes a wave that travels in the *negative z* direction, since as time increases, z must now *decrease* to keep the argument constant. Similar behavior is found for the wave current, but complications arise from line-dependent phase differences that occur between current and voltage. These issues are best addressed once we are familiar with complex analysis of sinusoidal signals.

10.5 COMPLEX ANALYSIS OF SINUSOIDAL WAVES

Expressing sinusoidal waves as complex functions is useful (and essentially indispensable) because it greatly eases the evaluation and visualization of phase that will be found to accumulate by way of many mechanisms. In addition, we will find many cases in which two or more sinusoidal waves must be combined to form a resultant wave—a task made much easier if complex analysis is used.

Expressing sinusoidal functions in complex form is based on the Euler identity:

$$e^{\pm jx} = \cos(x) \pm j\sin(x) \tag{32}$$

from which we may write the cosine and sine, respectively, as the real and imaginary parts of the complex exponent:

$$\cos(x) = \mathrm{Re}[e^{\pm jx}] = \frac{1}{2}(e^{jx} + e^{-jx}) = \frac{1}{2}e^{jx} + c.c. \tag{33a}$$

$$\sin(x) = \pm\mathrm{Im}[e^{\pm jx}] = \frac{1}{2j}(e^{jx} - e^{-jx}) = \frac{1}{2j}e^{jx} + c.c. \tag{33b}$$

where $j \equiv \sqrt{-1}$, and where $c.c.$ denotes the complex conjugate of the preceding term. The conjugate is formed by changing the sign of j wherever it appears in the complex expression.

We may next apply (33a) to our voltage wave function, Eq. (26):

$$\mathcal{V}(z, t) = |V_0|\cos[\omega t \pm \beta z + \phi] = \frac{1}{2}\underbrace{(|V_0|e^{j\phi})}_{V_0}e^{\pm j\beta z}e^{j\omega t} + c.c. \tag{34}$$

Note that we have arranged the phases in (34) such that we identify the *complex amplitude* of the wave as $V_0 = (|V_0|e^{j\phi})$. In future usage, a single symbol (V_0 in the

present example) will usually be used for the voltage or current amplitudes, with the understanding that these will generally be complex (having magnitude and phase).

Two additional definitions follow from Eq. (34). First, we define the *complex instantaneous* voltage as:

$$V_c(z, t) = V_0 e^{\pm j\beta z} e^{j\omega t} \qquad (35)$$

The *phasor* voltage is then formed by dropping the $e^{j\omega t}$ factor from the complex instantaneous form:

$$V_s(z) = V_0 e^{\pm j\beta z} \qquad (36)$$

The phasor voltage can be defined provided we have *sinusoidal steady-state* conditions—meaning that V_0 is independent of time. This has in fact been our assumption all along, because a time-varying amplitude would imply the existence of other frequency components in our signal. Again, we are treating only a single-frequency wave. The significance of the phasor voltage is that we are effectively letting time stand still and observing the stationary wave in space at $t = 0$. The processes of evaluating relative phases between various line positions and of combining multiple waves is made much simpler in phasor form. Again, this works only if all waves under consideration have the same frequency. With the definitions in (35) and (36), the real instantaneous voltage can be constructed using (34):

$$\mathcal{V}(z, t) = |V_0| \cos[\omega t \pm \beta z + \phi] = \text{Re}[V_c(z, t)] = \frac{1}{2} V_c + c.c. \qquad (37a)$$

Or, in terms of the phasor voltage:

$$\mathcal{V}(z, t) = |V_0| \cos[\omega t \pm \beta z + \phi] = \text{Re}[V_s(z) e^{j\omega t}] = \frac{1}{2} V_s(z) e^{j\omega t} + c.c. \qquad (37b)$$

In words, we may obtain our real sinusoidal voltage wave by multiplying the phasor voltage by $e^{j\omega t}$ (reincorporating the time dependence) and then taking the real part of the resulting expression. It is imperative that one becomes familiar with these relations and their meaning before proceeding further.

EXAMPLE 10.1

Two voltage waves having equal frequencies and amplitudes propagate in opposite directions on a lossless transmission line. Determine the total voltage as a function of time and position.

Solution. Because the waves have the same frequency, we can write their combination using their phasor forms. Assuming phase constant, β, and real amplitude, V_0, the two wave voltages combine in this way:

$$V_{sT}(z) = V_0 e^{-j\beta z} + V_0 e^{+j\beta z} = 2V_0 \cos(\beta z)$$

In real instantaneous form, this becomes

$$\mathcal{V}(z, t) = \text{Re}[2V_0 \cos(\beta z)e^{j\omega t}] = 2V_0 \cos(\beta z)\cos(\omega t)$$

We recognize this as a *standing wave,* in which the amplitude varies, as $\cos(\beta z)$, and oscillates in time, as $\cos(\omega t)$. Zeros in the amplitude (nulls) occur at fixed locations, $z_n = (m\pi)/(2\beta)$ where m is an odd integer. We extend the concept in Section 10.10, where we explore the *voltage standing wave ratio* as a measurement technique.

10.6 TRANSMISSION LINE EQUATIONS AND THEIR SOLUTIONS IN PHASOR FORM

We now apply our results of the previous section to the transmission line equations, beginning with the general wave equation, (11). This is rewritten as follows, for the real instantaneous voltage, $\mathcal{V}(z, t)$:

$$\frac{\partial^2 \mathcal{V}}{\partial z^2} = LC\frac{\partial^2 \mathcal{V}}{\partial t^2} + (LG + RC)\frac{\partial \mathcal{V}}{\partial t} + RG\mathcal{V} \tag{38}$$

We next substitute $\mathcal{V}(z, t)$ as given by the far right-hand side of (37b), noting that the complex conjugate term (*c.c.*) will form a separate redundant equation. We also use the fact that the operator $\partial/\partial t$, when applied to the complex form, is equivalent to multiplying by a factor of $j\omega$. After substitution, and after all time derivatives are taken, the factor $e^{j\omega t}$ divides out. We are left with the wave equation in terms of the phasor voltage:

$$\frac{d^2 V_s}{dz^2} = -\omega^2 LC V_s + j\omega(LG + RC)V_s + RG V_s \tag{39}$$

Rearranging terms leads to the simplified form:

$$\frac{d^2 V_s}{dz^2} = \underbrace{(R + j\omega L)}_{Z}\underbrace{(G + j\omega C)}_{Y} V_s = \gamma^2 V_s \tag{40}$$

where Z and Y, as indicated, are respectively the *net series impedance* and the *net shunt admittance* in the transmission line—both as per-unit-distance measures. The *propagation constant* in the line is defined as

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} = \sqrt{ZY} = \alpha + j\beta \tag{41}$$

The significance of the term will be explained in Section 10.7. For our immediate purposes, the solution of (40) will be

$$V_s(z) = V_0^+ e^{-\gamma z} + V_0^- e^{+\gamma z} \tag{42a}$$

The wave equation for current will be identical in form to (40). We therefore expect the phasor current to be in the form:

$$I_s(z) = I_0^+ e^{-\gamma z} + I_0^- e^{\gamma z} \qquad (42b)$$

The relation between the current and voltage waves is now found, as before, through the telegraphist's equations, (5) and (8). In a manner consistent with Eq. (37b), we write the sinusoidal current as

$$\mathcal{I}(z, t) = |I_0| \cos(\omega t \pm \beta z + \xi) = \frac{1}{2} \underbrace{(|I_0| e^{j\xi})}_{I_0} e^{\pm j\beta z} e^{j\omega t} + c.c. = \frac{1}{2} I_s(z) e^{j\omega t} + c.c.$$

$$(43)$$

Substituting the far right-hand sides of (37b) and (43) into (5) and (8) transforms the latter equations as follows:

$$\frac{\partial \mathcal{V}}{\partial z} = -\left(R\mathcal{I} + L\frac{\partial \mathcal{I}}{\partial t} \right) \quad \Rightarrow \quad \boxed{\frac{dV_s}{dz} = -(R + j\omega L)I_s = -ZI_s} \qquad (44a)$$

and

$$\frac{\partial \mathcal{I}}{\partial z} = -\left(G\mathcal{V} + C\frac{\partial \mathcal{V}}{\partial t} \right) \quad \Rightarrow \quad \boxed{\frac{dI_s}{dz} = -(G + j\omega C)V_s = -YV_s} \qquad (44b)$$

We can now substitute (42a) and (42b) into either (44a) or (44b) [we will use (44a)] to find:

$$-\gamma V_0^+ e^{-\gamma z} + \gamma V_0^- e^{\gamma z} = -Z(I_0^+ e^{-\gamma z} + I_0^- e^{\gamma z}) \qquad (45)$$

Next, equating coefficients of $e^{-\gamma z}$ and $e^{\gamma z}$, we find the general expression for the line characteristic impedance:

$$Z_0 = \frac{V_0^+}{I_0^+} = -\frac{V_0^-}{I_0^-} = \frac{Z}{\gamma} = \frac{Z}{\sqrt{ZY}} = \sqrt{\frac{Z}{Y}} \qquad (46)$$

Incorporating the expressions for Z and Y, we find the characteristic impedance in terms of our known line parameters:

$$\boxed{Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = |Z_0| e^{j\theta}} \qquad (47)$$

Note that with the voltage and current as given in (37b) and (43), we would identify the phase of the characteristic impedance, $\theta = \phi - \xi$.

EXAMPLE 10.2

A lossless transmission line is 80 cm long and operates at a frequency of 600 MHz. The line parameters are $L = 0.25\ \mu$H/m and $C = 100$ pF/m. Find the characteristic impedance, the phase constant, and the phase velocity.

Solution. Because the line is lossless, both R and G are zero. The characteristic impedance is

$$Z_0 = \sqrt{\frac{L}{C}} = \sqrt{\frac{0.25 \times 10^{-6}}{100 \times 10^{-12}}} = 50 \ \Omega$$

Because $\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)} = j\omega\sqrt{LC}$, we see that

$$\beta = \omega\sqrt{LC} = 2\pi(600 \times 10^6)\sqrt{(0.25 \times 10^{-6})(100 \times 10^{-12})} = 18.85 \ \text{rad/m}$$

Also,

$$v_p = \frac{\omega}{\beta} = \frac{2\pi(600 \times 10^6)}{18.85} = 2 \times 10^8 \ \text{m/s}$$

10.7 LOW-LOSS PROPAGATION

Having obtained the phasor forms of voltage and current in a general transmission line [Eqs. (42a) and (42b)], we can now look more closely at the significance of these results. First we incorporate (41) into (42a) to obtain

$$V_s(z) = V_0^+ e^{-\alpha z} e^{-j\beta z} + V_0^- e^{\alpha z} e^{j\beta z} \tag{48}$$

Next, multiplying (48) by $e^{j\omega t}$ and taking the real part gives the real instantaneous voltage:

$$\mathcal{V}(z, t) = V_0^+ e^{-\alpha z} \cos(\omega t - \beta z) + V_0^- e^{\alpha z} \cos(\omega t + \beta z) \tag{49}$$

In this exercise, we have assigned V_0^+ and V_0^- to be real. Eq. (49) is recognized as describing forward- and backward-propagating waves that diminish in amplitude with distance according to $e^{-\alpha z}$ for the forward wave, and $e^{\alpha z}$ for the backward wave. Both waves are said to *attenuate* with propagation distance at a rate determined by the *attenuation coefficient, α,* expressed in units of nepers/m [Np/m].[2]

The phase constant, β, found by taking the imaginary part of (41), is likely to be a somewhat complicated function, and will in general depend on R and G. Nevertheless, β is still defined as the ratio ω/v_p, and the wavelength is still defined as the distance that provides a phase shift of 2π rad, so that $\lambda = 2\pi/\beta$. By inspecting (41), we observe that losses in propagation are avoided (or $\alpha = 0$) only when $R = G = 0$. In that case, (41) gives $\gamma = j\beta = j\omega\sqrt{LC}$, and so $v_p = 1/\sqrt{LC}$, as we found before.

Expressions for α and β when losses are small can be readily obtained from (41). In the *low-loss approximation,* we require $R \ll \omega L$ and $G \ll \omega C$, a condition that

---

[2] The term *neper* was selected (by some poor speller) to honor John Napier, a Scottish mathematician who first proposed the use of logarithms.

is often true in practice. Before we apply these conditions, Eq. (41) can be written in the form:

$$\gamma = \alpha + j\beta = [(R + j\omega L)(G + j\omega C)]^{1/2}$$
$$= j\omega\sqrt{LC}\left[\left(1 + \frac{R}{j\omega L}\right)^{1/2}\left(1 + \frac{G}{j\omega C}\right)^{1/2}\right] \tag{50}$$

The low-loss approximation then allows us to use the first three terms in the binomial series:

$$\sqrt{1 + x} \doteq 1 + \frac{x}{2} - \frac{x^2}{8} \quad (x \ll 1) \tag{51}$$

We use (51) to expand the terms in large parentheses in (50), obtaining:

$$\gamma \doteq j\omega\sqrt{LC}\left[\left(1 + \frac{R}{j2\omega L} + \frac{R^2}{8\omega^2 L^2}\right)\left(1 + \frac{G}{j2\omega C} + \frac{G^2}{8\omega^2 C^2}\right)\right] \tag{52}$$

All products in (52) are then carried out, neglecting the terms involving RG^2, R^2G, and R^2G^2, as these will be negligible compared to all others. The result is

$$\gamma = \alpha + j\beta \doteq j\omega\sqrt{LC}\left[1 + \frac{1}{j2\omega}\left(\frac{R}{L} + \frac{G}{C}\right) + \frac{1}{8\omega^2}\left(\frac{R^2}{L^2} - \frac{2RG}{LC} + \frac{G^2}{C^2}\right)\right] \tag{53}$$

Now, separating real and imaginary parts of (53) yields α and β:

$$\alpha \doteq \frac{1}{2}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) \tag{54a}$$

and

$$\beta \doteq \omega\sqrt{LC}\left[1 + \frac{1}{8}\left(\frac{G}{\omega C} - \frac{R}{\omega L}\right)^2\right] \tag{54b}$$

We note that α scales in direct proportion to R and G, as would be expected. We also note that the terms in (54b) that involve R and G lead to a phase velocity, $v_p = \omega/\beta$, that is frequency-dependent. Moreover, the *group velocity, $v_g = d\omega/d\beta$,* will also depend on frequency, and will lead to signal distortion, as we will explore in Chapter 12. Note that with nonzero R and G, phase and group velocities that are constant with frequency can be obtained when $R/L = G/C$, known as *Heaviside's condition.* In this case, (54b) becomes $\beta \doteq \omega\sqrt{LC}$, and the line is said to be *distortionless.* Further complications occur when accounting for possible frequency dependencies within R, G, L, and C. Consequently, conditions of low-loss or distortion-free propagation will usually occur over limited frequency ranges. As a rule, loss increases with increasing frequency, mostly because of the increase in R with frequency. The nature of this latter effect, known as *skin effect* loss, requires field theory to understand

and quantify. We will study this in Chapter 11, and we will apply it to transmission line structures in Chapter 13.

Finally, we can apply the low-loss approximation to the characteristic impedance, Eq. (47). Using (51), we find

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} = \sqrt{\frac{j\omega L \left(1 + \frac{R}{j\omega L}\right)}{j\omega C \left(1 + \frac{G}{j\omega C}\right)}} \doteq \sqrt{\frac{L}{C} \left[\frac{\left(1 + \frac{R}{j2\omega L} + \frac{R^2}{8\omega^2 L^2}\right)}{\left(1 + \frac{G}{j2\omega C} + \frac{G^2}{8\omega^2 C^2}\right)}\right]} \quad (55)$$

Next, we multiply (55) by a factor of 1, in the form of the complex conjugate of the denominator of (55) divided by itself. The resulting expression is simplified by neglecting all terms on the order of $R^2 G$, $G^2 R$, and higher. Additionally, the approximation, $1/(1 + x) \doteq 1 - x$, where $x \ll 1$ is used. The result is

$$Z_0 \doteq \sqrt{\frac{L}{C}} \left\{1 + \frac{1}{2\omega^2}\left[\frac{1}{4}\left(\frac{R}{L} + \frac{G}{C}\right)^2 - \frac{G^2}{C^2}\right] + \frac{j}{2\omega}\left(\frac{G}{C} - \frac{R}{L}\right)\right\} \quad (56)$$

Note that when Heaviside's condition (again, $R/L = G/C$) holds, Z_0 simplifies to just $\sqrt{L/C}$, as is true when both R and G are zero.

EXAMPLE 10.3

Suppose in a certain transmission line $G = 0$, but R is finite valued and satisfies the low-loss requirement, $R \ll \omega L$. Use Eq. (56) to write the approximate magnitude and phase of Z_0.

Solution. With $G = 0$, the imaginary part of (56) is much greater than the second term in the real part [proportional to $(R/\omega L)^2$]. Therefore, the characteristic impedance becomes

$$Z_0(G = 0) \doteq \sqrt{\frac{L}{C}} \left(1 - j\frac{R}{2\omega L}\right) = |Z_0|e^{j\theta}$$

where $|Z_0| \doteq \sqrt{L/C}$, and $\theta = \tan^{-1}(-R/2\omega L)$.

D10.1. At an operating radian frequency of 500 Mrad/s, typical circuit values for a certain transmission line are: $R = 0.2$ Ω/m, $L = 0.25$ μH/m, $G = 10$ μS/m, and $C = 100$ pF/m. Find: (a) α; (b) β; (c) λ; (d) v_p; (e) Z_0.

Ans. 2.25 mNp/m; 2.50 rad/m; 2.51 m; 2×10^8 m/sec; $50.0 - j0.0350$ Ω

10.8 POWER TRANSMISSION AND THE USE OF DECIBELS IN LOSS CHARACTERIZATION

Having found the sinusoidal voltage and current in a lossy transmission line, we next evaluate the power transmitted over a specified distance as a function of voltage and current amplitudes. We start with the *instantaneous* power, given simply as the product of the real voltage and current. Consider the forward-propagating term in (49), where

again, the amplitude, $V_0^+ = |V_0|$, is taken to be real. The current waveform will be similar, but will generally be shifted in phase. Both current and voltage attenuate according to the factor $e^{-\alpha z}$. The instantaneous power therefore becomes:

$$\mathcal{P}(z, t) = \mathcal{V}(z, t)\mathcal{I}(z, t) = |V_0||I_0|e^{-2\alpha z} \cos(\omega t - \beta z) \cos(\omega t - \beta z + \theta) \quad (57)$$

Usually, the *time-averaged* power, $\langle P \rangle$, is of interest. We find this through:

$$\langle P \rangle = \frac{1}{T} \int_0^T |V_0||I_0|e^{-2\alpha z} \cos(\omega t - \beta z) \cos(\omega t - \beta z + \theta) dt \quad (58)$$

where $T = 2\pi/\omega$ is the time period for one oscillation cycle. Using a trigonometric identity, the product of cosines in the integrand can be written as the sum of individual cosines at the sum and difference frequencies:

$$\langle P \rangle = \frac{1}{T} \int_0^T \frac{1}{2}|V_0||I_0|e^{-2\alpha z}[\cos(2\omega t - 2\beta z + \theta) + \cos(\theta)]\, dt \quad (59)$$

The first cosine term integrates to zero, leaving the $\cos \theta$ term. The remaining integral easily evaluates as

$$\langle P \rangle = \frac{1}{2}|V_0||I_0|e^{-2\alpha z} \cos \theta = \frac{1}{2}\frac{|V_0|^2}{|Z_0|}e^{-2\alpha z} \cos \theta \,\, [\text{W}] \quad (60)$$

The same result can be obtained directly from the phasor voltage and current. We begin with these, expressed as

$$V_s(z) = V_0 e^{-\alpha z} e^{-j\beta z} \quad (61)$$

and

$$I_s(z) = I_0 e^{-\alpha z} e^{-j\beta z} = \frac{V_0}{Z_0} e^{-\alpha z} e^{-j\beta z} \quad (62)$$

where $Z_0 = |Z_0|e^{j\theta}$. We now note that the time-averaged power as expressed in (60) can be obtained from the phasor forms through:

$$\boxed{\langle P \rangle = \frac{1}{2}\text{Re}\{V_s I_s^*\}} \quad (63)$$

where again, the asterisk (*) denotes the complex conjugate (applied in this case to the current phasor only). Using (61) and (62) in (63), it is found that

$$\langle P \rangle = \frac{1}{2}\text{Re}\left\{ V_0 e^{-\alpha z} e^{-j\beta z} \frac{V_0^*}{|Z_0|e^{-j\theta}} e^{-\alpha z} e^{+j\beta z} \right\}$$

$$= \frac{1}{2}\text{Re}\left\{ \frac{V_0 V_0^*}{|Z_0|} e^{-2\alpha z} e^{j\theta} \right\} = \frac{1}{2}\frac{|V_0|^2}{|Z_0|} e^{-2\alpha z} \cos \theta \quad (64)$$

which we note is identical to the time-integrated result in (60). Eq. (63) applies to any single-frequency wave.

An important result of the preceding exercise is that power attenuates as $e^{-2\alpha z}$, or

$$\langle \mathcal{P}(z) \rangle = \langle \mathcal{P}(0) \rangle e^{-2\alpha z} \qquad (65)$$

Power drops at twice the exponential rate with distance as either voltage or current.

A convenient measure of power loss is in *decibel* units. This is based on expressing the power decrease as a power of 10. Specifically, we write

$$\frac{\langle \mathcal{P}(z) \rangle}{\langle \mathcal{P}(0) \rangle} = e^{-2\alpha z} = 10^{-\kappa \alpha z} \qquad (66)$$

where the constant, κ, is to be determined. Setting $\alpha z = 1$, we find

$$e^{-2} = 10^{-\kappa} \quad \Rightarrow \quad \kappa = \log_{10}(e^2) = 0.869 \qquad (67)$$

Now, by definition, the power loss in decibels (dB) is

$$\text{Power loss (dB)} = 10 \log_{10} \left[\frac{\langle \mathcal{P}(0) \rangle}{\langle \mathcal{P}(z) \rangle} \right] = 8.69 \alpha z \qquad (68)$$

where we note that inverting the power ratio in the argument of the log function [as compared to the ratio in (66)] yields a positive number for the dB loss. Also, noting that $\langle \mathcal{P} \rangle \propto |V_0|^2$, we may write, equivalently:

$$\text{Power loss (dB)} = 10 \log_{10} \left[\frac{\langle \mathcal{P}(0) \rangle}{\langle \mathcal{P}(z) \rangle} \right] = 20 \log_{10} \left[\frac{|V_0(0)|}{|V_0(z)|} \right] \qquad (69)$$

where $|V_0(z)| = |V_0(0)| e^{-\alpha z}$.

EXAMPLE 10.4

A 20-m length of transmission line is known to produce a 2.0-dB drop in power from end to end. (*a*) What fraction of the input power reaches the output? (*b*) What fraction of the input power reaches the midpoint of the line? (*c*) What exponential attenuation coefficient, α, does this represent?

Solution. (*a*) The power fraction will be

$$\frac{\langle \mathcal{P}(20) \rangle}{\langle \mathcal{P}(0) \rangle} = 10^{-0.2} = 0.63$$

(*b*) 2 dB in 20 m implies a loss rating of 0.2 dB/m. So, over a 10-m span, the loss is 1.0 dB. This represents the power fraction, $10^{-0.1} = 0.79$.

(*c*) The exponential attenuation coefficient is found through

$$\alpha = \frac{2.0 \text{ dB}}{(8.69 \text{ dB/Np})(20 \text{ m})} = 0.012 \text{ [Np/m]}$$

A final point addresses the question: Why use decibels? The most compelling reason is that when evaluating the accumulated loss for several lines and devices that

are all end-to-end connected, the net loss in dB for the entire span is just the sum of the dB losses of the individual elements.

> **D10.2.** Two transmission lines are to be joined end to end. Line 1 is 30 m long and is rated at 0.1 dB/m. Line 2 is 45 m long and is rated at 0.15 dB/m. The joint is not done well and imparts a 3-dB loss. What percentage of the input power reaches the output of the combination?
>
> **Ans.** 5.3%

10.9 WAVE REFLECTION AT DISCONTINUITIES

The concept of wave reflection was introduced in Section 10.1. As implied there, the need for a reflected wave originates from the necessity to satisfy all voltage and current boundary conditions at the ends of transmission lines and at locations at which two dissimilar lines are connected to each other. The consequences of reflected waves are usually less than desirable, in that some of the power that was intended to be transmitted to a load, for example, reflects and propagates back to the source. Conditions for achieving *no* reflected waves are therefore important to understand.

The basic reflection problem is illustrated in Figure 10.5. In it, a transmission line of characteristic impedance Z_0 is terminated by a load having complex impedance, $Z_L = R_L + jX_L$. If the line is lossy, then we know that Z_0 will also be complex. For convenience, we assign coordinates such that the load is at location $z = 0$. Therefore, the line occupies the region $z < 0$. A voltage wave is presumed to be incident on the load, and is expressed in phasor form for all z:

$$V_i(z) = V_{0i}e^{-\alpha z}e^{-j\beta z} \tag{70a}$$

When the wave reaches the load, a reflected wave is generated that back-propagates:

$$V_r(z) = V_{0r}e^{+\alpha z}e^{+j\beta z} \tag{70b}$$
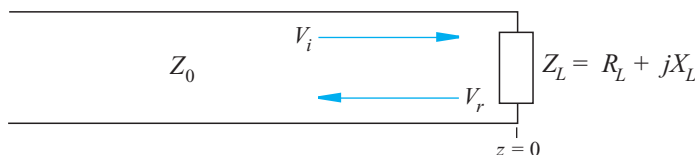
Figure 10.5 Voltage wave reflection from a complex load impedance.

The phasor voltage at the load is now the sum of the incident and reflected voltage phasors, evaluated at $z = 0$:

$$V_L = V_{0i} + V_{0r} \tag{71}$$

Additionally, the current through the load is the sum of the incident and reflected currents, also at $z = 0$:

$$I_L = I_{0i} + I_{0r} = \frac{1}{Z_0}[V_{0i} - V_{0r}] = \frac{V_L}{Z_L} = \frac{1}{Z_L}[V_{0i} + V_{0r}] \tag{72}$$

We can now solve for the ratio of the reflected voltage amplitude to the incident voltage amplitude, defined as the *reflection coefficient,* Γ:

$$\Gamma \equiv \frac{V_{0r}}{V_{0i}} = \frac{Z_L - Z_0}{Z_L + Z_0} = |\Gamma|e^{j\phi_r} \tag{73}$$

where we emphasize the complex nature of Γ—meaning that, in general, a reflected wave will experience a reduction in amplitude and a phase shift, relative to the incident wave.

Now, using (71) with (73), we may write

$$V_L = V_{0i} + \Gamma V_{0i} \tag{74}$$

from which we find the *transmission coefficient,* defined as the ratio of the load voltage amplitude to the incident voltage amplitude:

$$\tau \equiv \frac{V_L}{V_{0i}} = 1 + \Gamma = \frac{2Z_L}{Z_0 + Z_L} = |\tau|e^{j\phi_t} \tag{75}$$

A point that may at first cause some alarm is that if Γ is a positive real number, then $\tau > 1$; the voltage amplitude at the load is thus greater than the incident voltage. Although this would seem counterintuitive, it is not a problem because the load current will be lower than that in the incident wave. We will find that this always results in an average *power* at the load that is less than or equal to that in the incident wave. An additional point concerns the possibility of loss in the line. The incident wave amplitude that is used in (73) and (75) is always the amplitude that occurs *at the load*—after loss has occurred in propagating from the input.

Usually, the main objective in transmitting power to a load is to configure the line/load combination such that there is no reflection. The load therefore receives all the transmitted power. The condition for this is $\Gamma = 0$, which means that the load impedance must be equal to the line impedance. In such cases the load is said to be *matched* to the line (or vice versa). Various impedance-matching methods exist, many of which will be explored later in this chapter.

Finally, the fractions of the incident wave *power* that are reflected and dissipated by the load need to be determined. The incident power is found from (64), where this time we position the load at $z = L$, with the line input at $z = 0$.

$$\langle \mathcal{P}_i \rangle = \frac{1}{2}\text{Re}\left\{\frac{V_0 V_0^*}{|Z_0|}e^{-2\alpha L}e^{j\theta}\right\} = \frac{1}{2}\frac{|V_0|^2}{|Z_0|}e^{-2\alpha L}\cos\theta \tag{76a}$$

The reflected power is then found by substituting the reflected wave voltage into (76a), where the latter is obtained by multiplying the incident voltage by Γ:

$$\langle \mathcal{P}_r \rangle = \frac{1}{2}\text{Re}\left\{ \frac{(\Gamma V_0)(\Gamma^* V_0^*)}{|Z_0|} e^{-2\alpha L} e^{j\theta} \right\} = \frac{1}{2} \frac{|\Gamma|^2 |V_0|^2}{|Z_0|} e^{-2\alpha L} \cos\theta \qquad (76b)$$

The reflected power fraction at the load is now determined by the ratio of (76b) to (76a):

$$\boxed{\frac{\langle \mathcal{P}_r \rangle}{\langle \mathcal{P}_i \rangle} = \Gamma\Gamma^* = |\Gamma|^2} \qquad (77a)$$

The fraction of the incident power that is transmitted into the load (or dissipated by it) is therefore

$$\boxed{\frac{\langle \mathcal{P}_t \rangle}{\langle \mathcal{P}_i \rangle} = 1 - |\Gamma|^2} \qquad (77b)$$

The reader should be aware that the transmitted power fraction is *not* $|\tau|^2$, as one might be tempted to conclude.

In situations involving the connection of two semi-infinite transmission lines having different characteristic impedances, reflections will occur at the junction, with the second line being treated as the load. For a wave incident from line 1 (Z_{01}) to line 2 (Z_{02}), we find

$$\boxed{\Gamma = \frac{Z_{02} - Z_{01}}{Z_{02} + Z_{01}}} \qquad (78)$$

The fraction of the power that propagates into the second line is then $1 - |\Gamma|^2$.

EXAMPLE 10.5

A 50-Ω lossless transmission line is terminated by a load impedance, $Z_L = 50 - j75\ \Omega$. If the incident power is 100 mW, find the power dissipated by the load.

Solution. The reflection coefficient is

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{50 - j75 - 50}{50 - j75 + 50} = 0.36 - j0.48 = 0.60 e^{-j.93}$$

Then

$$\langle \mathcal{P}_t \rangle = (1 - |\Gamma|^2)\langle \mathcal{P}_i \rangle = [1 - (0.60)^2](100) = 64\ \text{mW}$$

EXAMPLE 10.6

Two lossy lines are to be joined end to end. The first line is 10 m long and has a loss rating of 0.20 dB/m. The second line is 15 m long and has a loss rating of 0.10 dB/m. The reflection coefficient at the junction (line 1 to line 2) is $\Gamma = 0.30$. The input

power (to line 1) is 100 mW. (*a*) Determine the total loss of the combination in dB. (*b*) Determine the power transmitted to the output end of line 2.

Solution. (*a*) The dB loss of the joint is

$$L_j(\text{dB}) = 10 \log_{10}\left(\frac{1}{1 - |\Gamma|^2}\right) = 10 \log_{10}\left(\frac{1}{1 - 0.09}\right) = 0.41 \text{ dB}$$

The total loss of the link in dB is now

$$L_t(\text{dB}) = (0.20)(10) + 0.41 + (0.10)(15) = 3.91 \text{ dB}$$

(*b*) The output power will be $P_{\text{out}} = 100 \times 10^{-0.391} = 41$ mW.

10.10 VOLTAGE STANDING WAVE RATIO

In many instances, characteristics of transmission line performance are amenable to measurement. Included in these are measurements of unknown load impedances, or input impedances of lines that are terminated by known or unknown load impedances. Such techniques rely on the ability to measure voltage amplitudes that occur as functions of position within a line, usually designed for this purpose. A typical apparatus consists of a *slotted line*, which is a lossless coaxial transmission line having a longitudinal gap in the outer conductor along its entire length. The line is positioned between the sinusoidal voltage source and the impedance that is to be measured. Through the gap in the slotted line, a voltage probe may be inserted to measure the voltage amplitude between the inner and outer conductors. As the probe is moved along the length of the line, the maximum and minimum voltage amplitudes are noted, and their ratio, known as the *voltage standing wave ratio*, or VSWR, is determined. The significance of this measurement and its utility form the subject of this section.

   To understand the meaning of the voltage measurements, we consider a few special cases. First, if the slotted line is terminated by a matched impedance, then no reflected wave occurs; the probe will indicate the same voltage amplitude at every point. Of course, the instantaneous voltages that the probe samples will differ in phase by $\beta(z_2 - z_1)$ rad as the probe is moved from $z = z_1$ to $z = z_2$, but the system is insensitive to the phase of the field. The equal-amplitude voltages are characteristic of an unattenuated traveling wave.

   Second, if the slotted line is terminated by an open or short circuit (or in general a purely imaginary load impedance), the total voltage in the line is a standing wave and, as was shown in Example 10.1, the voltage probe provides no output when it is located at the nodes; these occur periodically with half-wavelength spacing. As the probe position is changed, its output varies as $|\cos(\beta z + \phi)|$, where z is the distance from the load, and where the phase, ϕ, depends on the load impedance. For example,

if the load is a short circuit, the requirement of zero voltage at the short leads to a null occurring there, and so the voltage in the line will vary as $|\sin(\beta z)|$ (where $\phi = \pm\pi/2$).

A more complicated situation arises when the reflected voltage is neither 0 nor 100 percent of the incident voltage. Some energy is absorbed by the load and some is reflected. The slotted line, therefore, supports a voltage that is composed of both a traveling wave and a standing wave. It is customary to describe this voltage as a standing wave, even though a traveling wave is also present. We will see that the voltage does not have zero amplitude at any point for all time, and the degree to which the voltage is divided between a traveling wave and a true standing wave is expressed by the ratio of the maximum amplitude found by the probe to the minimum amplitude (VSWR). This information, along with the positions of the voltage minima or maxima with respect to that of the load, enable one to determine the load impedance. The VSWR also provides a measure of the quality of the termination. Specifically, a perfectly matched load yields a VSWR of exactly 1. A totally reflecting load produces an infinite VSWR.

To derive the specific form of the total voltage, we begin with the forward and backward-propagating waves that occur within the slotted line. The load is positioned at $z = 0$, and so all positions within the slotted line occur at negative values of z. Taking the input wave amplitude as V_0, the total phasor voltage is

$$V_{sT}(z) = V_0 e^{-j\beta z} + \Gamma V_0 e^{j\beta z} \tag{79}$$

The line, being lossless, has real characteristic impedance, Z_0. The load impedance, Z_L, is in general complex, which leads to a complex reflection coefficient:

$$\boxed{\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} = |\Gamma|e^{j\phi}} \tag{80}$$

If the load is a short circuit ($Z_L = 0$), ϕ is equal to π; if Z_L is real and less than Z_0, ϕ is also equal to π; and if Z_L is real and greater than Z_0, ϕ is zero. Using (80), we may rewrite (79) in the form:

$$V_{sT}(z) = V_0 \left(e^{-j\beta z} + |\Gamma|e^{j(\beta z + \phi)}\right) = V_0 e^{j\phi/2}\left(e^{-j\beta z}e^{-j\phi/2} + |\Gamma|e^{j\beta z}e^{j\phi/2}\right) \tag{81}$$

To express (81) in a more useful form, we can apply the algebraic trick of adding and subtracting the term $V_0(1 - |\Gamma|)e^{-j\beta z}$:

$$V_{sT}(z) = V_0(1 - |\Gamma|)e^{-j\beta z} + V_0|\Gamma|e^{j\phi/2}\left(e^{-j\beta z}e^{-j\phi/2} + e^{j\beta z}e^{j\phi/2}\right) \tag{82}$$

The last term in parentheses in (82) becomes a cosine, and we write

$$\boxed{V_{sT}(z) = V_0(1 - |\Gamma|)e^{-j\beta z} + 2V_0|\Gamma|e^{j\phi/2}\cos(\beta z + \phi/2)} \tag{83}$$

The important characteristics of this result are most easily seen by converting it to real instantaneous form:

$$\mathcal{V}(z, t) = \text{Re}[V_{sT}(z)e^{j\omega t}] = \underbrace{V_0(1 - |\Gamma|)\cos(\omega t - \beta z)}_{\text{traveling wave}}$$
$$+ \underbrace{2|\Gamma|V_0\cos(\beta z + \phi/2)\cos(\omega t + \phi/2)}_{\text{standing wave}} \qquad (84)$$

Equation (84) is recognized as the sum of a traveling wave of amplitude $(1 - |\Gamma|)V_0$ and a standing wave having amplitude $2|\Gamma|V_0$. We can visualize events as follows: The portion of the incident wave that reflects and back-propagates in the slotted line interferes with an equivalent portion of the incident wave to form a standing wave. The rest of the incident wave (which does not interfere) is the traveling wave part of (84). The maximum amplitude observed in the line is found where the amplitudes of the two terms in (84) add directly to give $(1 + |\Gamma|)V_0$. The minimum amplitude is found where the standing wave achieves a null, leaving only the traveling wave amplitude of $(1 - |\Gamma|)V_0$. The fact that the two terms in (84) combine in this way with the proper phasing is not immediately apparent, but the following arguments will show that this does occur.

To obtain the minimum and maximum voltage amplitudes, we may revisit the first part of Eq. (81):

$$V_{sT}(z) = V_0\left(e^{-j\beta z} + |\Gamma|e^{j(\beta z + \phi)}\right) \qquad (85)$$

First, the minimum voltage amplitude is obtained when the two terms in (85) subtract directly (having a phase difference of π). This occurs at locations

$$z_{\text{min}} = -\frac{1}{2\beta}(\phi + (2m + 1)\pi) \quad (m = 0, 1, 2, \ldots) \qquad (86)$$

Note again that all positions within the slotted line occur at negative values of z. Substituting (86) into (85) leads to the minimum amplitude:

$$V_{sT}(z_{\text{min}}) = V_0(1 - |\Gamma|) \qquad (87)$$

The same result is obtained by substituting (86) into the real voltage, (84). This produces a null in the standing wave part, and we obtain

$$\mathcal{V}(z_{\text{min}}, t) = \pm V_0(1 - |\Gamma|)\sin(\omega t + \phi/2) \qquad (88)$$

The voltage oscillates (through zero) in time, with amplitude $V_0(1 - |\Gamma|)$. The plus and minus signs in (88) apply to even and odd values of m in (86), respectively.

Next, the maximum voltage amplitude is obtained when the two terms in (85) add in-phase. This will occur at locations given by

$$z_{\text{max}} = -\frac{1}{2\beta}(\phi + 2m\pi) \quad (m = 0, 1, 2, \ldots) \qquad (89)$$

On substituting (89) into (85), we obtain

$$V_{sT}(z_{\max}) = V_0(1 + |\Gamma|) \tag{90}$$

As before, we may substitute (89) into the real instantaneous voltage (84). The effect is to produce a maximum in the standing wave part, which then adds in-phase to the running wave. The result is

$$\mathcal{V}(z_{\max}, t) = \pm V_0(1 + |\Gamma|) \cos(\omega t + \phi/2) \tag{91}$$

where the plus and minus signs apply to positive and negative values of m in (89), respectively. Again, the voltage oscillates through zero in time, with amplitude $V_0(1 + |\Gamma|)$.

Note that a voltage maximum is located at the load ($z = 0$) if $\phi = 0$; moreover, $\phi = 0$ when Γ is real and positive. This occurs for real Z_L when $Z_L > Z_0$. Thus there is a voltage maximum at the load when the load impedance is greater than Z_0 and both impedances are real. With $\phi = 0$, maxima also occur at $z_{\max} = -m\pi/\beta = -m\lambda/2$. For a zero-load impedance, $\phi = \pi$, and the maxima are found at $z_{\max} = -\pi/(2\beta), -3\pi/(2\beta)$, or $z_{\max} = -\lambda/4, -3\lambda/4$, and so forth.

The minima are separated by multiples of one half-wavelength (as are the maxima), and for a zero load impedance, the first minimum occurs when $-\beta z = 0$, or at the load. In general, a voltage minimum is found at $z = 0$ whenever $\phi = \pi$; this occurs if $Z_L < Z_0$ where Z_L is real. The general results are illustrated in Figure 10.6.
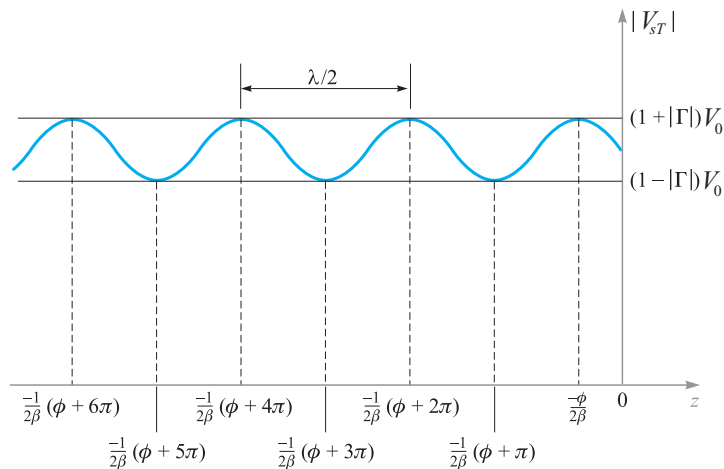
Figure 10.6 Plot of the magnitude of V_{sT} as found from Eq. (85) as a function of position, z, in front of the load (at $z = 0$). The reflection coefficient phase is ϕ, which leads to the indicated locations of maximum and minimum voltage amplitude, as found from Eqs. (86) and (89).

Finally, the voltage standing wave ratio is defined as:

$$s \equiv \frac{V_{sT}(z_{\max})}{V_{sT}(z_{\min})} = \frac{1 + |\Gamma|}{1 - |\Gamma|} \qquad (92)$$

Since the absolute voltage amplitudes have divided out, our measured VSWR permits the immediate evaluation of $|\Gamma|$. The phase of Γ is then found by measuring the location of the first maximum or minimum with respect to the load, and then using (86) or (89) as appropriate. Once Γ is known, the load impedance can be found, assuming Z_0 is known.

> **D10.3.** What voltage standing wave ratio results when $\Gamma = \pm 1/2$?
>
> **Ans.** 3

Slotted line measurements yield a VSWR of 5, a 15-cm spacing between successive voltage maxima, and the first maximum at a distance of 7.5 cm in front of the load. Determine the load impedance, assuming a 50-Ω impedance for the slotted line.

Solution. The 15-cm spacing between maxima is $\lambda/2$, implying a wavelength of 30 cm. Because the slotted line is air-filled, the frequency is $f = c/\lambda = 1$ GHz. The first maximum at 7.5 cm is thus at a distance of $\lambda/4$ from the load, which means that a voltage minimum occurs at the load. Thus Γ will be real and negative. We use (92) to write

$$|\Gamma| = \frac{s - 1}{s + 1} = \frac{5 - 1}{5 + 1} = \frac{2}{3}$$

So

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

which we solve for Z_L to obtain

$$Z_L = \frac{1}{5} Z_0 = \frac{50}{5} = 10 \ \Omega$$

10.11 TRANSMISSION LINES OF FINITE LENGTH

A new type of problem emerges when considering the propagation of sinusoidal voltages on finite-length lines that have loads that are not impedance matched. In such cases, numerous reflections occur at the load and at the generator, setting up a multiwave bidirectional voltage distribution in the line. As always, the objective is to determine the net power transferred to the load in steady state, but we must now include the effect of the numerous forward- and backward-reflected waves.
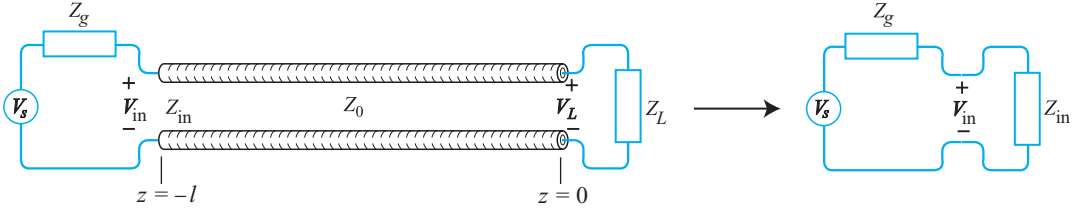
Figure 10.7 Finite-length transmission line configuration and its equivalent circuit.

Figure 10.7 shows the basic problem. The line, assumed to be lossless, has characteristic impedance Z_0 and is of length l. The sinusoidal voltage source at frequency ω provides phasor voltage V_s. Associated with the souce is a complex internal impedance, Z_g, as shown. The load impedance, Z_L, is also assumed to be complex and is located at $z = 0$. The line thus exists along the negative z axis. The easiest method of approaching the problem is not to attempt to analyze every reflection individually, but rather to recognize that in steady state, there will exist one net forward wave and one net backward wave, representing the superposition of all waves that are incident on the load and all waves that are reflected from it. We may thus write the total voltage in the line as

$$V_{sT}(z) = V_0^+ e^{-j\beta z} + V_0^- e^{j\beta z} \tag{93}$$

in which V_0^+ and V_0^- are complex amplitudes, composed respectively of the sum of all individual forward and backward wave amplitudes and phases. In a similar way, we may write the total current in the line:

$$I_{sT}(z) = I_0^+ e^{-j\beta z} + I_0^- e^{j\beta z} \tag{94}$$

We now define the *wave impedance, $Z_w(z)$,* as the ratio of the total phasor voltage to the total phasor current. Using (93) and (94), this becomes:

$$Z_w(z) \equiv \frac{V_{sT}(z)}{I_{sT}(z)} = \frac{V_0^+ e^{-j\beta z} + V_0^- e^{j\beta z}}{I_0^+ e^{-j\beta z} + I_0^- e^{j\beta z}} \tag{95}$$

We next use the relations $V_0^- = \Gamma V_0^+$, $I_0^+ = V_0^+/Z_0$, and $I_0^- = -V_0^-/Z_0$. Eq. (95) simplifies to

$$Z_w(z) = Z_0 \left[\frac{e^{-j\beta z} + \Gamma e^{j\beta z}}{e^{-j\beta z} - \Gamma e^{j\beta z}} \right] \tag{96}$$

Now, using the Euler identity, (32), and substituting $\Gamma = (Z_L - Z_0)/(Z_L + Z_0)$, Eq. (96) becomes

$$Z_w(z) = Z_0 \left[\frac{Z_L \cos(\beta z) - j Z_0 \sin(\beta z)}{Z_0 \cos(\beta z) - j Z_L \sin(\beta z)} \right] \tag{97}$$

The wave impedance at the line input is now found by evaluating (97) at $z = -l$, obtaining

$$Z_{\text{in}} = Z_0 \left[\frac{Z_L \cos(\beta l) + j Z_0 \sin(\beta l)}{Z_0 \cos(\beta l) + j Z_L \sin(\beta l)} \right] \tag{98}$$

This is the quantity that we need in order to create the equivalent circuit in Figure 10.7.

One special case is that in which the line length is a half-wavelength, or an integer multiple thereof. In that case,

$$\beta l = \frac{2\pi}{\lambda} \frac{m\lambda}{2} = m\pi \quad (m = 0, 1, 2, \ldots)$$

Using this result in (98), we find

$$Z_{\text{in}}(l = m\lambda/2) = Z_L \tag{99}$$

For a half-wave line, the equivalent circuit can be constructed simply by removing the line completely and placing the load impedance at the input. This simplification works, of course, provided the line length is indeed an integer multiple of a half-wavelength. Once the frequency begins to vary, the condition is no longer satisfied, and (98) must be used in its general form to find Z_{in}.

Another important special case is that in which the line length is an odd multiple of a quarter wavelength:

$$\beta l = \frac{2\pi}{\lambda}(2m + 1)\frac{\lambda}{4} = (2m + 1)\frac{\pi}{2} \quad (m = 0, 1, 2, \ldots)$$

Using this result in (98) leads to

$$Z_{\text{in}}(l = \lambda/4) = \frac{Z_0^2}{Z_L} \tag{100}$$

An immediate application of (100) is to the problem of joining two lines having different characteristic impedances. Suppose the impedances are (from left to right) Z_{01} and Z_{03}. At the joint, we may insert an additional line whose characteristic impedance is Z_{02} and whose length is $\lambda/4$. We thus have a sequence of joined lines whose impedances progress as Z_{01}, Z_{02}, and Z_{03}, in that order. A voltage wave is now incident from line 1 onto the joint between Z_{01} and Z_{02}. Now the effective load at the far end of line 2 is Z_{03}. The input impedance to line 2 at any frequency is now

$$Z_{\text{in}} = Z_{02} \frac{Z_{03} \cos \beta_2 l + j Z_{02} \sin \beta_2 l}{Z_{02} \cos \beta_2 l + j Z_{03} \sin \beta_2 l} \tag{101}$$

Then, since the length of line 2 is $\lambda/4$,

$$Z_{\text{in}}(\text{line 2}) = \frac{Z_{02}^2}{Z_{03}} \tag{102}$$

Reflections at the Z_{01}–Z_{02} interface will not occur if $Z_{\text{in}} = Z_{01}$. Therefore, we can match the junction (allowing complete transmission through the three-line sequence)

if Z_{02} is chosen so that

$$Z_{02} = \sqrt{Z_{01} Z_{03}} \qquad (103)$$

This technique is called *quarter-wave matching* and again is limited to the frequency (or narrow band of frequencies) such that $l \doteq (2m + 1)\lambda/4$. We will encounter more examples of these techniques when we explore electromagnetic wave reflection in Chapter 12. Meanwhile, further examples that involve the use of the input impedance and the VSWR are presented in Section 10.12.

10.12 SOME TRANSMISSION LINE EXAMPLES

In this section, we apply many of the results that we obtained in the previous sections to several typical transmission line problems. We simplify our work by restricting our attention to the lossless line.

Let us begin by assuming a two-wire 300 Ω line ($Z_0 = 300\ \Omega$), such as the lead-in wire from the antenna to a television or FM receiver. The circuit is shown in Figure 10.8. The line is 2 m long, and the values of L and C are such that the velocity on the line is 2.5×10^8 m/s. We will terminate the line with a receiver having an input resistance of 300 Ω and represent the antenna by its Thevenin equivalent $Z = 300\ \Omega$ in series with $V_s = 60$ V at 100 MHz. This antenna voltage is larger by a factor of about 10^5 than it would be in a practical case, but it also provides simpler values to work with; in order to think practical thoughts, divide currents or voltages by 10^5, divide powers by 10^{10}, and leave impedances alone.

Because the load impedance is equal to the characteristic impedance, the line is matched; the reflection coefficient is zero, and the standing wave ratio is unity. For the given velocity and frequency, the wavelength on the line is $v/f = 2.5$ m, and the phase constant is $2\pi/\lambda = 0.8\pi$ rad/m; the attenuation constant is zero. The electrical length of the line is $\beta l = (0.8\pi)2$, or 1.6π rad. This length may also be expressed as $288°$, or 0.8 wavelength.

The input impedance offered to the voltage source is 300 Ω, and since the internal impedance of the source is 300 Ω, the voltage at the input to the line is half of 60 V, or 30 V. The source is matched to the line and delivers the maximum available power
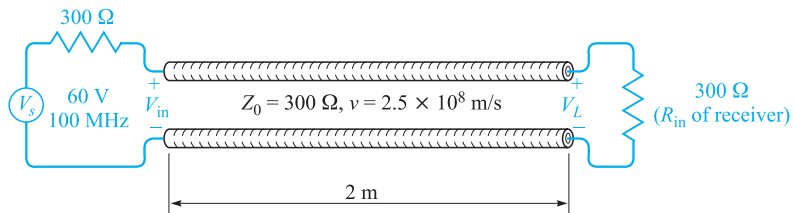
Figure 10.8   A transmission line that is matched at both ends produces no reflections and thus delivers maximum power to the load.

to the line. Because there is no reflection and no attenuation, the voltage at the load is 30 V, but it is delayed in phase by 1.6π rad. Thus,

$$V_{\text{in}} = 30\cos(2\pi\,10^8 t)\text{ V}$$

whereas

$$V_L = 30\cos(2\pi\,10^8 t - 1.6\pi)\text{ V}$$

The input current is

$$I_{\text{in}} = \frac{V_{\text{in}}}{300} = 0.1\cos(2\pi\,10^8 t)\text{ A}$$

while the load current is

$$I_L = 0.1\cos(2\pi\,10^8 t - 1.6\pi)\text{ A}$$

The average power delivered to the input of the line by the source must all be delivered to the load by the line,

$$P_{\text{in}} = P_L = \frac{1}{2}\times 30\times 0.1 = 1.5\text{ W}$$

Now let us connect a second receiver, also having an input resistance of 300 Ω, across the line in parallel with the first receiver. The load impedance is now 150 Ω, the reflection coefficient is

$$\Gamma = \frac{150 - 300}{150 + 300} = -\frac{1}{3}$$

and the standing wave ratio on the line is

$$s = \frac{1 + \frac{1}{3}}{1 - \frac{1}{3}} = 2$$

The input impedance is no longer 300 Ω, but is now

$$Z_{\text{in}} = Z_0\frac{Z_L\cos\beta l + jZ_0\sin\beta l}{Z_0\cos\beta l + jZ_L\sin\beta l} = 300\frac{150\cos 288° + j300\sin 288°}{300\cos 288° + j150\sin 288°}$$

$$= 510\angle{-23.8°} = 466 - j206\ \Omega$$

which is a capacitive impedance. Physically, this means that this length of line stores more energy in its electric field than in its magnetic field. The input current phasor is thus

$$I_{s,\text{in}} = \frac{60}{300 + 466 - j206} = 0.0756\angle 15.0°\text{ A}$$

and the power supplied to the line by the source is

$$P_{\text{in}} = \frac{1}{2}\times (0.0756)^2 \times 466 = 1.333\text{ W}$$

Since there are no losses in the line, 1.333 W must also be delivered to the load. Note that this is less than the 1.50 W which we were able to deliver to a matched load; moreover, this power must divide equally between two receivers, and thus each

receiver now receives only 0.667 W. Because the input impedance of each receiver is 300 Ω, the voltage across the receiver is easily found as

$$0.667 = \frac{1}{2} \frac{|V_{s,L}|^2}{300}$$

$$|V_{s,L}| = 20 \text{ V}$$

in comparison with the 30 V obtained across the single load.

Before we leave this example, let us ask ourselves several questions about the voltages on the transmission line. Where is the voltage a maximum and a minimum, and what are these values? Does the phase of the load voltage still differ from the input voltage by 288°? Presumably, if we can answer these questions for the voltage, we could do the same for the current.

Equation (89) serves to locate the voltage maxima at

$$z_{max} = -\frac{1}{2\beta} (\phi + 2m\pi) \quad (m = 0, 1, 2, \ldots)$$

where $\Gamma = |\Gamma| e^{j\phi}$. Thus, with $\beta = 0.8\pi$ and $\phi = \pi$, we find

$$z_{max} = -0.625 \quad \text{and} \quad -1.875 \text{ m}$$

while the minima are $\lambda/4$ distant from the maxima;

$$z_{min} = 0 \quad \text{and} \quad -1.25 \text{ m}$$

and we find that the load voltage (at $z = 0$) is a voltage minimum. This, of course, verifies the general conclusion we reached earlier: a voltage minimum occurs at the load if $Z_L < Z_0$, and a voltage maximum occurs if $Z_L > Z_0$, where both impedances are pure resistances.

The minimum voltage on the line is thus the load voltage, 20 V; the maximum voltage must be 40 V, since the standing wave ratio is 2. The voltage at the input end of the line is

$$V_{s,in} = I_{s,in} Z_{in} = (0.0756\angle 15.0°)(510\angle -23.8°) = 38.5\angle -8.8°$$

The input voltage is almost as large as the maximum voltage anywhere on the line because the line is about three-quarters of a wavelength long, a length which would place the voltage maximum at the input when $Z_L < Z_0$.

Finally, it is of interest to determine the load voltage in magnitude *and phase*. We begin with the total voltage in the line, using (93).

$$V_{sT} = \left(e^{-j\beta z} + \Gamma e^{j\beta z} \right) V_0^+ \tag{104}$$

We may use this expression to determine the voltage at any point on the line in terms of the voltage at any other point. Because we know the voltage at the input to the line, we let $z = -l$,

$$V_{s,in} = \left(e^{j\beta l} + \Gamma e^{-j\beta l} \right) V_0^+ \tag{105}$$

and solve for V_0^+,

$$V_0^+ = \frac{V_{s,in}}{e^{j\beta l} + \Gamma e^{-j\beta l}} = \frac{38.5\angle -8.8°}{e^{j1.6\pi} - \frac{1}{3} e^{-j1.6\pi}} = 30.0\angle 72.0° \text{ V}$$

We may now let $z = 0$ in (104) to find the load voltage,

$$V_{s,L} = (1 + \Gamma)V_0^+ = 20\angle 72° = 20\angle -288°$$

The amplitude agrees with our previous value. The presence of the reflected wave causes $V_{s,\text{in}}$ and $V_{s,L}$ to differ in phase by about $-279°$ instead of $-288°$.

In order to provide a slightly more complicated example, let us now place a purely capacitive impedance of $-j300\,\Omega$ in parallel with the two $300\,\Omega$ receivers. We are to find the input impedance and the power delivered to each receiver.

Solution. The load impedance is now $150\,\Omega$ in parallel with $-j300\,\Omega$, or

$$Z_L = \frac{150(-j300)}{150 - j300} = \frac{-j300}{1 - j2} = 120 - j60\,\Omega$$

We first calculate the reflection coefficient and the VSWR:

$$\Gamma = \frac{120 - j60 - 300}{120 - j60 + 300} = \frac{-180 - j60}{420 - j60} = 0.447\angle -153.4°$$

$$s = \frac{1 + 0.447}{1 - 0.447} = 2.62$$

Thus, the VSWR is higher and the mismatch is therefore worse. Let us next calculate the input impedance. The electrical length of the line is still $288°$, so that

$$Z_{\text{in}} = 300 \frac{(120 - j60)\cos 288° + j300 \sin 288°}{300 \cos 288° + j(120 - j60)\sin 288°} = 755 - j138.5\,\Omega$$

This leads to a source current of

$$I_{s,\text{in}} = \frac{V_{Th}}{Z_{Th} + Z_{\text{in}}} = \frac{60}{300 + 755 - j138.5} = 0.0564\angle 7.47°\,\text{A}$$

Therefore, the average power delivered to the input of the line is $P_{\text{in}} = \frac{1}{2}(0.0564)^2(755) = 1.200$ W. Since the line is lossless, it follows that $P_L = 1.200$ W, and each receiver gets only 0.6 W.

As a final example, let us terminate our line with a purely capacitive impedance, $Z_L = -j300\,\Omega$. We seek the reflection coefficient, the VSWR, and the power delivered to the load.

Solution. Obviously, we cannot deliver any average power to the load since it is a pure reactance. As a consequence, the reflection coefficient is

$$\Gamma = \frac{-j300 - 300}{-j300 + 300} = -j1 = 1\angle -90°$$

and the reflected wave is equal in amplitude to the incident wave. Hence, it should not surprise us to see that the VSWR is

$$s = \frac{1 + |-j1|}{1 - |-j1|} = \infty$$

and the input impedance is a pure reactance,

$$Z_{\text{in}} = 300 \frac{-j300 \cos 288° + j300 \sin 288°}{300 \cos 288° + j(-j300) \sin 288°} = j589$$

Thus, no average power can be delivered to the input impedance by the source, and therefore no average power can be delivered to the load.

Although we could continue to find numerous other facts and figures for these examples, much of the work may be done more easily for problems of this type by using graphical techniques. We encounter these in Section 10.13.

D10.4. A 50 W lossless line has a length of 0.4λ. The operating frequency is 300 MHz. A load $Z_L = 40 + j30$ Ω is connected at $z = 0$, and the Thevenin-equivalent source at $z = -l$ is $12\angle 0°$ V in series with $Z_{Th} = 50 + j0$ Ω. Find: (a) Γ; (b) s; (c) Z_{in}.

Ans. $0.333\angle 90°$; 2.00; $25.5 + j5.90$ Ω

D10.5. For the transmission line of Problem D10.4, also find: (a) the phasor voltage at $z = -l$; (b) the phasor voltage at $z = 0$; (c) the average power delivered to Z_L.

Ans. $4.14\angle 8.58°$ V; $6.32\angle -125.6°$ V; 0.320 W

10.13 GRAPHICAL METHODS: THE SMITH CHART

Transmission line problems often involve manipulations with complex numbers, making the time and effort required for a solution several times greater than are needed for a similar sequence of operations on real numbers. One means of reducing the labor without seriously affecting the accuracy is by using transmission-line charts. Probably the most widely used one is the Smith chart.[3]

Basically, this diagram shows curves of constant resistance and constant reactance; these may represent either an input impedance or a load impedance. The latter, of course, is the input impedance of a zero-length line. An indication of location along the line is also provided, usually in terms of the fraction of a wavelength from a voltage maximum or minimum. Although they are not specifically shown on the chart, the standing-wave ratio and the magnitude and angle of the reflection coefficient are very

---

[3] P. H. Smith, "Transmission Line Calculator," *Electronics,* vol. 12, pp. 29–31, January 1939.
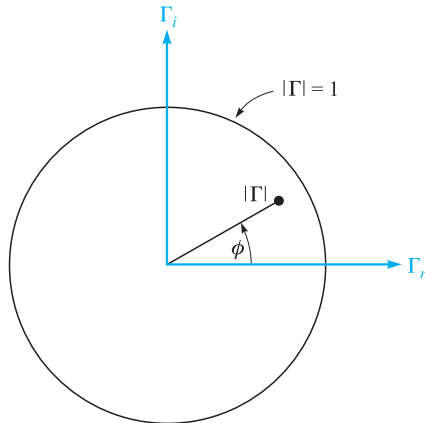
Figure 10.9 The polar coordinates of the Smith chart are the magnitude and phase angle of the reflection coefficient; the rectangular coordinates are the real and imaginary parts of the reflection coefficient. The entire chart lies within the circle $|\Gamma| = 1$.

quickly determined. As a matter of fact, the diagram is constructed within a circle of unit radius, using polar coordinates, with radius variable $|\Gamma|$ and counterclockwise angle variable ϕ, where $\Gamma = |\Gamma|e^{j\phi}$. Figure 10.9 shows this circle. Since $|\Gamma| < 1$, all our information must lie on or within the unit circle. Peculiarly enough, the reflection coefficient itself will not be plotted on the final chart, for these additional contours would make the chart very difficult to read.

The basic relationship upon which the chart is constructed is

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{106}$$

The impedances that we plot on the chart will be *normalized* with respect to the characteristic impedance. Let us identify the normalized load impedance as z_L,

$$z_L = r + jx = \frac{Z_L}{Z_0} = \frac{R_L + jX_L}{Z_0}$$

and thus

$$\Gamma = \frac{z_L - 1}{z_L + 1}$$

or

$$z_L = \frac{1 + \Gamma}{1 - \Gamma} \tag{107}$$

In polar form, we have used $|\Gamma|$ and ϕ as the magnitude and angle of Γ. With Γ_r and Γ_i as the real and imaginary parts of Γ, we write

$$\Gamma = \Gamma_r + j\Gamma_i \tag{108}$$

Thus

$$r + jx = \frac{1 + \Gamma_r + j\Gamma_i}{1 - \Gamma_r - j\Gamma_i} \tag{109}$$

The real and imaginary parts of this equation are

$$r = \frac{1 - \Gamma_r^2 - \Gamma_i^2}{(1 - \Gamma_r)^2 + \Gamma_i^2} \tag{110}$$

$$x = \frac{2\Gamma_i}{(1 - \Gamma_r)^2 + \Gamma_i^2} \tag{111}$$

After several lines of elementary algebra, we may write (110) and (111) in forms which readily display the nature of the curves on Γ_r, Γ_i axes,

$$\left(\Gamma_r - \frac{r}{1+r}\right)^2 + \Gamma_i^2 = \left(\frac{1}{1+r}\right)^2 \tag{112}$$

$$(\Gamma_r - 1)^2 + \left(\Gamma_i - \frac{1}{x}\right)^2 = \left(\frac{1}{x}\right)^2 \tag{113}$$

The first equation describes a family of circles, where each circle is associated with a specific value of resistance r. For example, if $r = 0$, the radius of this zero-resistance circle is seen to be unity, and it is centered at the origin ($\Gamma_r = 0$, $\Gamma_i = 0$). This checks, for a pure reactance termination leads to a reflection coefficient of unity magnitude. On the other hand, if $r = \infty$, then $z_L = \infty$ and we have $\Gamma = 1 + j0$. The circle described by (112) is centered at $\Gamma_r = 1$, $\Gamma_i = 0$ and has zero radius. It is therefore the point $\Gamma = 1 + j0$, as we decided it should be. As another example, the circle for $r = 1$ is centered at $\Gamma_r = 0.5$, $\Gamma_i = 0$ and has a radius of 0.5. This circle is shown in Figure 10.10, along with circles for $r = 0.5$ and $r = 2$. All circles are centered on the Γ_r axis and pass through the point $\Gamma = 1 + j0$.

Equation (113) also represents a family of circles, but each of these circles is defined by a particular value of x, rather than r. If $x = \infty$, then $z_L = \infty$, and $\Gamma = 1 + j0$ again. The circle described by (113) is centered at $\Gamma = 1 + j0$ and has zero radius; it is therefore the point $\Gamma = 1 + j0$. If $x = +1$, then the circle is centered at $\Gamma = 1 + j1$ and has unit radius. Only one-quarter of this circle lies within the boundary curve $|\Gamma| = 1$, as shown in Figure 10.11. A similar quarter-circle appears below the Γ_r axis for $x = -1$. The portions of other circles for $x = 0.5, -0.5, 2$, and -2 are also shown. The "circle" representing $x = 0$ is the Γ_r axis; this is also labeled in Figure 10.11.

The two families of circles both appear on the Smith chart, as shown in Figure 10.12. It is now evident that if we are given Z_L, we may divide by Z_0 to
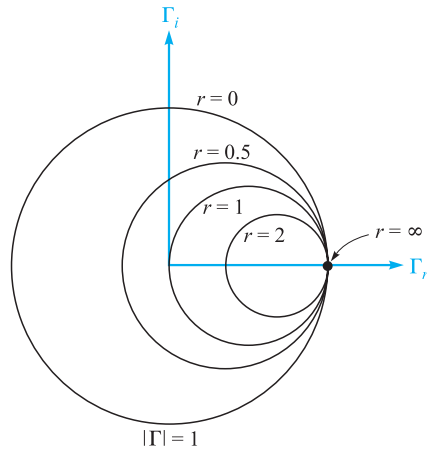
Figure 10.10 Constant-r circles are shown on the Γ_r, Γ_i plane. The radius of any circle is $1/(1+r)$.

obtain z_L, locate the appropriate r and x circles (interpolating as necessary), and determine Γ by the intersection of the two circles. Because the chart does not have concentric circles showing the values of $|\Gamma|$, it is necessary to measure the radial distance from the origin to the intersection with dividers or a compass and use an auxiliary scale to find $|\Gamma|$. The graduated line segment below the chart in Figure 10.12 serves this purpose. The angle of Γ is ϕ, and it is the counterclockwise angle from the Γ_r axis. Again, radial lines showing the angle would clutter up the chart
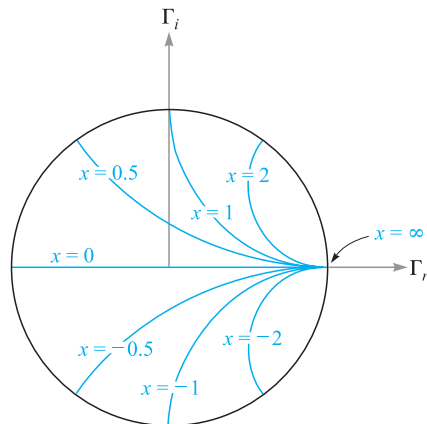
Figure 10.11 The portions of the circles of constant x lying within $|\Gamma| = 1$ are shown on the Γ_r, Γ_i axes. The radius of a given circle is $1/|x|$.
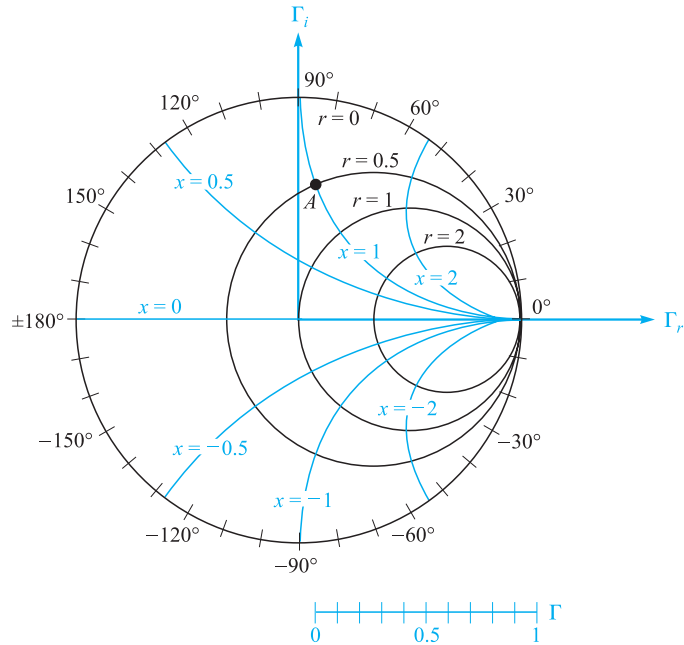
Figure 10.12   The Smith chart contains the constant-r circles and constant-x circles, an auxiliary radial scale to determine $|\Gamma|$, and an angular scale on the circumference for measuring ϕ.

badly, so the angle is indicated on the circumference of the circle. A straight line from the origin through the intersection may be extended to the perimeter of the chart. As an example, if $Z_L = 25 + j50\,\Omega$ on a 50 Ω line, $z_L = 0.5 + j1$, and point A on Figure 10.12 shows the intersection of the $r = 0.5$ and $x = 1$ circles. The reflection coefficient is approximately 0.62 at an angle ϕ of 83°.

The Smith chart is completed by adding a second scale on the circumference by which distance along the line may be computed. This scale is in wavelength units, but the values placed on it are not obvious. To obtain them, we first divide the voltage at any point along the line,

$$V_s = V_0^+(e^{-j\beta z} + \Gamma e^{j\beta z})$$

by the current

$$I_s = \frac{V_0^+}{Z_0}(e^{-j\beta z} - \Gamma e^{j\beta z})$$

obtaining the normalized input impedance

$$z_{\text{in}} = \frac{V_s}{Z_0 I_s} = \frac{e^{-j\beta z} + \Gamma e^{j\beta z}}{e^{-j\beta z} - \Gamma e^{j\beta z}}$$

Replacing z with $-l$ and dividing numerator and denominator by $e^{j\beta l}$, we have the general equation relating normalized input impedance, reflection coefficient, and

line length,

$$z_{in} = \frac{1 + \Gamma e^{-j2\beta l}}{1 - \Gamma e^{-j2\beta l}} = \frac{1 + |\Gamma|e^{j(\phi - 2\beta l)}}{1 - |\Gamma|e^{j(\phi - 2\beta l)}} \tag{114}$$

Note that when $l = 0$, we are located at the load, and $z_{in} = (1 + \Gamma)/(l - \Gamma) = z_L$, as shown by (107).

Equation (114) shows that the input impedance at any point $z = -l$ can be obtained by replacing Γ, the reflection coefficient of the load, by $\Gamma e^{-j2\beta l}$. That is, we decrease the angle of Γ by $2\beta l$ radians as we move from the load to the line input. Only the angle of Γ is changed; the magnitude remains constant.

Thus, as we proceed from the load z_L to the input impedance z_{in}, we move *toward* the generator a distance l on the transmission line, but we move through a *clockwise* angle of $2\beta l$ on the Smith chart. Since the magnitude of Γ stays constant, the movement toward the source is made along a constant-radius circle. One lap around the chart is accomplished whenever βl changes by π rad, or when l changes by one-half wavelength. This agrees with our earlier discovery that the input impedance of a half-wavelength lossless line is equal to the load impedance.

The Smith chart is thus completed by the addition of a scale showing a change of 0.5λ for one circumnavigation of the unit circle. For convenience, two scales are usually given, one showing an increase in distance for clockwise movement and the other an increase for counterclockwise travel. These two scales are shown in Figure 10.13. Note that the one marked "wavelengths toward generator" (wtg) shows increasing values of l/λ for clockwise travel, as described previously. The zero point of the wtg scale is rather arbitrarily located to the left. This corresponds to input impedances having phase angles of $0°$ and $R_L < Z_0$. We have also seen that voltage minima are always located here.

EXAMPLE 10.10

The use of the transmission line chart is best shown by example. Let us again consider a load impedance, $Z_L = 25 + j50\ \Omega$, terminating a 50-Ω line. The line length is 60 cm and the operating frequency is such that the wavelength on the line is 2 m. We desire the input impedance.

Solution. We have $z_L = 0.5 + j1$, which is marked as A on Figure 10.14, and we read $\Gamma = 0.62\angle 82°$. By drawing a straight line from the origin through A to the circumference, we note a reading of 0.135 on the wtg scale. We have $l/\lambda = 0.6/2 = 0.3$, and it is, therefore, 0.3λ from the load to the input. We therefore find z_{in} on the $|\Gamma| = 0.62$ circle opposite a wtg reading of $0.135 + 0.300 = 0.435$. This construction is shown in Figure 10.14, and the point locating the input impedance is marked B. The normalized input impedance is read as $0.28 - j0.40$, and thus $Z_{in} = 14 - j20$. A more accurate analytical calculation gives $Z_{in} = 13.7 - j20.2$.

Information concerning the location of the voltage maxima and minima is also readily obtained on the Smith chart. We already know that a maximum or minimum
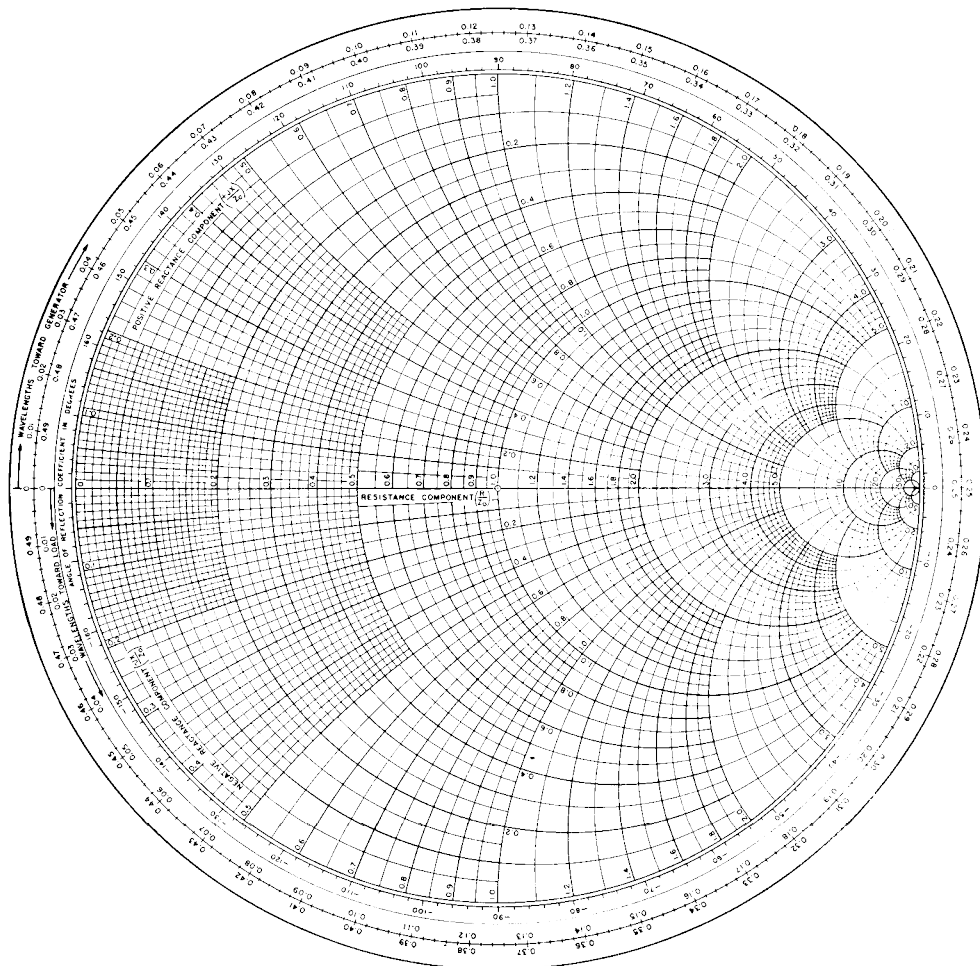
Figure 10.13 A photographic reduction of one version of a useful Smith chart (*courtesy of the Emeloid Company, Hillside, NJ*). For accurate work, larger charts are available wherever fine technical books are sold.

must occur at the load when Z_L is a pure resistance; if $R_L > Z_0$ there is a maximum at the load, and if $R_L < Z_0$ there is a minimum. We may extend this result now by noting that we could cut off the load end of a transmission line at a point where the input impedance is a pure resistance and replace that section with a resistance R_{in}; there would be no changes on the generator portion of the line. It follows, then, that the location of voltage maxima and minima must be at those points where Z_{in} is a pure resistance. Purely resistive input impedances must occur on the $x = 0$ line (the Γ_r axis) of the Smith chart. Voltage maxima or current minima occur when $r > 1$, or at wtg $= 0.25$, and voltage minima or current maxima occur when $r < 1$,
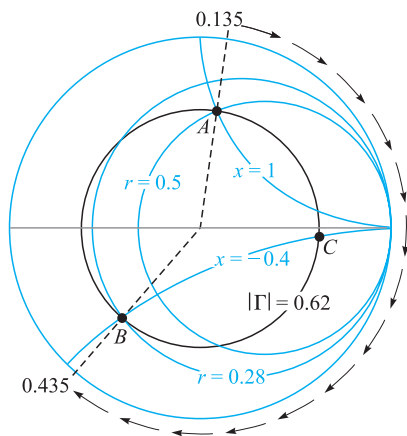
Figure 10.14 Normalized input impedance produced by a normalized load impedance $z_L = 0.5 + j1$ on a line 0.3λ long is $z_{in} = 0.28 - j0.40$.

or at wtg $= 0$. In Example 10.10, then, the maximum at wtg $= 0.250$ must occur $0.250 - 0.135 = 0.115$ wavelengths toward the generator from the load. This is a distance of 0.115×200, or 23 cm from the load.

We should also note that because the standing wave ratio produced by a resistive load R_L is either R_L/R_0 or R_0/R_L, whichever is greater than unity, the value of s may be read directly as the value of r at the intersection of the $|\Gamma|$ circle and the r axis, $r > 1$. In our example, this intersection is marked point C, and $r = 4.2$; thus, $s = 4.2$.

Transmission line charts may also be used for normalized admittances, although there are several slight differences in such use. We let $y_L = Y_L/Y_0 = g + jb$ and use the r circles as g circles and the x circles as b circles. The two differences are, first, the line segment where $g > 1$ and $b = 0$ corresponds to a voltage minimum; and second, $180°$ must be added to the angle of Γ as read from the perimeter of the chart. We shall use the Smith chart in this way in Section 10.14.

Special charts are also available for non-normalized lines, particularly 50 Ω charts and 20 mS charts.

> **D10.6.** A load $Z_L = 80 - j100 \ \Omega$ is located at $z = 0$ on a lossless 50-Ω line. The operating frequency is 200 MHz and the wavelength on the line is 2 m. (a) If the line is 0.8 m in length, use the Smith chart to find the input impedance. (b) What is s? (c) What is the distance from the load to the nearest voltage maximum? (d) What is the distance from the input to the nearest point at which the remainder of the line could be replaced by a pure resistance?
>
> **Ans.** $79 + j99 \ \Omega$: 4.50; 0.0397 m; 0.760 m
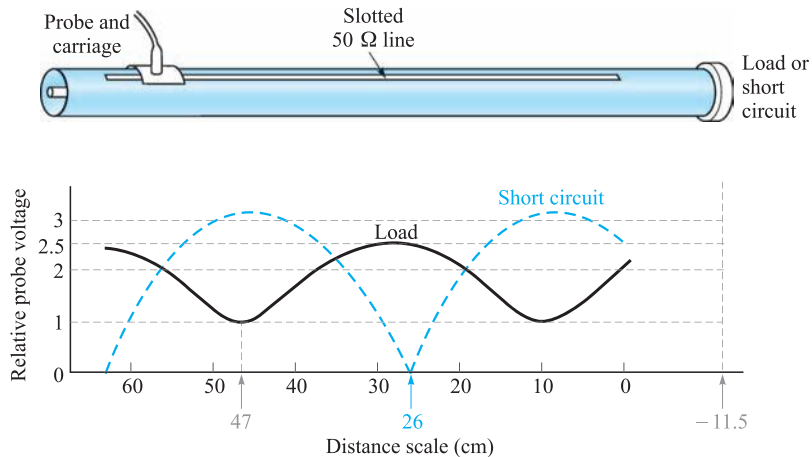
Figure 10.15   A sketch of a coaxial slotted line. The distance scale is on the slotted line. With the load in place, $s = 2.5$, and the minimum occurs at a scale reading of 47 cm. For a short circuit, the minimum is located at a scale reading of 26 cm. The wavelength is 75 cm.

We next consider two examples of practical transmission line problems. The first is the determination of load impedance from experimental data, and the second is the design of a single-stub matching network.

Let us assume that we have made experimental measurements on a 50 Ω slotted line that show there is a voltage standing wave ratio of 2.5. This has been determined by moving a sliding carriage back and forth along the line to determine maximum and minimum voltage readings. A scale provided on the track along which the carriage moves indicates that a *minimum* occurs at a scale reading of 47.0 cm, as shown in Figure 10.15. The zero point of the scale is arbitrary and does not correspond to the location of the load. The location of the minimum is usually specified instead of the maximum because it can be determined more accurately than that of the maximum; think of the sharper minima on a rectified sine wave. The frequency of operation is 400 MHz, so the wavelength is 75 cm. In order to pinpoint the location of the load, we remove it and replace it with a short circuit; the position of the minimum is then determined as 26.0 cm.

We know that the short circuit must be located an integral number of half-wavelengths from the minimum; let us arbitrarily locate it one half-wavelength away at $26.0 - 37.5 = -11.5$ cm on the scale. Since the short circuit has replaced the load, the load is also located at -11.5 cm. Our data thus show that the minimum is $47.0 - (-11.5) = 58.5$ cm from the load, or subtracting one-half wavelength, a minimum is 21.0 cm from the load. The voltage *maximum* is thus $21.0 - (37.5/2) = 2.25$ cm from the load, or $2.25/75 = 0.030$ wavelength from the load.

With this information, we can now turn to the Smith chart. At a voltage maximum, the input impedance is a pure resistance equal to $s R_0$; on a normalized basis, $z_{in} = 2.5$.
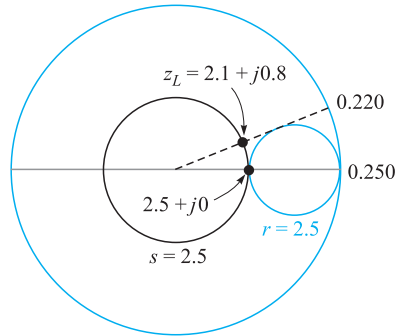
Figure 10.16   If $z_{in} = 2.5 + j0$ on a line 0.3 wavelengths long, then $z_L = 2.1 + j0.8$.

We therefore enter the chart at $z_{in} = 2.5$ and read 0.250 on the wtg scale. Subtracting 0.030 wavelength to reach the load, we find that the intersection of the $s = 2.5$ (or $|\Gamma| = 0.429$) circle and the radial line to 0.220 wavelength is at $z_L = 2.1 + j0.8$. The construction is sketched on the Smith chart of Figure 10.16. Thus $Z_L = 105 + j40 \, \Omega$, a value that assumes its location at a scale reading of -11.5 cm, or an integral number of half-wavelengths from that position. Of course, we may select the "location" of our load at will by placing the short circuit at the point that we wish to consider the load location. Since load locations are not well defined, it is important to specify the point (or plane) at which the load impedance is determined.

As a final example, let us try to match this load to the 50 Ω line by placing a short-circuited stub of length d_1 a distance d from the load (see Figure 10.17). The stub line has the same characteristic impedance as the main line. The lengths d and d_1 are to be determined.
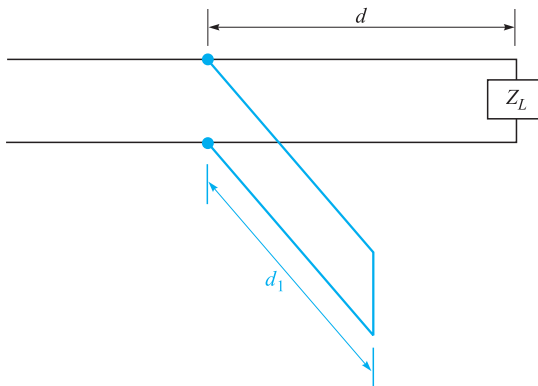
Figure 10.17   A short-circuited stub of length d_1, located at a distance d from a load Z_L, is used to provide a matched load to the left of the stub.

The input impedance to the stub is a pure reactance; when combined in parallel with the input impedance of the length d containing the load, the resultant input impedance must be $1 + j0$. Because it is much easier to combine admittances in parallel than impedances, let us rephrase our goal in admittance language: the input admittance of the length d containing the load must be $1 + jb_{in}$ for the addition of the input admittance of the stub jb_{stub} to produce a total admittance of $1 + j0$. Hence the stub admittance is $-jb_{in}$. We will therefore use the Smith chart as an admittance chart instead of an impedance chart.

The impedance of the load is $2.1 + j0.8$, and its location is at -11.5 cm. The admittance of the load is therefore $1/(2.1 + j0.8)$, and this value may be determined by adding one-quarter wavelength on the Smith chart, as Z_{in} for a quarter-wavelength line is R_0^2/Z_L, or $z_{in} = 1/z_L$, or $y_{in} = z_L$. Entering the chart (Figure 10.18) at $z_L = 2.1 + j0.8$, we read 0.220 on the wtg scale; we add (or subtract) 0.250 and find the admittance $0.41 - j0.16$ corresponding to this impedance. This point is still located on the $s = 2.5$ circle. Now, at what point or points on this circle is the real part of the admittance equal to unity? There are two answers, $1 + j0.95$ at wtg $= 0.16$, and $1 - j0.95$ at wtg $= 0.34$, as shown in Figure 10.18. We select the former value since this leads to the shorter stub. Hence $y_{stub} = -j0.95$, and the stub location corresponds to wtg $= 0.16$. Because the load admittance was found at wtg $= 0.470$, then we must move $(0.5 - 0.47) + 0.16 = 0.19$ wavelength to get to the stub location.

Finally, we may use the chart to determine the necessary length of the short-circuited stub. The input conductance is zero for any length of short-circuited stub, so we are restricted to the perimeter of the chart. At the short circuit, $y = \infty$ and wtg $= 0.250$. We find that $b_{in} = -0.95$ is achieved at wtg $= 0.379$, as shown in Figure 10.18. The stub is therefore $0.379 - 0.250 = 0.129$ wavelength, or 9.67 cm long.
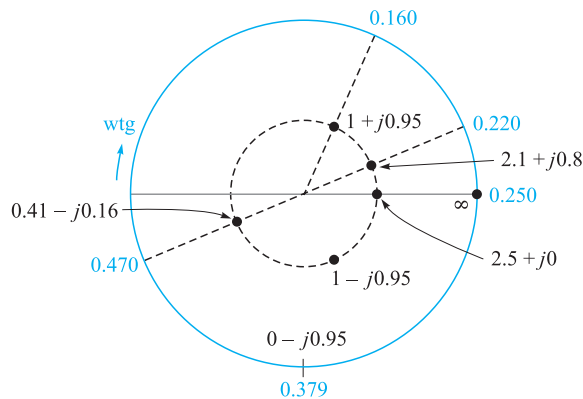
Figure 10.18 A normalized load, $z_L = 2.1 + j0.8$, is matched by placing a 0.129-wavelength short-circuited stub 0.19 wavelengths from the load.

D10.7. Standing wave measurements on a lossless 75-Ω line show maxima of 18 V and minima of 5 V. One minimum is located at a scale reading of 30 cm. With the load replaced by a short circuit, two adjacent minima are found at scale readings of 17 and 37 cm. Find: (*a*) s; (*b*) λ; (*c*) f; (*d*) Γ_L; (*e*) Z_L.

Ans. 3.60; 0.400 m; 750 MHz; $0.704\angle{-33.0}$; $77.9 + j104.7\,\Omega$

D10.8. A normalized load, $z_L = 2 - j1$, is located at $z = 0$ on a lossless 50-Ω line. Let the wavelength be 100 cm. (*a*) A short-circuited stub is to be located at $z = -d$. What is the shortest suitable value for d? (*b*) What is the shortest possible length of the stub? Find s: (*c*) on the main line for $z < -d$; (*d*) on the main line for $-d < z < 0$; (*e*) on the stub.

Ans. 12.5 cm; 12.5 cm; 1.00; 2.62; ∞

10.14 TRANSIENT ANALYSIS

Throughout most of this chapter, we have considered the operation of transmission lines under steady-state conditions, in which voltage and current were sinusoidal and at a single frequency. In this section we move away from the simple time-harmonic case and consider transmission line responses to voltage step functions and pulses, grouped under the general heading of *transients*. These situations were briefly considered in Section 10.2 with regard to switched voltages and currents. Line operation in transient mode is important to study because it allows us to understand how lines can be used to store and release energy (in pulse-forming applications, for example). Pulse propagation is important in general since digital signals, composed of sequences of pulses, are widely used.

We will confine our discussion to the propagation of transients in lines that are lossless and have no dispersion, so that the basic behavior and analysis methods may be learned. We must remember, however, that transient signals are necessarily composed of numerous frequencies, as Fourier analysis will show. Consequently, the question of dispersion in the line arises, since, as we have found, line propagation constants and reflection coefficients at complex loads will be frequency-dependent. So, in general, pulses are likely to broaden with propagation distance, and pulse shapes may change when reflecting from a complex load. These issues will not be considered in detail here, but they are readily addressed when the precise frequency dependences of β and Γ are known. In particular, $\beta(\omega)$ can be found by evaluating the imaginary part of γ, as given in Eq. (41), which would in general include the frequency dependences of R, C, G, and L arising from various mechanisms. For example, the skin effect (which affects both the conductor resistance and the internal inductance) will result in frequency-dependent R and L. Once $\beta(\omega)$ is known, pulse broadening can be evaluated using the methods to be presented in Chapter 12.

We begin our basic discussion of transients by considering a lossless transmission line of length l terminated by a matched load, $R_L = Z_0$, as shown in Figure 10.19*a*.
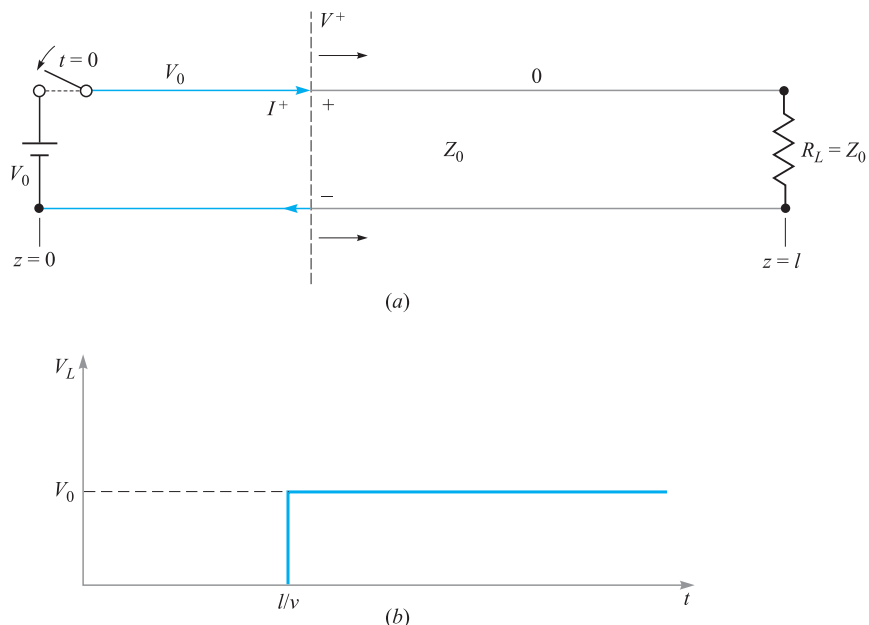
Figure 10.19 (a) Closing the switch at time $t = 0$ initiates voltage and current waves V^+ and I^+. The leading edge of both waves is indicated by the dashed line, which propagates in the lossless line toward the load at velocity v. In this case, $V^+ = V_0$; the line voltage is V^+ everywhere to the left of the leading edge, where current is $I^+ = V^+/Z_0$. To the right of the leading edge, voltage and current are both zero. Clockwise current, indicated here, is treated as positive and will occur when V^+ is positive. (b) Voltage across the load resistor as a function of time, showing the one-way transit time delay, l/v.

At the front end of the line is a battery of voltage V_0, which is connected to the line by closing a switch. At time $t = 0$, the switch is closed, and the line voltage at $z = 0$ becomes equal to the battery voltage. This voltage, however, does not appear across the load until adequate time has elapsed for the propagation delay. Specifically, at $t = 0$, a voltage wave is initiated in the line at the battery end, which then propagates toward the load. The leading edge of the wave, labeled V^+ in Figure 10.19, is of value $V^+ = V_0$. It can be thought of as a propagating step function, because at all points to the left of V^+, the line voltage is V_0; at all points to the right (not yet reached by the leading edge), the line voltage is zero. The wave propagates at velocity v, which in general is the group velocity in the line.[4] The wave reaches the load at time $t = l/v$

---

[4] Because we have a step function (composed of many frequencies) as opposed to a sinusoid at a single frequency, the wave will propagate at the group velocity. In a lossless line with no dispersion as considered in this section, $\beta = \omega\sqrt{LC}$, where L and C are constant with frequency. In this case, we would find that the group and phase velocities are equal; that is, $d\omega/d\beta = \omega/\beta = v = 1/\sqrt{LC}$. We will thus write the velocity as v, knowing it to be both v_p and v_g.

and then does not reflect, as the load is matched. The transient phase is thus over, and the load voltage is equal to the battery voltage. A plot of load voltage as a function of time is shown in Figure 10.19*b*, indicating the propagation delay of $t = l/v$.

Associated with the voltage wave V^+ is a current wave whose leading edge is of value I^+. This wave is a propagating step function as well, whose value at all points to the left of V^+ is $I^+ = V^+/Z_0$; at all points to the right, current is zero. A plot of current through the load as a function of time will thus be identical in form to the voltage plot of Figure 10.19*b*, except that the load current at $t = l/v$ will be $I_L = V^+/Z_0 = V_0/R_L$.

We next consider a more general case, in which the load of Figure 10.19*a* is again a resistor but is *not matched* to the line ($R_L \neq Z_0$). Reflections will thus occur at the load, complicating the problem. At $t = 0$, the switch is closed as before and a voltage wave, $V_1^+ = V_0$, propagates to the right. Upon reaching the load, however, the wave will now reflect, producing a back-propagating wave, V_1^-. The relation between V_1^- and V_1^+ is through the reflection coefficient at the load:

$$\frac{V_1^-}{V_1^+} = \Gamma_L = \frac{R_L - Z_0}{R_L + Z_0} \tag{115}$$

As V_1^- propagates back toward the battery, it leaves behind its leading edge a total voltage of $V_1^+ + V_1^-$. Voltage V_1^+ exists everywhere ahead of the V_1^- wave until it reaches the battery, whereupon the entire line now is charged to voltage $V_1^+ + V_1^-$. At the battery, the V_1^- wave reflects to produce a new forward wave, V_2^+. The ratio of V_2^+ and V_1^- is found through the reflection coefficient at the battery:

$$\frac{V_2^+}{V_1^-} = \Gamma_g = \frac{Z_g - Z_0}{Z_g + Z_0} = \frac{0 - Z_0}{0 + Z_0} = -1 \tag{116}$$

where the impedance at the generator end, Z_g, is that of the battery, or zero.

V_2^+ (equal to $-V_1^-$) now propagates to the load, where it reflects to produce backward wave $V_2^- = \Gamma_L V_2^+$. This wave then returns to the battery, where it reflects with $\Gamma_g = -1$, and the process repeats. Note that with each round trip the wave voltage is reduced in magnitude because $|\Gamma_L| < 1$. Because of this the propagating wave voltages will eventually approach zero, and steady state is reached.

The voltage across the load resistor can be found at any given time by summing the voltage waves that have reached the load and have reflected from it up to that time. After many round trips, the load voltage will be, in general,

$$V_L = V_1^+ + V_1^- + V_2^+ + V_2^- + V_3^+ + V_3^- + \cdots$$
$$= V_1^+ \left(1 + \Gamma_L + \Gamma_g \Gamma_L + \Gamma_g \Gamma_L^2 + \Gamma_g^2 \Gamma_L^2 + \Gamma_g^2 \Gamma_L^3 + \cdots\right)$$

With a simple factoring operation, the preceding equation becomes

$$V_L = V_1^+ (1 + \Gamma_L)\left(1 + \Gamma_g \Gamma_L + \Gamma_g^2 \Gamma_L^2 + \cdots\right) \tag{117}$$
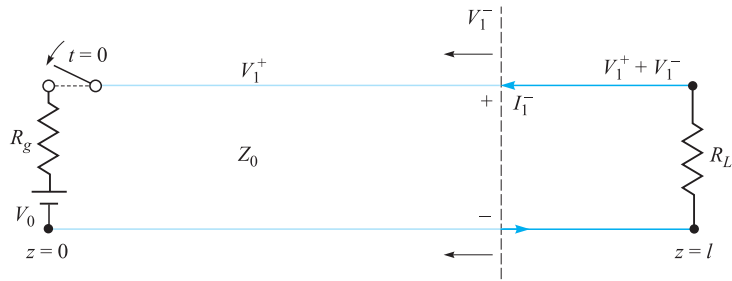
Figure 10.20   With series resistance at the battery location, voltage division occurs when the switch is closed, such that $V_0 = V_{rg} + V_1^+$. Shown is the first reflected wave, which leaves voltage $V_1^+ + V_1^-$ behind its leading edge. Associated with the wave is current I_1^-, which is $-V_1^-/Z_0$. Counterclockwise current is treated as negative and will occur when V_1^- is positive.

Allowing time to approach infinity, the second term in parentheses in (117) becomes the power series expansion for the expression $1/(1 - \Gamma_g\Gamma_L)$. Thus, in steady state we obtain

$$V_L = V_1^+ \left(\frac{1 + \Gamma_L}{1 - \Gamma_g\Gamma_L} \right) \tag{118}$$

In our present example, $V_1^+ = V_0$ and $\Gamma_g = -1$. Substituting these into (118), we find the expected result in steady state: $V_L = V_0$.

A more general situation would involve a nonzero impedance at the battery location, as shown in Figure 10.20. In this case, a resistor of value R_g is positioned in series with the battery. When the switch is closed, the battery voltage appears across the series combination of R_g and the line characteristic impedance, Z_0. The value of the initial voltage wave, V_1^+, is thus found through simple voltage division, or

$$V_1^+ = \frac{V_0 Z_0}{R_g + Z_0} \tag{119}$$

With this initial value, the sequence of reflections and the development of the voltage across the load occurs in the same manner as determined by (117), with the steady-state value determined by (118). The value of the reflection coefficient at the generator end, determined by (116), is $\Gamma_g = (R_g - Z_0)/(R_g + Z_0)$.

A useful way of keeping track of the voltage at any point in the line is through a *voltage reflection diagram*. Such a diagram for the line of Figure 10.20 is shown in Figure 10.21a. It is a two-dimensional plot in which position on the line, z, is shown on the horizontal axis. Time is plotted on the vertical axis and is conveniently expressed as it relates to position and velocity through $t = z/v$. A vertical line, located at $z = l$, is drawn, which, together with the ordinate, defines the z axis boundaries of the transmission line. With the switch located at the battery position, the initial voltage wave, V_1^+, starts at the origin, or lower-left corner of the diagram ($z = t = 0$). The location of the leading edge of V_1^+ as a function of time is shown as the diagonal line
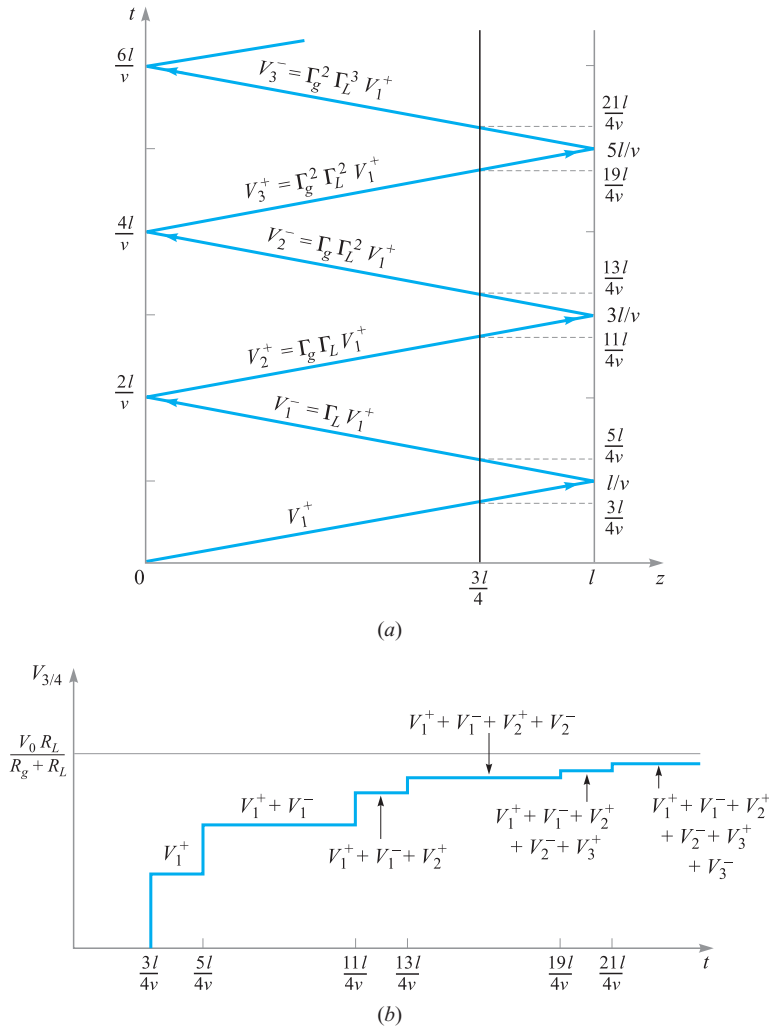
Figure 10.21 (a) Voltage reflection diagram for the line of Figure 10.20. A reference line, drawn at $z = 3l/4$, is used to evaluate the voltage at that position as a function of time. (b) The line voltage at $z = 3l/4$ as determined from the reflection diagram of (a). Note that the voltage approaches the expected $V_0 R_L/(R_g + R_L)$ as time approaches infinity.

that joins the origin to the point along the right-hand vertical line that corresponds to time $t = l/v$ (the one-way transit time). From there (the load location), the position of the leading edge of the reflected wave, V_1^-, is shown as a "reflected" line that joins the $t = l/v$ point on the right boundary to the $t = 2l/v$ point on the ordinate. From there (at the battery location), the wave reflects again, forming V_2^+, shown as a line parallel to that for V_1^+. Subsequent reflected waves are shown, and their values are labeled.

The voltage as a function of time at a given position in the line can now be determined by adding the voltages in the waves as they intersect a vertical line drawn at the desired location. This addition is performed starting at the bottom of the diagram ($t = 0$) and progressing upward (in time). Whenever a voltage wave crosses the vertical line, its value is added to the total at that time. For example, the voltage at a location three-fourths the distance from the battery to the load is plotted in Figure 10.21b. To obtain this plot, the line $z = (3/4)l$ is drawn on the diagram. Whenever a wave crosses this line, the voltage in the wave is added to the voltage that has accumulated at $z = (3/4)l$ over all earlier times. This general procedure enables one to easily determine the voltage at any specific time and location. In doing so, the terms in (117) that have occurred up to the chosen time are being added, but with information on the time at which each term appears.

Line current can be found in a similar way through a *current reflection diagram*. It is easiest to construct the current diagram directly from the voltage diagram by determining a value for current that is associated with each voltage wave. In dealing with current, it is important to keep track of the *sign* of the current because it relates to the voltage waves and their polarities. Referring to Figures 10.19a and 10.20, we use the convention in which current associated with a *forward-z* traveling voltage wave of positive polarity is positive. This would result in current that flows in the clockwise direction, as shown in Figure 10.19a. Current associated with a *backward-z* traveling voltage wave of positive polarity (thus flowing counterclockwise) is negative. Such a case is illustrated in Figure 10.20. In our two-dimensional transmission-line drawings, we assign positive polarity to voltage waves propagating in *either* direction if the upper conductor carries a positive charge and the lower conductor a negative charge. In Figures 10.19a and 10.20, both voltage waves are of positive polarity, so their associated currents will be net positive for the forward wave and net negative for the backward wave. In general, we write

$$I^+ = \frac{V^+}{Z_0} \tag{120}$$

and

$$I^- = -\frac{V^-}{Z_0} \tag{121}$$

Finding the current associated with a backward-propagating voltage wave immediately requires a minus sign, as (121) indicates.

Figure 10.22a shows the current reflection diagram that is derived from the voltage diagram of Figure 10.21a. Note that the current values are labeled in terms of the voltage values, with the appropriate sign added as per (120) and (121). Once the current diagram is constructed, current at a given location and time can be found in exactly the same manner as voltage is found using the voltage diagram. Figure 10.22b shows the current as a function of time at the $z = (3/4)l$ position, determined by summing the current wave values as they cross the vertical line drawn at that location.
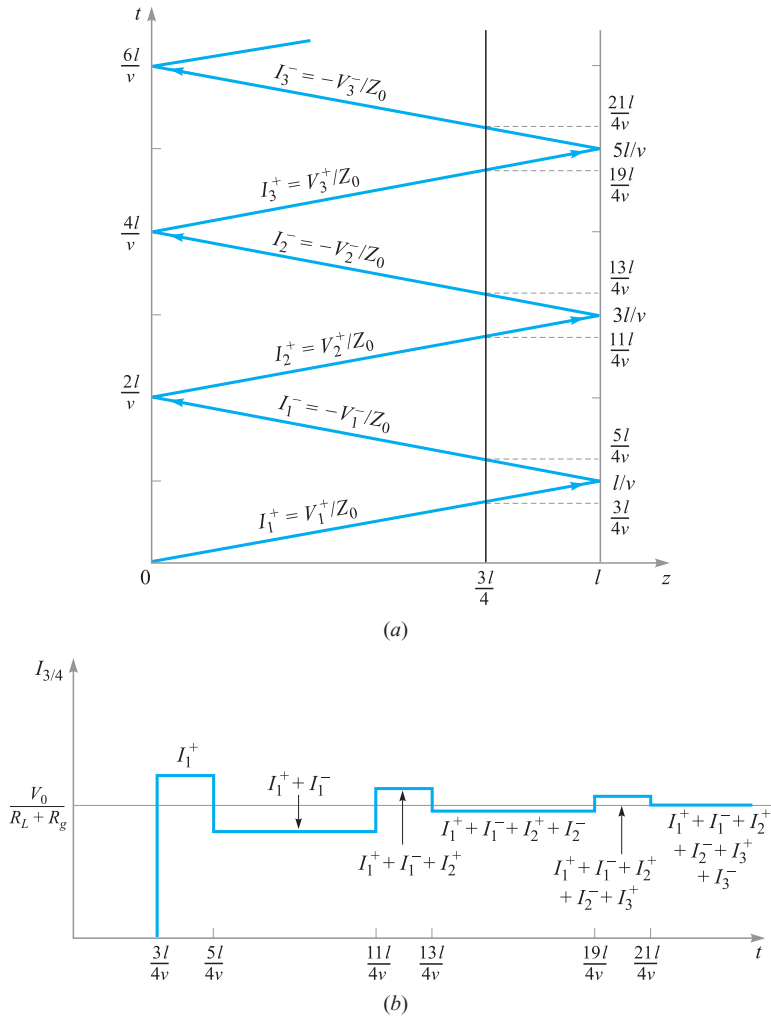
Figure 10.22 (a) Current reflection diagram for the line of Figure 10.20 as obtained from the voltage diagram of Figure 10.21a. (b) Current at the $z = 3l/4$ position as determined from the current reflection diagram, showing the expected steady-state value of $V_0/(R_L + R_g)$.

In Figure 10.20, $R_g = Z_0 = 50\,\Omega$, $R_L = 25\,\Omega$, and the battery voltage is $V_0 = 10$ V. The switch is closed at time $t = 0$. Determine the voltage at the load resistor and the current in the battery as functions of time.

Solution. Voltage and current reflection diagrams are shown in Figure 10.23a and b. At the moment the switch is closed, half the battery voltage appears across the
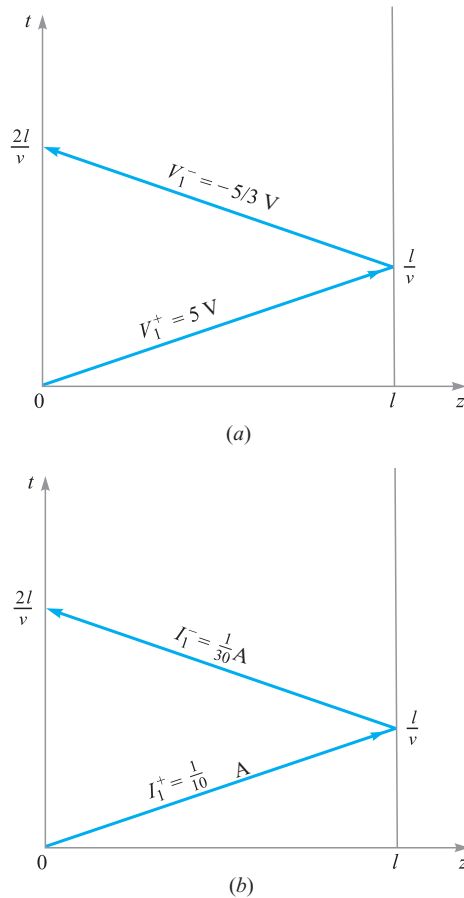
Figure 10.23   Voltage (*a*) and current
(*b*) reflection diagrams for Example 10.11.

50-Ω resistor, with the other half comprising the initial voltage wave. Thus $V_1^+ = (1/2)V_0 = 5$ V. The wave reaches the 25-Ω load, where it reflects with reflection coefficient

$$\Gamma_L = \frac{25 - 50}{25 + 50} = -\frac{1}{3}$$

So $V_1^- = -(1/3)V_1^+ = -5/3$ V. This wave returns to the battery, where it encounters reflection coefficient $\Gamma_g = 0$. Thus, no further waves appear; steady state is reached.

Once the voltage wave values are known, the current reflection diagram can be constructed. The values for the two current waves are

$$I_1^+ = \frac{V_1^+}{Z_0} = \frac{5}{50} = \frac{1}{10} \, \text{A}$$

and

$$I_1^- = -\frac{V_1^-}{Z_0} = -\left(-\frac{5}{3}\right)\left(\frac{1}{50}\right) = \frac{1}{30} \text{ A}$$

Note that no attempt is made here to derive I_1^- from I_1^+. They are both obtained independently from their respective voltages.

The voltage at the load as a function of time is now found by summing the voltages along the vertical line at the load position. The resulting plot is shown in Figure 10.24a. Current in the battery is found by summing the currents along the vertical axis, with the resulting plot shown as Figure 10.24b. Note that in steady state, we treat the circuit as lumped, with the battery in series with the 50- and 25-Ω resistors. Therefore, we expect to see a steady-state current through the battery (and everywhere else) of

$$I_B(\text{steady state}) = \frac{10}{50 + 25} = \frac{1}{7.5} \text{ A}$$

(a)

(b)
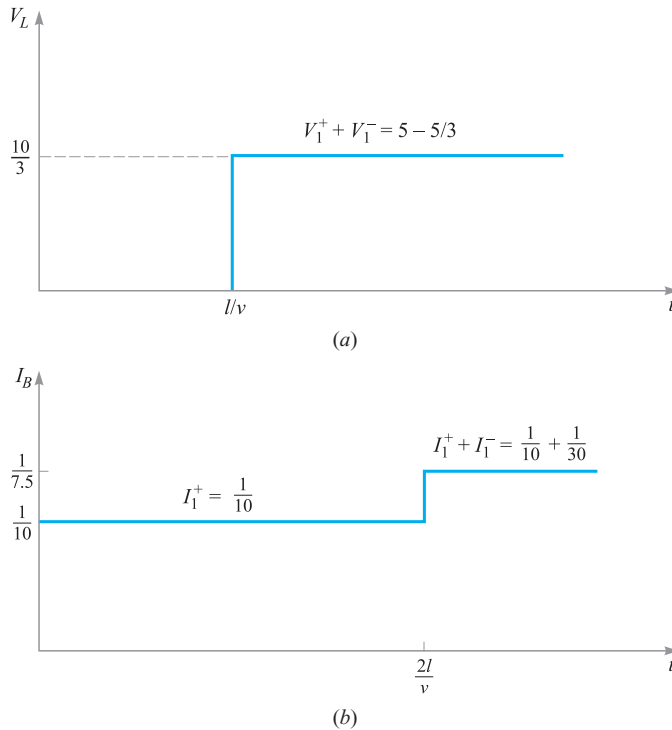
Figure 10.24 Voltage across the load (a) and current in the battery (b) as determined from the reflection diagrams of Figure 10.23 (Example 10.11).
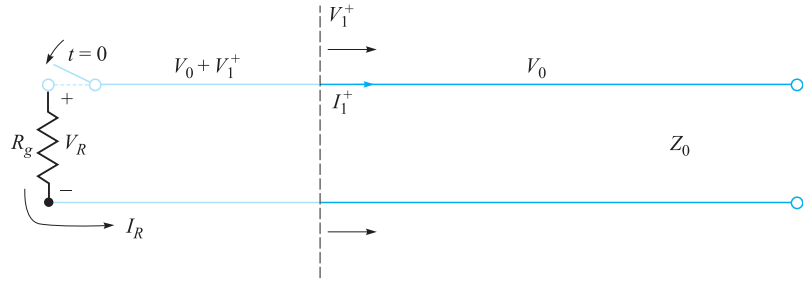
Figure 10.25   In an initially charged line, closing the switch as shown initiates a voltage wave of opposite polarity to that of the initial voltage. The wave thus depletes the line voltage and will fully discharge the line in one round trip if $R_g = Z_0$.

This value is also found from the current reflection diagram for $t > 2l/v$. Similarly, the steady-state load voltage should be

$$V_L(\text{steady state}) = V_0 \frac{R_L}{R_g + R_L} = \frac{(10)(25)}{50 + 25} = \frac{10}{3} \text{ V}$$

which is found also from the voltage reflection diagram for $t > l/v$.

Another type of transient problem involves lines that are *initially charged*. In these cases, the manner in which the line discharges through a load is of interest. Consider the situation shown in Figure 10.25, in which a charged line of characteristic impedance Z_0 is discharged through a resistor of value R_g when a switch at the resistor location is closed.[5] We consider the resistor at the $z = 0$ location; the other end of the line is open (as would be necessary) and is located at $z = l$.

When the switch is closed, current I_R begins to flow through the resistor, and the line discharge process begins. This current does not immediately flow everywhere in the transmission line but begins at the resistor and establishes its presence at more distant parts of the line as time progresses. By analogy, consider a long line of automobiles at a red light. When the light turns green, the cars at the front move through the intersection first, followed successively by those further toward the rear. The point that divides cars in motion and those standing still is, in fact, a wave that propagates toward the back of the line. In the transmission line, the flow of charge progresses in a similar way. A voltage wave, V_1^+, is initiated and propagates to the right. To the left of its leading edge, charge is in motion; to the right of the leading edge, charge is stationary and carries its original density. Accompanying the charge in motion to the left of V_1^+ is a drop in the charge density as the discharge process occurs, and so the line voltage to the left of V_1^+ is partially reduced. This voltage will be given by the sum of the initial voltage, V_0, and V_1^+, which means that V_1^+ must

---

[5] Even though this is a load resistor, we will call it R_g because it is located at the front (generator) end of the line.

in fact be negative (or of opposite sign to V_0). The line discharge process is analyzed by keeping track of V_1^+ as it propagates and undergoes multiple reflections at the two ends. Voltage and current reflection diagrams are used for this purpose in much the same way as before.

Referring to Figure 10.25, we see that for positive V_0 the current flowing through the resistor will be counterclockwise and hence negative. We also know that continuity requires that the resistor current be equal to the current associated with the voltage wave, or

$$ I_R = I_1^+ = \frac{V_1^+}{Z_0} $$

Now the resistor voltage will be

$$ V_R = V_0 + V_1^+ = -I_R R_g = -I_1^+ R_g = -\frac{V_1^+}{Z_0} R_g $$

where the minus signs arise from the fact that V_R (having positive polarity) is produced by the negative current, I_R. We solve for V_1^+ to obtain

$$ V_1^+ = \frac{-V_0 Z_0}{Z_0 + R_g} \tag{122} $$

Having found V_1^+, we can set up the voltage and current reflection diagrams. The diagram for voltage is shown in Figure 10.26. Note that the initial condition of voltage V_0 everywhere on the line is accounted for by assigning voltage V_0 to the horizontal axis of the voltage diagram. The diagram is otherwise drawn as before, but with $\Gamma_L = 1$ (at the open-circuited load end). Variations in how the line discharges thus depend on the resistor value at the switch end, R_g, which determines the reflection coefficient, Γ_g, at that location. The current reflection diagram is derived from the voltage diagram in the usual way. There is no initial current to consider.
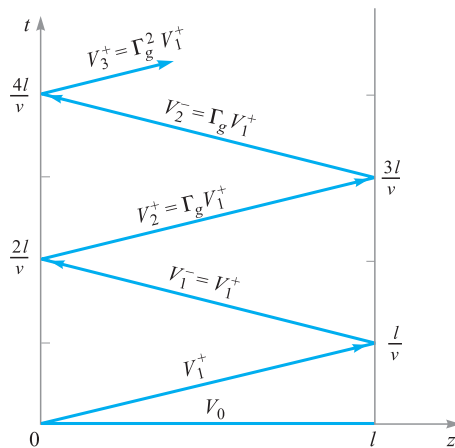
Figure 10.26 Voltage reflection diagram for the charged line of Figure 10.25, showing the initial condition of V_0 everywhere on the line at $t = 0$.
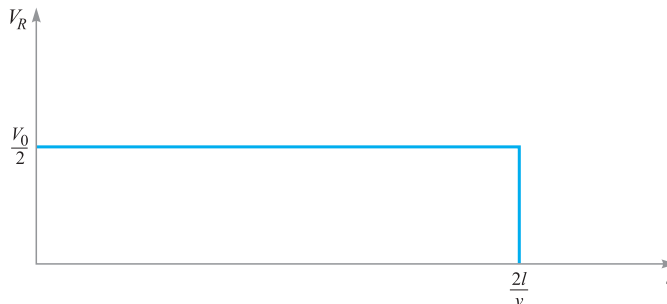
Figure 10.27   Voltage across the resistor as a function of time, as
determined from the reflection diagram of Figure 10.26, in which
$R_g = Z_0$ ($\Gamma = 0$).

A special case of practical importance is that in which the resistor is matched
to the line, or $R_g = Z_0$. In this case, Eq. (122) gives $V_1^+ = -V_0/2$. The line fully
discharges in one round trip of V_1^+ and produces a voltage across the resistor of value
$V_R = V_0/2$, which persists for time $T = 2l/v$. The resistor voltage as a function
of time is shown in Figure 10.27. The transmission line in this application is known
as a *pulse-forming line*; pulses that are generated in this way are well formed and
of low noise, provided the switch is sufficiently fast. Commercial units are available
that are capable of generating high-voltage pulses of widths on the order of a few
nanoseconds, using thyratron-based switches.

When the resistor is not matched to the line, full discharge still occurs, but does
so over several reflections, leading to a complicated pulse shape.

EXAMPLE 10.12

In the charged line of Figure 10.25, the characteristic impedance is $Z_0 = 100 \, \Omega$, and
$R_g = 100/3 \, \Omega$. The line is charged to an initial voltage, $V_0 = 160$ V, and the switch is
closed at time $t = 0$. Determine and plot the voltage and current through the resistor
for time $0 < t < 8l/v$ (four round trips).

Solution. With the given values of R_g and Z_0, Eq. (47) gives $\Gamma_g = -1/2$. Then,
with $\Gamma_L = 1$, and using (122), we find

$$V_1^+ = V_1^- = -3/4V_0 = -120 \text{ V}$$
$$V_2^+ = V_2^- = \Gamma_g V_1^- = +60 \text{ V}$$
$$V_3^+ = V_3^- = \Gamma_g V_2^- = -30 \text{ V}$$
$$V_4^+ = V_4^- = \Gamma_g V_3^- = +15 \text{ V}$$

Using these values on the voltage reflection diagram, we evaluate the voltage in time
at the resistor location by moving up the left-hand vertical axis, adding voltages as
we progress, and beginning with $V_0 + V_1^+$ at $t = 0$. Note that when we add voltages
along the vertical axis, we are encountering the intersection points between incident

and reflected waves, which occur (in time) at each integer multiple of $2l/v$. So, when moving up the axis, we add the voltages of *both* waves to our total at each occurrence. The voltage within each time interval is thus:

$$
\begin{aligned}
V_R &= V_0 + V_1^+ = 40 \text{ V} & (0 < t < 2l/v) \\
&= V_0 + V_1^+ + V_1^- + V_2^+ = -20 \text{ V} & (2l/v < t < 4l/v) \\
&= V_0 + V_1^+ + V_1^- + V_2^+ + V_2^- + V_3^+ = 10 \text{ V} & (4l/v < t < 6l/v) \\
&= V_0 + V_1^+ + V_1^- + V_2^+ + V_2^- + V_3^+ + V_3^- + V_4^+ = -5 \text{ V} & (6l/v < t < 8l/v)
\end{aligned}
$$

The resulting voltage plot over the desired time range is shown in Figure 10.28*a*.
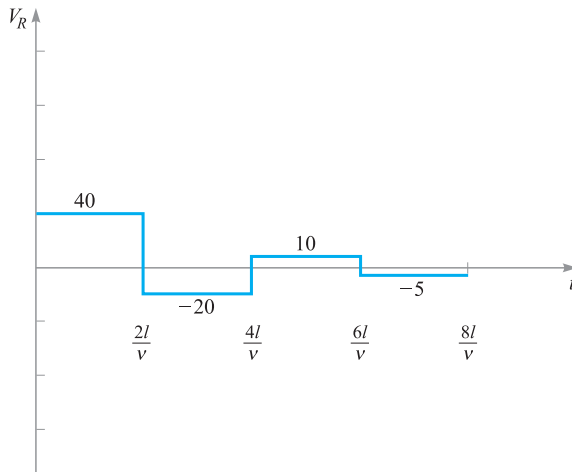
(a)
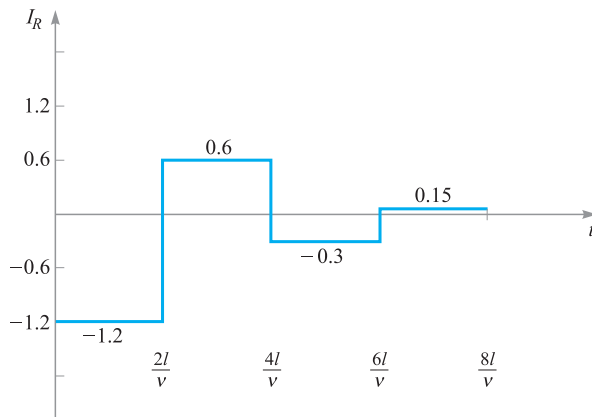
(b)

Figure 10.28 Resistor voltage (*a*) and current (*b*) as functions of time for the line of Figure 10.25, with values as specified in Example 10.12.

The current through the resistor is most easily obtained by dividing the voltages in Figure 10.28a by $-R_g$. As a demonstration, we can also use the current diagram of Figure 10.22a to obtain this result. Using (120) and (121), we evaluate the current waves as follows:

$$I_1^+ = V_1^+/Z_0 = -1.2 \text{ A}$$
$$I_1^- = -V_1^-/Z_0 = +1.2 \text{ A}$$
$$I_2^+ = -I_2^- = V_2^+/Z_0 = +0.6 \text{ A}$$
$$I_3^+ = -I_3^- = V_3^+/Z_0 = -0.30 \text{ A}$$
$$I_4^+ = -I_4^- = V_4^+/Z_0 = +0.15 \text{ A}$$

Using these values on the current reflection diagram, Figure 10.22a, we add up currents in the resistor in time by moving up the left-hand axis, as we did with the voltage diagram. The result is shown in Figure 10.28b. As a further check to the correctness of our diagram construction, we note that current at the open end of the line $(Z = l)$ must always be zero. Therefore, summing currents up the right-hand axis must give a zero result for all time. The reader is encouraged to verify this.

REFERENCES

1.   White, H. J., P. R. Gillette, and J. V. Lebacqz. "The Pulse-Forming Network." Chapter 6 in *Pulse Generators*, edited by G. N, Glasoe and J. V. Lebacqz. New York: Dover, 1965.

2.   Brown, R. G., R. A. Sharpe, W. L. Hughes, and R. E. Post. *Lines, Waves, and Antennas*. 2d ed. New York: The Ronald Press Company, 1973. Transmission lines are covered in the first six chapters, with numerous examples.

3.   Cheng, D. K. *Field and Wave Electromagnetics*. 2d ed. Reading, Mass.: Addison-Wesley, 1989. Provides numerous examples of Smith chart problems and transients.

4.   Seshadri, S. R. *Fundamentals of Transmission Lines and Electromagnetic Fields*. Reading, Mass.: Addison-Wesley, 1971.

CHAPTER 10   PROBLEMS

Quizzes

10.1   The parameters of a certain transmission line operating at $\omega = 6 \times 10^8$ rad/s are $L = 0.35 \,\mu\text{H/m}$, $C = 40$ pF/m, $G = 75 \,\mu\text{S/m}$, and $R = 17 \,\Omega/\text{m}$. Find $\gamma, \alpha, \beta, \lambda$, and Z_0.

10.2   A sinusoidal wave on a transmission line is specified by voltage and current in phasor form:

$$V_s(z) = V_0 \, e^{\alpha z} \, e^{j\beta z} \quad \text{and} \quad I_s(z) = I_0 \, e^{\alpha z} \, e^{j\beta z} \, e^{j\phi}$$

where V_0 and I_0 are both real. (a) In which direction does this wave propagate and why? (b) It is found that $\alpha = 0$, $Z_0 = 50 \,\Omega$, and the wave velocity is $v_p = 2.5 \times 10^8$ m/s, with $\omega = 10^8$ s^{-1}. Evaluate R, G, L, C, λ, and ϕ.

10.3 The characteristic impedance of a certain lossless transmission line is 72 Ω. If $L = 0.5\,\mu\text{H/m}$, find (a) C; (b) v_p; (c) β if $f = 80$ MHz. (d) The line is terminated with a load of 60 Ω. Find Γ and s.

10.4 A sinusoidal voltage wave of amplitude V_0, frequency ω, and phase constant β propagates in the forward z direction toward the open load end in a lossless transmission line of characteristic impedance Z_0. At the end, the wave totally reflects with zero phase shift, and the reflected wave now interferes with the incident wave to yield a standing wave pattern over the line length (as per Example 10.1). Determine the standing wave pattern for the *current* in the line. Express the result in real instantaneous form and simplify.

10.5 Two characteristics of a certain lossless transmission line are $Z_0 = 50\,\Omega$ and $\gamma = 0 + j0.2\pi\ \text{m}^{-1}$ at $f = 60$ MHz (a) find L and C for the line. (b) A load $Z_L = 60 + j80\,\Omega$ is located at $z = 0$. What is the shortest distance from the load to a point at which $Z_{\text{in}} = R_{\text{in}} + j0$?

10.6 A 50-Ω load is attached to a 50-m section of the transmission line of Problem 10.1, and a 100-W signal is fed to the input end of the line. (a) Evaluate the distributed line loss in dB/m. (b) Evaluate the reflection coefficient at the load. (c) Evaluate the power that is dissipated by the load resistor. (d) What power drop in dB does the dissipated power in the load represent when compared to the original input power? (e) On partial reflection from the load, how much power returns to the input and what dB drop does this represent when compared to the original 100-W input power?

10.7 A transmitter and receiver are connected using a cascaded pair of transmission lines. At the operating frequency, line 1 has a measured loss of 0.1 dB/m, and line 2 is rated at 0.2 dB/m. The link is composed of 40 m of line 1 joined to 25 m of line 2. At the joint, a splice loss of 2 dB is measured. If the transmitted power is 100 mW, what is the received power?

10.8 An absolute measure of power is the dBm scale, in which power is specified in decibels relative to one milliwatt. Specifically, $P(\text{dBm}) = 10\log_{10}[P(\text{mW})/1\,\text{mW}]$. Suppose that a receiver is rated as having a *sensitivity* of -20 dBm, indicating the *mimimum* power that it must receive in order to adequately interpret the transmitted electronic data. Suppose this receiver is at the load end of a 50-Ω transmission line having 100-m length and loss rating of 0.09 dB/m. The receiver impedance is 75 Ω, and so is not matched to the line. What is the minimum required input power to the line in (a) dBm, (b) mW?

10.9 A sinusoidal voltage source drives the series combination of an impedance, $Z_g = 50 - j50\,\Omega$, and a lossless transmission line of length L, shorted at the load end. The line characteristic impedance is 50 Ω, and wavelength λ is measured on the line. (a) Determine, in terms of wavelength, the shortest line length that will result in the voltage source driving a total impedance of

50 Ω. (*b*) Will other line lengths meet the requirements of part (*a*)? If so, what are they?

10.10 Two lossless transmission lines having different characteristic impedances are to be joined end to end. The impedances are $Z_{01} = 100 \ \Omega$ and $Z_{03} = 25 \ \Omega$. The operating frequency is 1 GHz. (*a*) Find the required characteristic impedance, Z_{02}, of a quarter-wave section to be inserted between the two, which will impedance-match the joint, thus allowing total power transmission through the three lines. (*b*) The capacitance per unit length of the intermediate line is found to be 100 pF/m. Find the shortest length in meters of this line that is needed to satisfy the impedance-matching condition. (*c*) With the three-segment setup as found in parts (*a*) and (*b*), the frequency is now doubled to 2 GHz. Find the input impedance at the line-1-to-line-2 junction, seen by waves incident from line 1. (*d*) Under the conditions of part (*c*), and with power incident from line 1, evaluate the standing wave ratio that will be measured in line 1, and the fraction of the incident power from line 1 that is reflected and propagates back to the line 1 input.

10.11 A transmission line having primary constants L, C, R, and G has length ℓ and is terminated by a load having complex impedance $R_L + jX_L$. At the input end of the line, a dc voltage source, V_0, is connected. Assuming all parameters are known at zero frequency, find the steady-state power dissipated by the load if (*a*) $R = G = 0$; (*b*) $R \neq 0, G = 0$; (*c*) $R = 0$, $G \neq 0$; (*d*) $R \neq 0, G \neq 0$.

10.12 In a circuit in which a sinusoidal voltage source drives its internal impedance in series with a load impedance, it is known that maximum power transfer to the load occurs when the source and load impedances form a complex conjugate pair. Suppose the source (with its internal impedance) now drives a complex load of impedance $Z_L = R_L + jX_L$ that has been moved to the end of a lossless transmission line of length ℓ having characteristic impedance Z_0. If the source impedance is $Z_g = R_g + jX_g$, write an equation that can be solved for the required line length, ℓ, such that the displaced load will receive the maximum power.

10.13 The incident voltage wave on a certain lossless transmission line for which $Z_0 = 50 \ \Omega$ and $v_p = 2 \times 10^8$ m/s is $V^+(z, t) = 200 \cos(\omega t - \pi z)$ V. (*a*) Find ω. (*b*) Find $I^+(z, t)$. The section of line for which $z > 0$ is replaced by a load $Z_L = 50 + j30 \ \Omega$ at $z = 0$. Find: (*c*) Γ_L; (*d*) $V_s^-(z)$; (*e*) V_s at $z = -2.2$ m.

10.14 A lossless transmission line having characteristic impedance $Z_0 = 50 \ \Omega$ is driven by a source at the input end that consists of the series combination of a 10-V sinusoidal generator and a 50-Ω resistor. The line is one-quarter wavelength long. At the other end of the line, a load impedance, $Z_L = 50 - j50 \ \Omega$ is attached. (*a*) Evaluate the input impedance to the line
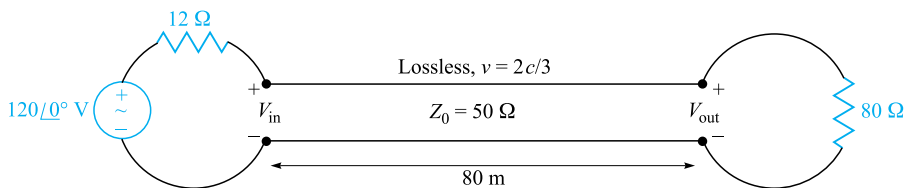
Figure 10.29 See Problem 10.15.

seen by the voltage source-resistor combination; (b) evaluate the power that is dissipated by the load; (c) evaluate the voltage amplitude that appears across the load.

10.15 For the transmission line represented in Figure 10.29, find $V_{s,\text{out}}$ if $f =$ (a) 60 Hz; (b) 500 kHz.

10.16 A 100-Ω lossless transmission line is connected to a second line of 40-Ω impedance, whose length is $\lambda/4$. The other end of the short line is terminated by a 25-Ω resistor. A sinusoidal wave (of frequency f) having 50 W average power is incident from the 100-Ω line. (a) Evaluate the input impedance to the quarter-wave line. (b) Determine the steady-state power that is dissipated by the resistor. (c) Now suppose that the operating frequency is lowered to one-half its original value. Determine the new input impedance, Z'_{in}, for this case. (d) For the new frequency, calculate the power in watts that returns to the input end of the line after reflection.

10.17 Determine the average power absorbed by each resistor in Figure 10.30.

10.18 The line shown in Figure 10.31 is lossless. Find s on both sections 1 and 2.

10.19 A lossless transmission line is 50 cm in length and operates at a frequency of 100 MHz. The line parameters are $L = 0.2\,\mu$H/m and $C = 80$ pF/m. The line is terminated in a short circuit at $z = 0$, and there is a load $Z_L = 50 + j20\,\Omega$ across the line at location $z = -20$ cm. What average power is delivered to Z_L if the input voltage is $100\angle 0°$ V?

10.20 (a) Determine s on the transmission line of Figure 10.32. Note that the dielectric is air. (b) Find the input impedance. (c) If $\omega L = 10\,\Omega$, find I_s. (d) What value of L will produce a maximum value for $|I_s|$ at $\omega = 1$
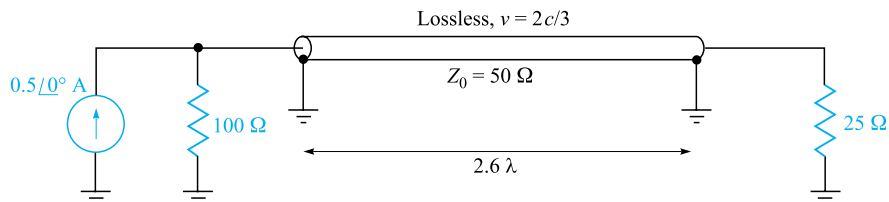
Figure 10.30 See Problem 10.17.
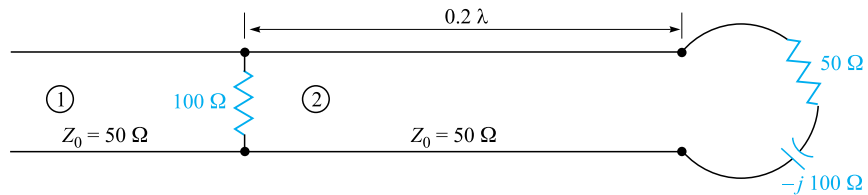
Figure 10.31   See Problem 10.18.

Grad/s? For this value of L, calculate the average power (e) supplied by the source; (f) delivered to $Z_L = 40 + j30$ Ω.

10.21 A lossless line having an air dielectric has a characteristic impedance of 400 Ω. The line is operating at 200 MHz and $Z_{in} = 200 - j200$ Ω. Use analytic methods or the Smith chart (or both) to find (a) s; (b) Z_L, if the line is 1 m long; (c) the distance from the load to the nearest voltage maximum.

10.22 A lossless 75-Ω line is terminated by an unknown load impedance. A VSWR of 10 is measured, and the first voltage minimum occurs at 0.15 wavelengths in front of the load. Using the Smith chart, find (a) the load impedance; (b) the magnitude and phase of the reflection coefficient; (c) the shortest length of line necessary to achieve an entirely resistive input impedance.

10.23 The normalized load on a lossless transmission line is $2 + j1$. Let $\lambda = 20$ m and make use of the Smith chart to find (a) the shortest distance from the load to a point at which $z_{in} = r_{in} + j0$, where $r_{in} > 0$; (b) z_{in} at this point. (c) The line is cut at this point and the portion containing z_L is thrown away. A resistor $r = r_{in}$ of part (a) is connected across the line. What is s on the remainder of the line? (d) What is the shortest distance from this resistor to a point at which $z_{in} = 2 + j1$?

10.24 With the aid of the Smith chart, plot a curve of $|Z_{in}|$ versus l for the transmission line shown in Figure 10.33. Cover the range $0 < l/\lambda < 0.25$.

10.25 A 300-Ω transmission line is short-circuited at $z = 0$. A voltage maximum, $|V|_{max} = 10$ V, is found at $z = -25$ cm, and the minimum voltage, $|V|_{min} = 0$, is at $z = -50$ cm. Use the Smith chart to find Z_L (with the short circuit
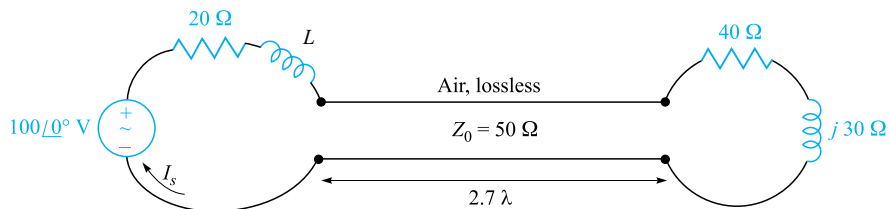
Figure 10.32   See Problem 10.20.
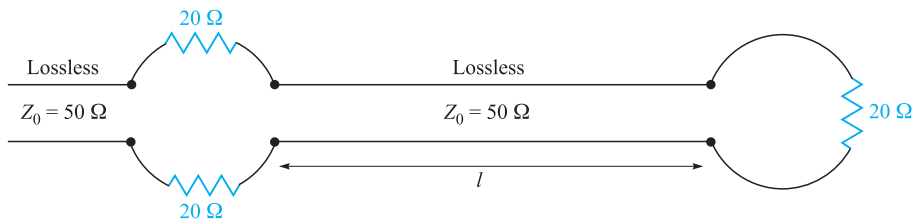
Figure 10.33 See Problem 10.24.

replaced by the load) if the voltage readings are (a) $|V|_{max} = 12$ V at $z = -5$ cm, and $|V|_{min} = 5$ V; (b) $|V|_{max} = 17$ V at $z = -20$ cm, and $|V|_{min} = 0$.

10.26 A 50-Ω lossless line is of length 1.1 λ. It is terminated by an unknown load impedance. The input end of the 50-Ω line is attached to the load end of a lossless 75-Ω line. A VSWR of 4 is measured on the 75-Ω line, on which the first voltage maximum occurs at a distance of 0.2 λ in front of the junction between the two lines. Use the Smith chart to find the unknown load impedance.

10.27 The characteristic admittance ($Y_0 = 1/Z_0$) of a lossless transmission line is 20 mS. The line is terminated in a load $Y_L = 40 - j20$ mS. Use the Smith chart to find (a) s; (b) Y_{in} if $l = 0.15\lambda$; (c) the distance in wavelengths from Y_L to the nearest voltage maximum.

10.28 The wavelength on a certain lossless line is 10 cm. If the normalized input impedance is $z_{in} = 1 + j2$, use the Smith chart to determine (a) s; (b) z_L, if the length of the line is 12 cm; (c) x_L, if $z_L = 2 + jx_L$ where $x_L > 0$.

10.29 A standing wave ratio of 2.5 exists on a lossless 60 Ω line. Probe measurements locate a voltage minimum on the line whose location is marked by a small scratch on the line. When the load is replaced by a short circuit, the minima are 25 cm apart, and one minimum is located at a point 7 cm toward the source from the scratch. Find Z_L.

10.30 A two-wire line constructed of lossless wire of circular cross section is gradually flared into a coupling loop that looks like an egg beater. At the point X, indicated by the arrow in Figure 10.34, a short circuit is placed
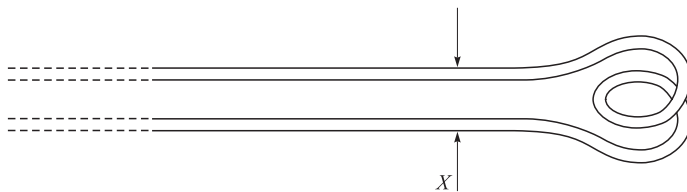
Figure 10.34 See Problem 10.30.

across the line. A probe is moved along the line and indicates that the first voltage minimum to the left of X is 16 cm from X. With the short circuit removed, a voltage minimum is found 5 cm to the left of X, and a voltage maximum is located that is 3 times the voltage of the minimum. Use the Smith chart to determine (a) f; (b) s; (c) the normalized input impedance of the egg beater as seen looking to the right at point X.

10.31 In order to compare the relative sharpness of the maxima and minima of a standing wave, assume a load $z_L = 4 + j0$ is located at $z = 0$. Let $|V|_{min} = 1$ and $\lambda = 1$ m. Determine the width of the (a) minimum where $|V| < 1.1$; (b) maximum where $|V| > 4/1.1$.

10.32 In Figure 10.17, let $Z_L = 250\ \Omega$, $Z_0 = 50\ \Omega$, find the shortest attachment distance d and the shortest length d_1 of a short-circuited stub line that will provide a perfect match on the main line to the left of the stub. Express all answers in wavelengths.

10.33 In Figure 10.17, let $Z_L = 40 - j10\ \Omega$, $Z_0 = 50\ \Omega$, $f = 800$ MHz, and $v = c$. (a) Find the shortest length d_1 of a short-circuited stub, and the shortest distance d that it may be located from the load to provide a perfect match on the main line to the left of the stub. (b) Repeat for an open-circuited stub.

10.34 The lossless line shown in Figure 10.35 is operating with $\lambda = 100$ cm. If $d_1 = 10$ cm, $d = 25$ cm, and the line is matched to the left of the stub, what is Z_L?

10.35 A load, $Z_L = 25 + j75\ \Omega$, is located at $z = 0$ on a lossless two-wire line for which $Z_0 = 50\ \Omega$ and $v = c$. (a) If $f = 300$ MHz, find the shortest distance d ($z = -d$) at which the input admittance has a real part equal to $1/Z_0$ and a negative imaginary part. (b) What value of capacitance C should be connected across the line at that point to provide unity standing wave ratio on the remaining portion of the line?

10.36 The two-wire lines shown in Figure 10.36 are all lossless and have $Z_0 = 200\ \Omega$. Find d and the shortest possible value for d_1 to provide a matched load if $\lambda = 100$ cm.
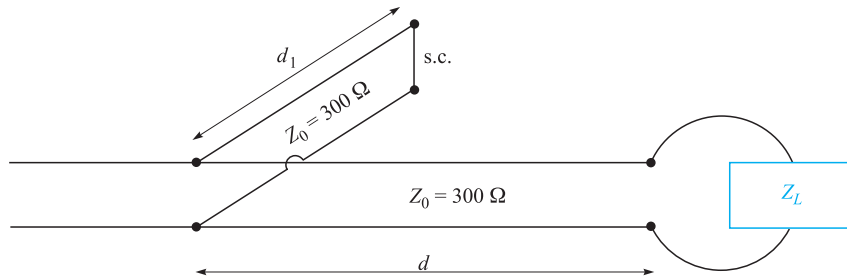
Figure 10.35 See Problem 10.34.
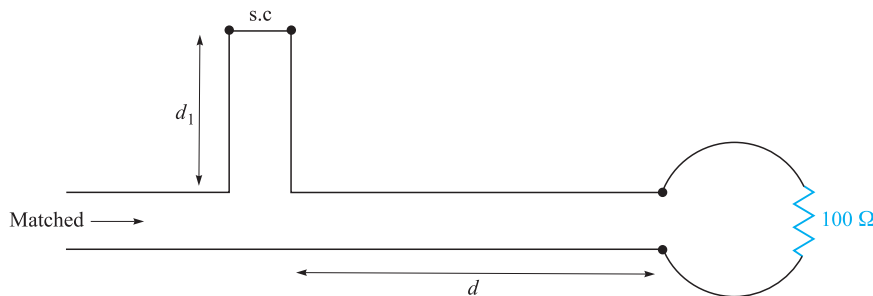
Figure 10.36 See Problem 10.36.

10.37 In the transmission line of Figure 10.20, $R_g = Z_0 = 50\ \Omega$, and $R_L = 25\ \Omega$. Determine and plot the voltage at the load resistor and the current in the battery as functions of time by constructing appropriate voltage and current reflection diagrams.

10.38 Repeat Problem 10.37, with $Z_0 = 50\ \Omega$, and $R_L = R_g = 25\ \Omega$. Carry out the analysis for the time period $0 < t < 8l/v$.

10.39 In the transmission line of Figure 10.20, $Z_0 = 50\ \Omega$, and $R_L = R_g = 25\ \Omega$. The switch is closed at $t = 0$ and *is opened again* at time $t = l/4v$, thus creating a rectangular voltage *pulse* in the line. Construct an appropriate voltage reflection diagram for this case and use it to make a plot of the voltage at the load resistor as a function of time for $0 < t < 8l/v$ (note that the effect of opening the switch is to initiate a second voltage wave, whose value is such that it leaves a net current of zero in its wake).

10.40 In the charged line of Figure 10.25, the characteristic impedance is $Z_0 = 100\ \Omega$, and $R_g = 300\ \Omega$. The line is charged to initial voltage, $V_0 = 160$ V, and the switch is closed at $t = 0$. Determine and plot the voltage and current through the resistor for time $0 < t < 8l/v$ (four round-trips). This problem accompanies Example 10.12 as the other special case of the basic charged-line problem, in which now $R_g > Z_0$.

10.41 In the transmission line of Figure 10.37, the switch is located *midway* down the line and is closed at $t = 0$. Construct a voltage reflection diagram for this case, where $R_L = Z_0$. Plot the load resistor voltage as a function of time.
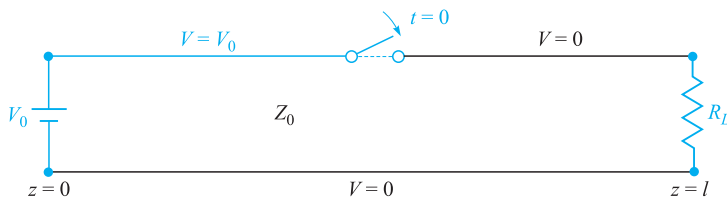
Figure 10.37 See Problem 10.41.
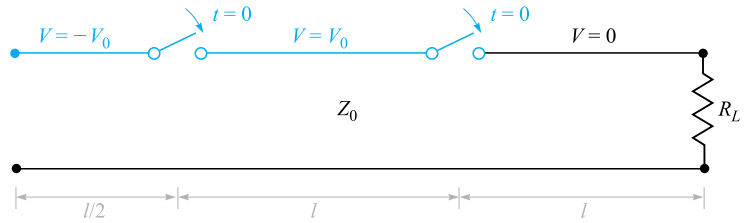
Figure 10.38 See Problem 10.42.
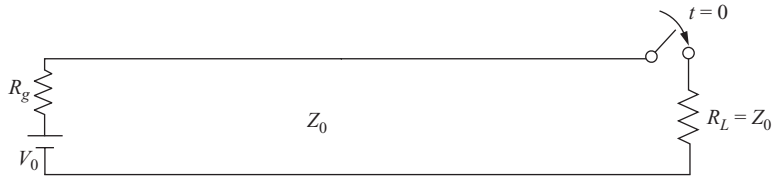
Figure 10.39 See Problem 10.43.

10.42 A simple *frozen wave generator* is shown in Figure 10.38. Both switches are closed simultaneously at $t = 0$. Construct an appropriate voltage reflection diagram for the case in which $R_L = Z_0$. Determine and plot the load resistor voltage as a function of time.

10.43 In Figure 10.39, $R_L = Z_0$ and $R_g = Z_0/3$. The switch is closed at $t = 0$. Determine and plot as functions of time (*a*) the voltage across R_L; (*b*) the voltage across R_g; (*c*) the current through the battery.

The Uniform Plane Wave

This chapter is concerned with the application of Maxwell's equations to the problem of electromagnetic wave propagation. The uniform plane wave represents the simplest case, and while it is appropriate for an introduction, it is of great practical importance. Waves encountered in practice can often be assumed to be of this form. In this study, we will explore the basic principles of electromagnetic wave propagation, and we will come to understand the physical processes that determine the speed of propagation and the extent to which attenuation may occur. We will derive and use the Poynting theorem to find the power carried by a wave. Finally, we will learn how to describe wave polarization. ∎

11.1 WAVE PROPAGATION IN FREE SPACE

We begin with a quick study of Maxwell's equations, in which we look for clues of wave phenomena. In Chapter 10, we saw how voltages and currents propagate as waves in transmission lines, and we know that the existence of voltages and currents implies the existence of electric and magnetic fields. So we can identify a transmission line as a structure that confines the fields while enabling them to travel along its length as waves. It can be argued that it is the fields that generate the voltage and current in a transmission line wave, and—if there is no structure on which the voltage and current can exist—the fields will exist nevertheless, and will propagate. In free space, the fields are not bounded by any confining structure, and so they may assume *any* magnitude and direction, as initially determined by the device (such as an antenna) that generates them.

When considering electromagnetic waves in free space, we note that the medium is *sourceless* ($\rho_v = \mathbf{J} = 0$). Under these conditions, Maxwell's equations may be

written in terms of **E** and **H** only as

$$\nabla \times \mathbf{H} = \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \qquad (1)$$

$$\nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t} \qquad (2)$$

$$\nabla \cdot \mathbf{E} = 0 \qquad (3)$$

$$\nabla \cdot \mathbf{H} = 0 \qquad (4)$$

Now let us see whether wave motion can be inferred from these four equations without actually solving them. Equation (1) states that if electric field **E** is changing with time at some point, then magnetic field **H** has curl at that point; therefore **H** varies spatially in a direction normal to its orientation direction. Also, if **E** is changing with time, then **H** will in general also change with time, although not necessarily in the same way. Next, we see from Eq. (2) that a time-varying **H** generates **E**, which, having curl, varies spatially in the direction normal to its orientation. We now have once more a changing electric field, our original hypothesis, but this field is present a small distance away from the point of the original disturbance. We might guess (correctly) that the velocity with which the effect moves away from the original point is the velocity of light, but this must be checked by a more detailed examination of Maxwell's equations.

We postulate the existence of a *uniform plane wave,* in which both fields, **E** and **H**, lie in the *transverse plane*—that is, the plane whose normal is the direction of propagation. Furthermore, and by definition, both fields are of constant magnitude in the transverse plane. For this reason, such a wave is sometimes called a *transverse electromagnetic* (TEM) wave. The required spatial variation of both fields in the direction normal to their orientations will therefore occur only in the direction of travel—or normal to the transverse plane. Assume, for example, that $\mathbf{E} = E_x \mathbf{a}_x$, or that the electric field is *polarized* in the x direction. If we further assume that wave travel is in the z direction, we allow spatial variation of **E** only with z. Using Eq. (2), we note that with these restrictions, the curl of **E** reduces to a single term:

$$\nabla \times \mathbf{E} = \frac{\partial E_x}{\partial z} \mathbf{a}_y = -\mu_0 \frac{\partial \mathbf{H}}{\partial t} = -\mu_0 \frac{\partial H_y}{\partial t} \mathbf{a}_y \qquad (5)$$

The direction of the curl of **E** in (5) determines the direction of **H**, which we observe to be along the y direction. Therefore, in a uniform plane wave, the directions of **E** and **H** and the direction of travel are mutually orthogonal. Using the y-directed magnetic field, and the fact that it varies only in z, simplifies Eq. (1) to read

$$\nabla \times \mathbf{H} = -\frac{\partial H_y}{\partial z} \mathbf{a}_x = \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} = \epsilon_0 \frac{\partial E_x}{\partial t} \mathbf{a}_x \qquad (6)$$

Equations (5) and (6) can be more succinctly written:

$$\frac{\partial E_x}{\partial z} = -\mu_0 \frac{\partial H_y}{\partial t} \tag{7}$$

$$\frac{\partial H_y}{\partial z} = -\epsilon_0 \frac{\partial E_x}{\partial t} \tag{8}$$

These equations compare directly with the telegraphist's equations for the lossless transmission line [Eqs. (20) and (21) in Chapter 10]. Further manipulations of (7) and (8) proceed in the same manner as was done with the telegraphist's equations. Specifically, we differentiate (7) with respect to z, obtaining:

$$\frac{\partial^2 E_x}{\partial z^2} = -\mu_0 \frac{\partial^2 H_y}{\partial t \partial z} \tag{9}$$

Then, (8) is differentiated with respect to t:

$$\frac{\partial^2 H_y}{\partial z \partial t} = -\epsilon_0 \frac{\partial^2 E_x}{\partial t^2} \tag{10}$$

Substituting (10) into (9) results in

$$\frac{\partial^2 E_x}{\partial z^2} = \mu_0 \epsilon_0 \frac{\partial^2 E_x}{\partial t^2} \tag{11}$$

This equation, in direct analogy to Eq. (13) in Chapter 10, we identify as the wave equation for our x-polarized TEM electric field in free space. From Eq. (11), we further identify the propagation velocity:

$$v = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = 3 \times 10^8 \text{ m/s} = c \tag{12}$$

where c denotes the velocity of light in free space. A similar procedure, involving differentiating (7) with t and (8) with z, yields the wave equation for the magnetic field; it is identical in form to (11):

$$\frac{\partial^2 H_y}{\partial z^2} = \mu_0 \epsilon_0 \frac{\partial^2 H_y}{\partial t^2} \tag{13}$$

As was discussed in Chapter 10, the solution to equations of the form of (11) and (13) will be forward- and backward-propagating waves having the general form [in this case for Eq. (11)]:

$$E_x(z, t) = f_1(t - z/v) + f_2(t + z/v) \tag{14}$$

where again f_1 amd f_2 can be any function whose argument is of the form $t \pm z/v$.

From here, we immediately specialize to sinusoidal functions of a specified frequency and write the solution to (11) in the form of forward- and backward-propagating

Animations

cosines. Because the waves are sinusoidal, we denote their velocity as the *phase velocity, v_p*. The waves are written as:

$$E_x(z, t) = \mathcal{E}_x(z, t) + \mathcal{E}'_x(z, t)$$

$$= |E_{x0}| \cos[\omega(t - z/v_p) + \phi_1] + |E'_{x0}| \cos[\omega(t + z/v_p) + \phi_2]$$

$$= \underbrace{|E_{x0}| \cos[\omega t - k_0 z + \phi_1]}_{\text{forward } z \text{ travel}} + \underbrace{|E'_{x0}| \cos[\omega t + k_0 z + \phi_2]}_{\text{backward } z \text{ travel}} \qquad (15)$$

In writing the second line of (15), we have used the fact that the waves are traveling in free space, in which case the phase velocity, $v_p = c$. Additionally, the *wavenumber* in free space in defined as

$$k_0 \equiv \frac{\omega}{c} \text{ rad/m} \qquad (16)$$

In a manner consistant with our transmission line studies, we refer to the solutions expressed in (15) as the *real instantaneous* forms of the electric field. They are the mathematical representations of what one would experimentally measure. The terms ωt and $k_0 z$, appearing in (15), have units of angle and are usually expressed in radians. We know that ω is the radian time frequency, measuring phase shift *per unit time;* it has units of *rad/s.* In a similar way, we see that k_0 will be interpreted as a *spatial frequency*, which in the present case measures the phase shift *per unit distance* along the z direction in rad/m. We note that k_0 is the phase constant for lossless propagation of uniform plane waves in free space. The *wavelength* in free space is the distance over which the spatial phase shifts by 2π radians, assuming fixed time, or

$$k_0 z = k_0 \lambda = 2\pi \quad \rightarrow \quad \lambda = \frac{2\pi}{k_0} \quad \text{(free space)} \qquad (17)$$

The manner in which the waves propagate is the same as we encountered in transmission lines. Specifically, suppose we consider some point (such as a wave crest) on the forward-propagating cosine function of Eq. (15). For a crest to occur, the argument of the cosine must be an integer multiple of 2π. Considering the mth crest of the wave, the condition becomes

$$k_0 z = 2m\pi$$

So let us now consider the point on the cosine that we have chosen, and see what happens as time is allowed to increase. Our requirement is that the entire cosine argument be the same multiple of 2π for all time, in order to keep track of the chosen point. Our condition becomes

$$\omega t - k_0 z = \omega(t - z/c) = 2m\pi \qquad (18)$$

As time increases, the position z must also increase in order to satisfy (18). The wave crest (and the entire wave) moves in the positive z direction at phase velocity c (in free space). Using similar reasoning, the wave in Eq. (15) having cosine argument $(\omega t + k_0 z)$ describes a wave that moves in the negative z direction, since as time

increases, z must now decrease to keep the argument constant. For simplicity, we will restrict our attention in this chapter to only the positive z traveling wave.

As was done for transmission line waves, we express the real instantaneous fields of Eq. (15) in terms of their phasor forms. Using the forward-propagating field in (15), we write:

$$\mathcal{E}_x(z, t) = \underbrace{\frac{1}{2}|E_{x0}|e^{j\phi_1}}_{E_{x0}} e^{-jk_0 z} e^{j\omega t} + c.c. = \frac{1}{2}E_{xs}e^{j\omega t} + c.c. = \text{Re}[E_{xs}e^{j\omega t}] \quad (19)$$

where $c.c.$ denotes the complex conjugate, and where we identify the *phasor electric field* as $E_{xs} = E_{x0}e^{-jk_0 z}$. As indicated in (19), E_{x0} is the *complex* amplitude (which includes the phase, ϕ_1).

EXAMPLE 11.1

Let us express $\mathcal{E}_y(z, t) = 100 \cos(10^8 t - 0.5z + 30°)$ V/m as a phasor.

Solution. We first go to exponential notation,

$$\mathcal{E}_y(z, t) = \text{Re}\left[100 e^{j(10^8 t - 0.5z + 30°)}\right]$$

and then drop Re and suppress $e^{j10^8 t}$, obtaining the phasor

$$E_{ys}(z) = 100 e^{-j0.5z + j30°}$$

Note that a mixed nomenclature is used for the angle in this case; that is, $0.5z$ is in radians, while $30°$ is in degrees. Given a scalar component or a vector expressed as a phasor, we may easily recover the time-domain expression.

EXAMPLE 11.2

Given the complex amplitude of the electric field of a uniform plane wave, $\mathbf{E}_0 = 100\mathbf{a}_x + 20\angle 30°\mathbf{a}_y$ V/m, construct the phasor and real instantaneous fields if the wave is known to propagate in the forward z direction in free space and has frequency of 10 MHz.

Solution. We begin by constructing the general phasor expression:

$$\mathbf{E}_s(z) = \left[100\mathbf{a}_x + 20e^{j30°}\mathbf{a}_y\right]e^{-jk_0 z}$$

where $k_0 = \omega/c = 2\pi \times 10^7/3 \times 10^8 = 0.21$ rad/m. The real instantaneous form is then found through the rule expressed in Eq. (19):

$$\begin{aligned}
\mathcal{E}(z, t) &= \text{Re}\left[100 e^{-j0.21z} e^{j2\pi \times 10^7 t}\mathbf{a}_x + 20e^{j30°}e^{-j0.21z}e^{j2\pi \times 10^7 t}\mathbf{a}_y\right] \\
&= \text{Re}\left[100 e^{j(2\pi \times 10^7 t - 0.21z)}\mathbf{a}_x + 20e^{j(2\pi \times 10^7 t - 0.21z + 30°)}\mathbf{a}_y\right] \\
&= 100\cos\left(2\pi \times 10^7 t - 0.21z\right)\mathbf{a}_x + 20\cos\left(2\pi \times 10^7 t - 0.21z + 30°\right)\mathbf{a}_y
\end{aligned}$$

It is evident that taking the partial derivative of any field quantity with respect to time is equivalent to multiplying the corresponding phasor by $j\omega$. As an example, we can express Eq. (8) (using sinusoidal fields) as

$$\frac{\partial \mathcal{H}_y}{\partial z} = -\epsilon_0 \frac{\partial \mathcal{E}_x}{\partial t} \tag{20}$$

where, in a manner consistent with (19):

$$\mathcal{E}_x(z, t) = \frac{1}{2} E_{xs}(z) e^{j\omega t} + c.c. \qquad \text{and} \qquad \mathcal{H}_y(z, t) = \frac{1}{2} H_{ys}(z) e^{j\omega t} + c.c. \tag{21}$$

On substituting the fields in (21) into (20), the latter equation simplifies to

$$\frac{d H_{ys}(z)}{dz} = -j\omega\epsilon_0 E_{xs}(z) \tag{22}$$

In obtaining this equation, we note first that the complex conjugate terms in (21) produce their own separate equation, redundant with (22); second, the $e^{j\omega t}$ factors, common to both sides, have divided out; third, the partial derivative with z becomes the total derivative, since the phasor, H_{ys}, depends only on z.

We next apply this result to Maxwell's equations, to obtain them in phasor form. Substituting the field as expressed in (21) into Eqs. (1) through (4) results in

$$\nabla \times \mathbf{H}_s = j\omega\epsilon_0 \mathbf{E}_s \tag{23}$$
$$\nabla \times \mathbf{E}_s = -j\omega\mu_0 \mathbf{H}_s \tag{24}$$
$$\nabla \cdot \mathbf{E}_s = 0 \tag{25}$$
$$\nabla \cdot \mathbf{H}_s = 0 \tag{26}$$

It should be noted that (25) and (26) are no longer independent relationships, for they can be obtained by taking the divergence of (23) and (24), respectively.

Eqs. (23) through (26) may be used to obtain the sinusoidal steady-state vector form of the wave equation in free space. We begin by taking the curl of both sides of (24):

$$\nabla \times \nabla \times \mathbf{E}_s = -j\omega\mu_0 \nabla \times \mathbf{H}_s = \nabla(\nabla \cdot \mathbf{E}_s) - \nabla^2 \mathbf{E}_s \tag{27}$$

where the last equality is an identity, which defines the *vector Laplacian* of $\mathbf{E}_s$:

$$\nabla^2 \mathbf{E}_s = \nabla(\nabla \cdot \mathbf{E}_s) - \nabla \times \nabla \times \mathbf{E}_s$$

From (25), we note that $\nabla \cdot \mathbf{E}_s = 0$. Using this, and substituting (23) in (27), we obtain

$$\nabla^2 \mathbf{E}_s = -k_0^2 \mathbf{E}_s \tag{28}$$

where again, $k_0 = \omega/c = \omega\sqrt{\mu_0\epsilon_0}$. Equation (28) is known as the vector Helmholtz equation in free space.[1] It is fairly formidable when expanded, even in rectangular coordinates, for three scalar phasor equations result (one for each vector component), and each equation has four terms. The x component of (28) becomes, still using the del-operator notation,

$$\nabla^2 E_{xs} = -k_0^2 E_{xs} \qquad (29)$$

and the expansion of the operator leads to the second-order partial differential equation

$$\frac{\partial^2 E_{xs}}{\partial x^2} + \frac{\partial^2 E_{xs}}{\partial y^2} + \frac{\partial^2 E_{xs}}{\partial z^2} = -k_0^2 E_{xs}$$

Again, assuming a uniform plane wave in which E_{xs} does not vary with x or y, the two corresponding derivatives are zero, and we obtain

$$\frac{d^2 E_{xs}}{dz^2} = -k_0^2 E_{xs} \qquad (30)$$

the solution of which we already know:

$$E_{xs}(z) = E_{x0}e^{-jk_0z} + E'_{x0}e^{jk_0z} \qquad (31)$$

Let us now return to Maxwell's equations, (23) through (26), and determine the form of the **H** field. Given $\mathbf{E}_s$, $\mathbf{H}_s$ is most easily obtained from (24):

$$\nabla \times \mathbf{E}_s = -j\omega\mu_0\mathbf{H}_s \qquad (24)$$

which is greatly simplified for a single E_{xs} component varying only with z,

$$\frac{dE_{xs}}{dz} = -j\omega\mu_0 H_{ys}$$

Using (31) for E_{xs}, we have

$$H_{ys} = -\frac{1}{j\omega\mu_0}\left[(-jk_0)E_{x0}e^{-jk_0z} + (jk_0)E'_{x0}e^{jk_0z}\right]$$

$$= E_{x0}\sqrt{\frac{\epsilon_0}{\mu_0}}e^{-jk_0z} - E'_{x0}\sqrt{\frac{\epsilon_0}{\mu_0}}e^{jk_0z} = H_{y0}e^{-jk_0z} + H'_{y0}e^{jk_0z} \qquad (32)$$

In real instantaneous form, this becomes

$$H_y(z,t) = E_{x0}\sqrt{\frac{\epsilon_0}{\mu_0}}\cos(\omega t - k_0z) - E'_{x0}\sqrt{\frac{\epsilon_0}{\mu_0}}\cos(\omega t + k_0z) \qquad (33)$$

where E_{x0} and E'_{x0} are assumed real.

---

[1] Hermann Ludwig Ferdinand von Helmholtz (1821–1894) was a professor at the University of Berlin working in the fields of physiology, electrodynamics, and optics. Hertz was one of his students.

In general, we find from (32) that the electric and magnetic field amplitudes of the forward-propagating wave in free space are related through

$$E_{x0} = \sqrt{\frac{\mu_0}{\epsilon_0}} H_{y0} = \eta_0 H_{y0} \qquad (34a)$$

We also find the backward-propagating wave amplitudes are related through

$$E'_{x0} = -\sqrt{\frac{\mu_0}{\epsilon_0}} H'_{y0} = -\eta_0 H'_{y0} \qquad (34b)$$

where the *intrinsic impedance* of free space is defined as

$$\eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 377 \doteq 120\pi \ \Omega \qquad (35)$$

The dimension of η_0 in ohms is immediately evident from its definition as the ratio of E (in units of V/m) to H (in units of A/m). It is in direct analogy to the characteristic impedance, Z_0, of a transmission line, where we defined the latter as the ratio of voltage to current in a traveling wave. We note that the difference between (34a) and (34b) is a minus sign. This is consistent with the transmission line analogy that led to Eqs. (25a) and (25b) in Chapter 10. Those equations accounted for the definitions of positive and negative current associated with forward and backward voltage waves. In a similar way, Eq. (34a) specifies that in a forward-z propagating uniform plane wave whose electric field vector lies in the positive x direction at a given point in time and space, the magnetic field vector lies in the *positive y* direction at the same space and time coordinates. In the case of a backward-z propagating wave having a positive x-directed electric field, the magnetic field vector lies in the *negative y* direction. The physical significance of this has to do with the definition of power flow in the wave, as specified through the Poynting vector, $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ (in watts/m^2). The cross product of $\mathbf{E}$ with $\mathbf{H}$ must give the correct wave propagation direction, and so the need for the minus sign in (34b) is apparent. Issues relating to power transmission will be addressed in Section 11.3.

Some feeling for the way in which the fields vary in space may be obtained from Figures 11.1a and 11.1b. The electric field intensity in Figure 11.1a is shown at $t = 0$, and the instantaneous value of the field is depicted along three lines, the z axis and arbitrary lines parallel to the z axis in the $x = 0$ and $y = 0$ planes. Since the field is uniform in planes perpendicular to the z axis, the variation along all three of the lines is the same. One complete cycle of the variation occurs in a wavelength, λ. The values of H_y at the same time and positions are shown in Figure 11.1b.

A uniform plane wave cannot exist physically, for it extends to infinity in two dimensions at least and represents an infinite amount of energy. The distant field of a transmitting antenna, however, is essentially a uniform plane wave in some limited region; for example, a radar signal impinging on a distant target is closely a uniform plane wave.
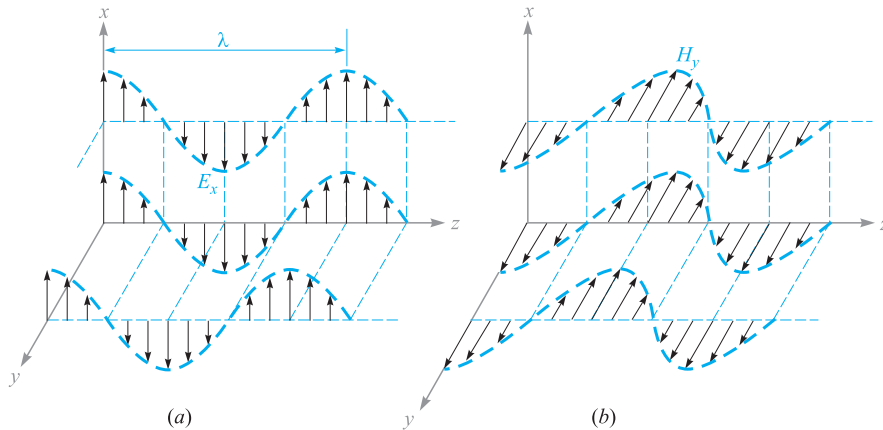
Figure 11.1 (a) Arrows represent the instantaneous values of $E_{x0} \cos[\omega(t - z/c)]$ at $t = 0$ along the z axis, along an arbitrary line in the $x = 0$ plane parallel to the z axis, and along an arbitrary line in the $y = 0$ plane parallel to the z axis. (b) Corresponding values of H_y are indicated. Note that E_x and H_y are in phase at any point in time.

Although we have considered only a wave varying sinusoidally in time and space, a suitable combination of solutions to the wave equation may be made to achieve a wave of any desired form, but which satisfies (14). The summation of an infinite number of harmonics through the use of a Fourier series can produce a periodic wave of square or triangular shape in both space and time. Nonperiodic waves may be obtained from our basic solution by Fourier integral methods. These topics are among those considered in the more advanced books on electromagnetic theory.

Animations

D11.1. The electric field amplitude of a uniform plane wave propagating in the $\mathbf{a}_z$ direction is 250 V/m. If $\mathbf{E} = E_x \mathbf{a}_x$ and $\omega = 1.00$ Mrad/s, find: (a) the frequency; (b) the wavelength; (c) the period; (d) the amplitude of $\mathbf{H}$.

Ans. 159 kHz; 1.88 km; 6.28 μs; 0.663 A/m

D11.2. Let $\mathbf{H}_s = (2\angle-40°\mathbf{a}_x - 3\angle20°\mathbf{a}_y)e^{-j0.07z}$ A/m for a uniform plane wave traveling in free space. Find: (a) ω; (b) H_x at $P(1, 2, 3)$ at $t = 31$ ns; (c) $|\mathbf{H}|$ at $t = 0$ at the origin.

Ans. 21.0 Mrad/s; 1.934 A/m; 3.22 A/m

11.2 WAVE PROPAGATION IN DIELECTRICS

We now extend our analytical treatment of the uniform plane wave to propagation in a dielectric of permittivity ϵ and permeability μ. The medium is assumed to be homogeneous (having constant μ and ϵ with position) and isotropic (in which μ and

ϵ are invariant with field orientation). The Helmholtz equation is

$$\nabla^2 \mathbf{E}_s = -k^2 \mathbf{E}_s \tag{36}$$

where the wavenumber is a function of the material properties, as described by μ and ϵ:

$$k = \omega\sqrt{\mu\epsilon} = k_0\sqrt{\mu_r\epsilon_r} \tag{37}$$

For E_{xs} we have

$$\frac{d^2 E_{xs}}{dz^2} = -k^2 E_{xs} \tag{38}$$

An important feature of wave propagation in a dielectric is that k can be complex-valued, and as such it is referred to as the complex *propagation constant*. A general solution of (38), in fact, allows the possibility of a complex k, and it is customary to write it in terms of its real and imaginary parts in the following way:

$$jk = \alpha + j\beta \tag{39}$$

A solution to (38) will be:

$$E_{xs} = E_{x0}e^{-jkz} = E_{x0}e^{-\alpha z}e^{-j\beta z} \tag{40}$$

Multiplying (40) by $e^{j\omega t}$ and taking the real part yields a form of the field that can be more easily visualized:

$$E_x = E_{x0}e^{-\alpha z}\cos(\omega t - \beta z) \tag{41}$$

We recognize this as a uniform plane wave that propagates in the forward z direction with phase constant β, but which (for positive α) loses amplitude with increasing z according to the factor $e^{-\alpha z}$. Thus the general effect of a complex-valued k is to yield a traveling wave that changes its amplitude with distance. If α is positive, it is called the *attenuation coefficient*. If α is negative, the wave grows in amplitude with distance, and α is called the *gain coefficient*. The latter effect would occur, for example, in laser amplifiers. In the present and future discussions in this book, we will consider only passive media, in which one or more loss mechanisms are present, thus producing a positive α.

The attenuation coefficient is measured in nepers per meter (Np/m) so that the exponent of e can be measured in the dimensionless units of nepers. Thus, if $\alpha = 0.01$ Np/m, the crest amplitude of the wave at $z = 50$ m will be $e^{-0.5}/e^{-0} = 0.607$ of its value at $z = 0$. In traveling a distance $1/\alpha$ in the $+z$ direction, the amplitude of the wave is reduced by the familiar factor of e^{-1}, or 0.368.

The ways in which physical processes in a material can affect the wave electric field are described through a *complex permittivity* of the form

$$\epsilon = \epsilon' - j\epsilon'' = \epsilon_0(\epsilon_r' - j\epsilon_r'') \tag{42}$$

Two important mechanisms that give rise to a complex permittivity (and thus result in wave losses) are bound electron or ion oscillations and dipole relaxation, both of which are discussed in Appendix E. An additional mechanism is the conduction of free electrons or holes, which we will explore at length in this chapter.

Losses arising from the response of the medium to the magnetic field can occur as well, and these are modeled through a *complex permeability, $\mu = \mu' - j\mu'' = \mu_0(\mu_r' - j\mu_r'')$.* Examples of such media include *ferrimagnetic* materials, or *ferrites*. The magnetic response is usually very weak compared to the dielectric response in most materials of interest for wave propagation; in such materials $\mu \approx \mu_0$. Consequently, our discussion of loss mechanisms will be confined to those described through the complex permittivity, and we will assume that μ is entirely real in our treatment.

We can substitute (42) into (37), which results in

$$k = \omega\sqrt{\mu(\epsilon' - j\epsilon'')} = \omega\sqrt{\mu\epsilon'}\sqrt{1 - j\frac{\epsilon''}{\epsilon'}} \tag{43}$$

Note the presence of the second radical factor in (43), which becomes unity (and real) as ϵ'' vanishes. With nonzero ϵ'', k is complex, and so losses occur which are quantified through the attenuation coefficient, α, in (39). The phase constant, β (and consequently the wavelength and phase velocity), will also be affected by ϵ''. α and β are found by taking the real and imaginary parts of jk from (43). We obtain:

$$\alpha = \text{Re}\{jk\} = \omega\sqrt{\frac{\mu\epsilon'}{2}}\left(\sqrt{1 + \left(\frac{\epsilon''}{\epsilon'}\right)^2} - 1\right)^{1/2} \tag{44}$$

$$\beta = \text{Im}\{jk\} = \omega\sqrt{\frac{\mu\epsilon'}{2}}\left(\sqrt{1 + \left(\frac{\epsilon''}{\epsilon'}\right)^2} + 1\right)^{1/2} \tag{45}$$

We see that a nonzero α (and hence loss) results if the imaginary part of the permittivity, ϵ'', is present. We also observe in (44) and (45) the presence of the ratio ϵ''/ϵ', which is called the *loss tangent*. The meaning of the term will be demonstrated when we investigate the specific case of conductive media. The practical importance of the ratio lies in its magnitude compared to unity, which enables simplifications to be made in (44) and (45).

Whether or not losses occur, we see from (41) that the wave phase velocity is given by

$$v_p = \frac{\omega}{\beta} \tag{46}$$

The wavelength is the distance required to effect a phase change of 2π radians

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

which leads to the fundamental definition of wavelength,

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

(47)

Because we have a uniform plane wave, the magnetic field is found through

$$H_{ys} = \frac{E_{x0}}{\eta} e^{-\alpha z} e^{-j\beta z}$$

where the intrinsic impedance is now a complex quantity,

$$\eta = \sqrt{\frac{\mu}{\epsilon' - j\epsilon''}} = \sqrt{\frac{\mu}{\epsilon'}} \frac{1}{\sqrt{1 - j(\epsilon''/\epsilon')}}$$

(48)

The electric and magnetic fields are no longer in phase.

A special case is that of a lossless medium, or *perfect dielectric,* in which $\epsilon'' = 0$, and so $\epsilon = \epsilon'$. From (44), this leads to $\alpha = 0$, and from (45),

$$\beta = \omega\sqrt{\mu\epsilon'} \qquad \text{(lossless medium)}$$

(49)

With $\alpha = 0$, the real field assumes the form

$$E_x = E_{x0}\cos(\omega t - \beta z)$$

(50)

We may interpret this as a wave traveling in the $+z$ direction at a phase velocity v_p, where

$$v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{\mu\epsilon'}} = \frac{c}{\sqrt{\mu_r \epsilon_r'}}$$

The wavelength is

$$\lambda = \frac{2\pi}{\beta} = \frac{2\pi}{\omega\sqrt{\mu\epsilon'}} = \frac{1}{f\sqrt{\mu\epsilon'}} = \frac{c}{f\sqrt{\mu_r \epsilon_r'}} = \frac{\lambda_0}{\sqrt{\mu_r \epsilon_r'}} \qquad \text{(lossless medium)}$$

(51)

where λ_0 is the free space wavelength. Note that $\mu_r \epsilon_r' > 1$, and therefore the wavelength is shorter and the velocity is lower in all real media than they are in free space.

Associated with E_x is the magnetic field intensity

$$H_y = \frac{E_{x0}}{\eta}\cos(\omega t - \beta z)$$

where the intrinsic impedance is

$$\eta = \sqrt{\frac{\mu}{\epsilon}}$$

(52)

The two fields are once again perpendicular to each other, perpendicular to the direction of propagation, and in phase with each other everywhere. Note that when **E** is crossed into **H**, the resultant vector is in the direction of propagation. We shall see the reason for this when we discuss the Poynting vector.

<div style="text-align: right">

EXAMPLE 11.3

</div>

Let us apply these results to a 1-MHz plane wave propagating in fresh water. At this frequency, losses in water are negligible, which means that we can assume that $\epsilon'' \doteq 0$. In water, $\mu_r = 1$ and at 1 MHz, $\epsilon_r' = 81$.

Solution. We begin by calculating the phase constant. Using (45) with $\epsilon'' = 0$, we have

$$\beta = \omega\sqrt{\mu\epsilon'} = \omega\sqrt{\mu_0\epsilon_0}\sqrt{\epsilon_r'} = \frac{\omega\sqrt{\epsilon_r'}}{c} = \frac{2\pi \times 10^6\sqrt{81}}{3.0 \times 10^8} = 0.19 \text{ rad/m}$$

Using this result, we can determine the wavelength and phase velocity:

$$\lambda = \frac{2\pi}{\beta} = \frac{2\pi}{.19} = 33 \text{ m}$$

$$v_p = \frac{\omega}{\beta} = \frac{2\pi \times 10^6}{.19} = 3.3 \times 10^7 \text{ m/s}$$

The wavelength in air would have been 300 m. Continuing our calculations, we find the intrinsic impedance using (48) with $\epsilon'' = 0$:

$$\eta = \sqrt{\frac{\mu}{\epsilon'}} = \frac{\eta_0}{\sqrt{\epsilon_r'}} = \frac{377}{9} = 42 \text{ } \Omega$$

If we let the electric field intensity have a maximum amplitude of 0.1 V/m, then

$$E_x = 0.1 \cos(2\pi 10^6 t - .19z) \text{ V/m}$$

$$H_y = \frac{E_x}{\eta} = (2.4 \times 10^{-3}) \cos(2\pi 10^6 t - .19z) \text{ A/m}$$

D11.3. A 9.375-GHz uniform plane wave is propagating in polyethylene (see Appendix C). If the amplitude of the electric field intensity is 500 V/m and the material is assumed to be lossless, find: (*a*) the phase constant; (*b*) the wavelength in the polyethylene; (*c*) the velocity of propagation; (*d*) the intrinsic impedance; (*e*) the amplitude of the magnetic field intensity.

Ans. 295 rad/m; 2.13 cm; 1.99×10^8 m/s; 251 Ω; 1.99 A/m

EXAMPLE 11.4

We again consider plane wave propagation in water, but at the much higher microwave frequency of 2.5 GHz. At frequencies in this range and higher, dipole relaxation and resonance phenomena in the water molecules become important.[2] Real and imaginary parts of the permittivity are present, and both vary with frequency. At frequencies below that of visible light, the two mechanisms together produce a value of ϵ'' that increases with increasing frequency, reaching a maximum in the vicinity of 10^{13} Hz. ϵ' decreases with increasing frequency, reaching a minimum also in the vicinity of 10^{13} Hz. Reference 3 provides specific details. At 2.5 GHz, dipole relaxation effects dominate. The permittivity values are $\epsilon_r' = 78$ and $\epsilon_r'' = 7$. From (44), we have

$$\alpha = \frac{(2\pi \times 2.5 \times 10^9)\sqrt{78}}{(3.0 \times 10^8)\sqrt{2}} \left(\sqrt{1 + \left(\frac{7}{78}\right)^2} - 1 \right)^{1/2} = 21 \text{ Np/m}$$

This first calculation demonstrates the operating principle of the *microwave oven*. Almost all foods contain water, and so they can be cooked when incident microwave radiation is absorbed and converted into heat. Note that the field will attenuate to a value of e^{-1} times its initial value at a distance of $1/\alpha = 4.8$ cm. This distance is called the *penetration depth* of the material, and of course it is frequency-dependent. The 4.8 cm depth is reasonable for cooking food, since it would lead to a temperature rise that is fairly uniform throughout the depth of the material. At much higher frequencies, where ϵ'' is larger, the penetration depth decreases, and too much power is absorbed at the surface; at lower frequencies, the penetration depth increases, and not enough overall absorption occurs. Commercial microwave ovens operate at frequencies in the vicinity of 2.5 GHz.

Using (45), in a calculation very similar to that for α, we find $\beta = 464$ rad/m. The wavelength is $\lambda = 2\pi/\beta = 1.4$ cm, whereas in free space this would have been $\lambda_0 = c/f = 12$ cm.

Using (48), the intrinsic impedance is found to be

$$\eta = \frac{377}{\sqrt{78}} \frac{1}{\sqrt{1 - j(7/78)}} = 43 + j1.9 = 43\angle 2.6° \; \Omega$$

and E_x leads H_y in time by 2.6° at every point.

We next consider the case of conductive materials, in which currents are formed by the motion of free electrons or holes under the influence of an electric field. The governing relation is $\mathbf{J} = \sigma\mathbf{E}$, where σ is the material conductivity. With finite conductivity, the wave loses power through resistive heating of the material. We look for an interpretation of the complex permittivity as it relates to the conductivity.

---

[2] These mechanisms and how they produce a complex permittivity are described in Appendix D. Additionally, the reader is referred to pp. 73–84 in Reference 1 and pp. 678–82 in Reference 2 for general treatments of relaxation and resonance effects on wave propagation. Discussions and data that are specific to water are presented in Reference 3, pp. 314–16.

Consider the Maxwell curl equation (23) which, using (42), becomes:

$$\nabla \times \mathbf{H}_s = j\omega(\epsilon' - j\epsilon'')\mathbf{E}_s = \omega\epsilon''\mathbf{E}_s + j\omega\epsilon'\mathbf{E}_s \qquad (53)$$

This equation can be expressed in a more familiar way, in which conduction current is included:

$$\nabla \times \mathbf{H}_s = \mathbf{J}_s + j\omega\epsilon\mathbf{E}_s \qquad (54)$$

We next use $\mathbf{J}_s = \sigma\mathbf{E}_s$, and interpret ϵ in (54) as ϵ'. The latter equation becomes:

$$\boxed{\nabla \times \mathbf{H}_s = (\sigma + j\omega\epsilon')\mathbf{E}_s = \mathbf{J}_{\sigma s} + \mathbf{J}_{ds}} \qquad (55)$$

which we have expressed in terms of conduction current density, $\mathbf{J}_{\sigma s} = \sigma\mathbf{E}_s$, and displacement current density, $\mathbf{J}_{ds} = j\omega\epsilon'\mathbf{E}_s$. Comparing Eqs. (53) and (55), we find that in a conductive medium:

$$\boxed{\epsilon'' = \frac{\sigma}{\omega}} \qquad (56)$$

Let us now turn our attention to the case of a dielectric material in which the loss is very small. The criterion by which we should judge whether or not the loss is small is the magnitude of the loss tangent, ϵ''/ϵ'. This parameter will have a direct influence on the attenuation coefficient, α, as seen from Eq. (44). In the case of conducting media, to which (56) applies, the loss tangent becomes $\sigma/\omega\epsilon'$. By inspecting (55), we see that the ratio of conduction current density to displacement current density magnitudes is

$$\boxed{\frac{J_{\sigma s}}{J_{ds}} = \frac{\epsilon''}{j\epsilon'} = \frac{\sigma}{j\omega\epsilon'}} \qquad (57)$$

That is, these two vectors point in the same direction in space, but they are 90° out of phase in time. Displacement current density leads conduction current density by 90°, just as the current through a capacitor leads the current through a resistor in parallel with it by 90° in an ordinary electric circuit. This phase relationship is shown in Figure 11.2. The angle θ (not to be confused with the polar angle in spherical coordinates) may therefore be identified as the angle by which the displacement current density leads the total current density, and

$$\boxed{\tan\theta = \frac{\epsilon''}{\epsilon'} = \frac{\sigma}{\omega\epsilon'}} \qquad (58)$$

The reasoning behind the term *loss tangent* is thus evident. Problem 11.16 at the end of the chapter indicates that the Q of a capacitor (its quality factor, not its charge) that incorporates a lossy dielectric is the reciprocal of the loss tangent.

If the loss tangent is small, then we may obtain useful approximations for the attenuation and phase constants, and the intrinsic impedance. The criterion for a small loss tangent is $\epsilon''/\epsilon' \ll 1$, which we say identifies the medium as a *good dielectric*.
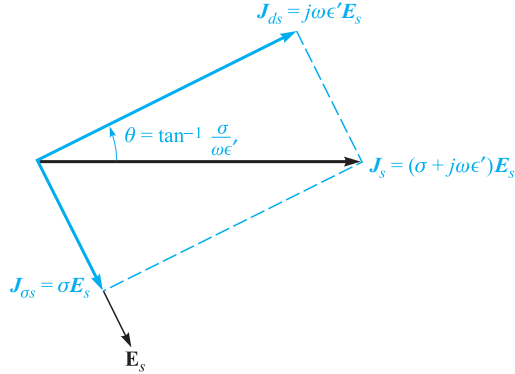
Figure 11.2 The time-phase relationship between $\mathbf{J}_{ds}$, $\mathbf{J}_{\sigma s}$, $\mathbf{J}_s$, and $\mathbf{E}_s$. The tangent of θ is equal to $\sigma/\omega\epsilon'$, and $90° - \theta$ is the common power-factor angle, or the angle by which $\mathbf{J}_s$ leads $\mathbf{E}_s$.

Considering a conductive material, for which $\epsilon'' = \sigma/\omega$, (43) becomes

$$jk = j\omega\sqrt{\mu\epsilon'}\sqrt{1 - j\frac{\sigma}{\omega\epsilon'}} \tag{59}$$

We may expand the second radical using the binomial theorem

$$(1 + x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots$$

where $|x| \ll 1$. We identify x as $-j\sigma/\omega\epsilon'$ and n as $1/2$, and thus

$$jk = j\omega\sqrt{\mu\epsilon'}\left[1 - j\frac{\sigma}{2\omega\epsilon'} + \frac{1}{8}\left(\frac{\sigma}{\omega\epsilon'}\right)^2 + \cdots\right] = \alpha + j\beta$$

Now, for a good dielectric,

$$\alpha = \text{Re}(jk) \doteq j\omega\sqrt{\mu\epsilon'}\left(-j\frac{\sigma}{2\omega\epsilon'}\right) = \frac{\sigma}{2}\sqrt{\frac{\mu}{\epsilon'}} \tag{60a}$$

and

$$\beta = \text{Im}(jk) \doteq \omega\sqrt{\mu\epsilon'}\left[1 + \frac{1}{8}\left(\frac{\sigma}{\omega\epsilon'}\right)^2\right] \tag{60b}$$

Equations (60a) and (60b) can be compared directly with the transmission line α and β under low-loss conditions, as expressed in Eqs. (54a) and (55b) in Chapter 10. In this comparison, we associate σ with G, μ with L, and ϵ with C. Note that in plane wave propagation in media with no boundaries, there can be no quantity that is analogous to the transmission line conductor resistance parameter, R. In many cases,

the second term in (60b) is small enough, so that

$$\beta \doteq \omega\sqrt{\mu\epsilon'} \tag{61}$$

Applying the binomial expansion to (48), we obtain, for a good dielectric

$$\eta \doteq \sqrt{\frac{\mu}{\epsilon'}}\left[1 - \frac{3}{8}\left(\frac{\sigma}{\omega\epsilon'}\right)^2 + j\frac{\sigma}{2\omega\epsilon'}\right] \tag{62a}$$

or

$$\eta \doteq \sqrt{\frac{\mu}{\epsilon'}}\left(1 + j\frac{\sigma}{2\omega\epsilon'}\right) \tag{62b}$$

The conditions under which these approximations can be used depend on the desired accuracy, measured by how much the results deviate from those given by the exact formulas, (44) and (45). Deviations of no more than a few percent occur if $\sigma/\omega\epsilon' < 0.1$.

EXAMPLE 11.5

As a comparison, we repeat the computations of Example 11.4, using the approximation formulas (60a), (61), and (62b).

Solution. First, the loss tangent in this case is $\epsilon''/\epsilon' = 7/78 = 0.09$. Using (60), with $\epsilon'' = \sigma/\omega$, we have

$$\alpha \doteq \frac{\omega\epsilon''}{2}\sqrt{\frac{\mu}{\epsilon'}} = \frac{1}{2}(7 \times 8.85 \times 10^{12})(2\pi \times 2.5 \times 10^9)\frac{377}{\sqrt{78}} = 21 \text{ cm}^{-1}$$

We then have, using (61b),

$$\beta \doteq (2\pi \times 2.5 \times 10^9)\sqrt{78}/(3 \times 10^8) = 464 \text{ rad/m}$$

Finally, with (62b),

$$\eta \doteq \frac{377}{\sqrt{78}}\left(1 + j\frac{7}{2 \times 78}\right) = 43 + j1.9$$

These results are identical (within the accuracy limitations as determined by the given numbers) to those of Example 11.4. Small deviations will be found, as the reader can verify by repeating the calculations of both examples and expressing the results to four or five significant figures. As we know, this latter practice would not be meaningful because the given parameters were not specified with such accuracy. Such is often the case, since measured values are not always known with high precision. Depending on how precise these values are, one can sometimes use a more relaxed judgment on when the approximation formulas can be used by allowing loss tangent values that can be larger than 0.1 (but still less than 1).

D11.4. Given a nonmagnetic material having $\epsilon'_r = 3.2$ and $\sigma = 1.5 \times 10^{-4}$ S/m, find numerical values at 3 MHz for the (*a*) loss tangent; (*b*) attenuation constant; (*c*) phase constant; (*d*) intrinsic impedance.

Ans. 0.28; 0.016 Np/m; 0.11 rad/m; $207\angle 7.8°$ Ω

D11.5. Consider a material for which $\mu_r = 1$, $\epsilon'_r = 2.5$, and the loss tangent is 0.12. If these three values are constant with frequency in the range 0.5 MHz $\leq$ $f \leq$ 100 MHz, calculate: (*a*) σ at 1 and 75 MHz; (*b*) λ at 1 and 75 MHz; (*c*) v_p at 1 and 75 MHz.

Ans. 1.67×10^{-5} and 1.25×10^{-3} S/m; 190 and 2.53 m; 1.90×10^8 m/s twice

11.3 POYNTING'S THEOREM AND WAVE POWER

In order to find the power flow associated with an electromagnetic wave, it is necessary to develop a power theorem for the electromagnetic field known as the Poynting theorem. It was originally postulated in 1884 by an English physicist, John H. Poynting.

The development begins with one of Maxwell's curl equations, in which we assume that the medium may be conductive:

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{63}$$

Next, we take the scalar product of both sides of (63) with **E**,

$$\mathbf{E} \cdot \nabla \times \mathbf{H} = \mathbf{E} \cdot \mathbf{J} + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \tag{64}$$

We then introduce the following vector identity, which may be proved by expansion in rectangular coordinates:

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = -\mathbf{E} \cdot \nabla \times \mathbf{H} + \mathbf{H} \cdot \nabla \times \mathbf{E} \tag{65}$$

Using (65) in the left side of (64) results in

$$\mathbf{H} \cdot \nabla \times \mathbf{E} - \nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{J} \cdot \mathbf{E} + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} \tag{66}$$

where the curl of the electric field is given by the other Maxwell curl equation:

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

Therefore

$$-\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} - \nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{J} \cdot \mathbf{E} + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t}$$

or

$$-\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{J} \cdot \mathbf{E} + \epsilon \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} + \mu \mathbf{H} \cdot \frac{\partial \mathbf{H}}{\partial t} \tag{67}$$

The two time derivatives in (67) can be rearranged as follows:

$$\epsilon \mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} = \frac{\partial}{\partial t}\left(\frac{1}{2}\mathbf{D} \cdot \mathbf{E}\right) \tag{68a}$$

and

$$\mu \mathbf{H} \cdot \frac{\partial \mathbf{H}}{\partial t} = \frac{\partial}{\partial t}\left(\frac{1}{2}\mathbf{B} \cdot \mathbf{H}\right) \tag{68b}$$

With these, Eq. (67) becomes

$$-\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \mathbf{J} \cdot \mathbf{E} + \frac{\partial}{\partial t}\left(\frac{1}{2}\mathbf{D} \cdot \mathbf{E}\right) + \frac{\partial}{\partial t}\left(\frac{1}{2}\mathbf{B} \cdot \mathbf{H}\right) \tag{69}$$

Finally, we integrate (69) throughout a volume:

$$-\int_{\text{vol}} \nabla \cdot (\mathbf{E} \times \mathbf{H})\, dv = \int_{\text{vol}} \mathbf{J} \cdot \mathbf{E}\, dv + \int_{\text{vol}} \frac{\partial}{\partial t}\left(\frac{1}{2}\mathbf{D} \cdot \mathbf{E}\right) dv + \int_{\text{vol}} \frac{\partial}{\partial t}\left(\frac{1}{2}\mathbf{B} \cdot \mathbf{H}\right) dv$$

The divergence theorem is then applied to the left-hand side, thus converting the volume integral there into an integral over the surface that encloses the volume. On the right-hand side, the operations of spatial integration and time differentiation are interchanged. The final result is

$$-\oint_{\text{area}} (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{S} = \int_{\text{vol}} \mathbf{J} \cdot \mathbf{E}\, dv + \frac{d}{dt}\int_{\text{vol}} \frac{1}{2}\mathbf{D} \cdot \mathbf{E}\, dv + \frac{d}{dt}\int_{\text{vol}} \frac{1}{2}\mathbf{B} \cdot \mathbf{H}\, dv \tag{70}$$

Equation (70) is known as Poynting's theorem. On the right-hand side, the first integral is the total (but instantaneous) ohmic power dissipated within the volume. The second integral is the total energy stored in the electric field, and the third integral is the stored energy in the magnetic field.[3] Since time derivatives are taken of the second and third integrals, those results give the time rates of increase of energy stored within the volume, or the instantaneous power going to increase the stored energy. The sum of the expressions on the right must therefore be the total power flowing *into* this volume, and so the total power flowing *out* of the volume is

$$\oint_{\text{area}} (\mathbf{E} \times \mathbf{H}) \cdot d\mathbf{S} \quad \text{W} \tag{71}$$

where the integral is over the closed surface surrounding the volume. The cross product $\mathbf{E} \times \mathbf{H}$ is known as the Poynting vector, $\mathbf{S}$,

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \quad \text{W/m}^2 \tag{72}$$

which is interpreted as an instantaneous power density, measured in watts per square meter (W/m^2). The direction of the vector $\mathbf{S}$ indicates the direction of the instantaneous

---

[3] This is the expression for magnetic field energy that we have been anticipating since Chapter 8.

power flow at a point, and many of us think of the Poynting vector as a "pointing" vector. This homonym, while accidental, is correct.[4]

Because **S** is given by the cross product of **E** and **H**, the direction of power flow at any point is normal to both the **E** and **H** vectors. This certainly agrees with our experience with the uniform plane wave, for propagation in the $+z$ direction was associated with an E_x and H_y component,

$$E_x \mathbf{a}_x \times H_y \mathbf{a}_y = S_z \mathbf{a}_z$$

In a perfect dielectric, the **E** and **H** field amplitudes are given by

$$E_x = E_{x0} \cos(\omega t - \beta z)$$

$$H_y = \frac{E_{x0}}{\eta} \cos(\omega t - \beta z)$$

where η is real. The power density amplitude is therefore

$$S_z = \frac{E_{x0}^2}{\eta} \cos^2(\omega t - \beta z) \tag{73}$$

In the case of a lossy dielectric, E_x and H_y are not in time phase. We have

$$E_x = E_{x0} e^{-\alpha z} \cos(\omega t - \beta z)$$

If we let

$$\eta = |\eta| \angle \theta_\eta$$

then we may write the magnetic field intensity as

$$H_y = \frac{E_{x0}}{|\eta|} e^{-\alpha z} \cos(\omega t - \beta z - \theta_\eta)$$

Thus,

$$S_z = E_x H_y = \frac{E_{x0}^2}{|\eta|} e^{-2\alpha z} \cos(\omega t - \beta z) \cos(\omega t - \beta z - \theta_\eta) \tag{74}$$

Because we are dealing with a sinusoidal signal, the time-average power density, $\langle S_z \rangle$, is the quantity that will ultimately be measured. To find this, we integrate (74) over one cycle and divide by the period $T = 1/f$. Additionally, the identity $\cos A \cos B = 1/2 \cos(A + B) + 1/2 \cos(A - B)$ is applied to the integrand, and we obtain:

$$\langle S_z \rangle = \frac{1}{T} \int_0^T \frac{1}{2} \frac{E_{x0}^2}{|\eta|} e^{-2\alpha z} [\cos(2\omega t - 2\beta z - 2\theta_\eta) + \cos \theta_\eta] \, dt \tag{75}$$

---

[4] Note that the vector symbol **S** is used for the Poynting vector, and is not to be confused with the differential area vector, $d\mathbf{S}$. The latter, as we know, is the product of the outward normal to the surface and the differential area.

The second-harmonic component of the integrand in (75) integrates to zero, leaving only the contribution from the dc component. The result is

$$\langle S_z \rangle = \frac{1}{2} \frac{E_{x0}^2}{|\eta|} e^{-2\alpha z} \cos \theta_\eta \tag{76}$$

Note that the power density attenuates as $e^{-2\alpha z}$, whereas E_x and H_y fall off as $e^{-\alpha z}$.

We may finally observe that the preceding expression can be obtained very easily by using the phasor forms of the electric and magnetic fields. In vector form, this is

$$\langle \mathbf{S} \rangle = \frac{1}{2} \text{Re}(\mathbf{E}_s \times \mathbf{H}_s^*) \quad \text{W/m}^2 \tag{77}$$

In the present case

$$\mathbf{E}_s = E_{x0} e^{-j\beta z} \mathbf{a}_x$$

and

$$\mathbf{H}_s^* = \frac{E_{x0}}{\eta^*} e^{+j\beta z} \mathbf{a}_y = \frac{E_{x0}}{|\eta|} e^{j\theta} e^{+j\beta z} \mathbf{a}_y$$

where E_{x0} has been assumed real. Eq. (77) applies to any sinusoidal electromagnetic wave and gives both the magnitude and direction of the time-average power density.

D11.6. At frequencies of 1, 100, and 3000 MHz, the dielectric constant of ice made from pure water has values of 4.15, 3.45, and 3.20, respectively, while the loss tangent is 0.12, 0.035, and 0.0009, also respectively. If a uniform plane wave with an amplitude of 100 V/m at $z = 0$ is propagating through such ice, find the time-average power density at $z = 0$ and $z = 10$ m for each frequency.

Ans. 27.1 and 25.7 W/m^2; 24.7 and 6.31 W/m^2; 23.7 and 8.63 W/m^2

11.4 PROPAGATION IN GOOD CONDUCTORS: SKIN EFFECT

As an additional study of propagation with loss, we will investigate the behavior of a *good conductor* when a uniform plane wave is established in it. Such a material satisfies the general high-loss criterion, in which the loss tangent, $\epsilon''/\epsilon' \gg 1$. Applying this to a good conductor leads to the more specific criterion, $\sigma/(\omega \epsilon') \gg 1$. As before, we have an interest in losses that occur on wave transmission *into* a good conductor, and we will find new approximations for the phase constant, attenuation coefficient, and intrinsic impedance. New to us, however, is a modification of the basic problem, appropriate for good conductors. This concerns waves associated with electromagnetic fields existing in an external dielectric that adjoins the conductor surface; in this case, the waves propagate *along* the surface. That portion of the overall field that exists within the conductor will suffer dissipative loss arising from the conduction currents it generates. The overall field therefore attenuates with increasing distance

of travel along the surface. This is the mechanism for the resistive transmission line loss that we studied in Chapter 10, and which is embodied in the line resistance parameter, R.

As implied, a good conductor has a high conductivity and large conduction currents. The energy represented by the wave traveling through the material therefore decreases as the wave propagates because ohmic losses are continuously present. When we discussed the loss tangent, we saw that the ratio of conduction current density to the displacement current density in a conducting material is given by $\sigma/\omega\epsilon'$. Choosing a poor metallic conductor and a very high frequency as a conservative example, this ratio[5] for nichrome ($\sigma \doteq 10^6$) at 100 MHz is about 2×10^8. We therefore have a situation where $\sigma/\omega\epsilon' \gg 1$, and we should be able to make several very good approximations to find α, β, and η for a good conductor.

The general expression for the propagation constant is, from (59),

$$jk = j\omega\sqrt{\mu\epsilon'}\sqrt{1 - j\frac{\sigma}{\omega\epsilon'}}$$

which we immediately simplify to obtain

$$jk = j\omega\sqrt{\mu\epsilon'}\sqrt{-j\frac{\sigma}{\omega\epsilon'}}$$

or

$$jk = j\sqrt{-j\omega\mu\sigma}$$

But

$$-j = 1\,\angle{-90°}$$

and

$$\sqrt{1\,\angle{-90°}} = 1\,\angle{-45°} = \frac{1}{\sqrt{2}}(1 - j)$$

Therefore

$$jk = j(1 - j)\sqrt{\frac{\omega\mu\sigma}{2}} = (1 + j)\sqrt{\pi f \mu\sigma} = \alpha + j\beta \tag{78}$$

Hence

$$\boxed{\alpha = \beta = \sqrt{\pi f \mu\sigma}} \tag{79}$$

Regardless of the parameters μ and σ of the conductor or of the frequency of the applied field, α and β are equal. If we again assume only an E_x component traveling in the $+z$ direction, then

$$E_x = E_{x0}e^{-z\sqrt{\pi f \mu\sigma}}\cos\left(\omega t - z\sqrt{\pi f \mu\sigma}\right) \tag{80}$$

---

[5] It is customary to take $\epsilon' = \epsilon_0$ for metallic conductors.

We may tie this field in the conductor to an external field at the conductor surface. We let the region $z > 0$ be the good conductor and the region $z < 0$ be a perfect dielectric. At the boundary surface $z = 0$, (80) becomes

$$E_x = E_{x0} \cos \omega t \qquad (z = 0)$$

This we shall consider as the source field that establishes the fields within the conductor. Since displacement current is negligible,

$$\mathbf{J} = \sigma \mathbf{E}$$

Thus, the conduction current density at any point within the conductor is directly related to $\mathbf{E}$:

$$J_x = \sigma E_x = \sigma E_{x0} e^{-z\sqrt{\pi f \mu \sigma}} \cos\left(\omega t - z\sqrt{\pi f \mu \sigma}\right) \qquad (81)$$

Equations (80) and (81) contain a wealth of information. Considering first the negative exponential term, we find an exponential decrease in the conduction current density and electric field intensity with penetration into the conductor (away from the source). The exponential factor is unity at $z = 0$ and decreases to $e^{-1} = 0.368$ when

$$z = \frac{1}{\sqrt{\pi f \mu \sigma}}$$

This distance is denoted by δ and is termed the *depth of penetration*, or the *skin depth*,

$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} = \frac{1}{\alpha} = \frac{1}{\beta} \qquad (82)$$

It is an important parameter in describing conductor behavior in electromagnetic fields. To get some idea of the magnitude of the skin depth, let us consider copper, $\sigma = 5.8 \times 10^7$ S/m, at several different frequencies. We have

$$\delta_{\text{Cu}} = \frac{0.066}{\sqrt{f}}$$

At a power frequency of 60 Hz, $\delta_{\text{Cu}} = 8.53$ mm. Remembering that the power density carries an exponential term $e^{-2\alpha z}$, we see that the power density is multiplied by a factor of $0.368^2 = 0.135$ for every 8.53 mm of distance into the copper.

At a microwave frequency of 10,000 MHz, δ is 6.61×10^{-4} mm. Stated more generally, all fields in a good conductor such as copper are essentially zero at distances greater than a few skin depths from the surface. Any current density or electric field intensity established at the surface of a good conductor decays rapidly as we progress into the conductor. Electromagnetic energy is not transmitted in the interior of a conductor; it travels in the region surrounding the conductor, while the conductor merely guides the waves. We will consider guided propagation in more detail in Chapter 13.

Suppose we have a copper bus bar in the substation of an electric utility company which we wish to have carry large currents, and we therefore select dimensions of 2 by 4 inches. Then much of the copper is wasted, for the fields are greatly reduced in

one skin depth, about 8.5 mm.[6] A hollow conductor with a wall thickness of about 12 mm would be a much better design. Although we are applying the results of an analysis for an infinite planar conductor to one of finite dimensions, the fields are attenuated in the finite-size conductor in a similar (but not identical) fashion.

The extremely short skin depth at microwave frequencies shows that only the surface coating of the guiding conductor is important. A piece of glass with an evaporated silver surface 3 μm thick is an excellent conductor at these frequencies.

Next, let us determine expressions for the velocity and wavelength within a good conductor. From (82), we already have

$$\alpha = \beta = \frac{1}{\delta} = \sqrt{\pi f \mu \sigma}$$

Then, as

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

we find the wavelength to be

$$\lambda = 2\pi \delta \tag{83}$$

Also, recalling that

$$\nu_p = \frac{\omega}{\beta}$$

we have

$$\boxed{\nu_p = \omega \delta} \tag{84}$$

For copper at 60 Hz, $\lambda = 5.36$ cm and $\nu_p = 3.22$ m/s, or about 7.2 mi/h! A lot of us can run faster than that. In free space, of course, a 60 Hz wave has a wavelength of 3100 mi and travels at the velocity of light.

EXAMPLE 11.6

Let us again consider wave propagation in water, but this time we will consider seawater. The primary difference between seawater and fresh water is of course the salt content. Sodium chloride dissociates in water to form Na^+ and Cl^- ions, which, being charged, will move when forced by an electric field. Seawater is thus conductive, and so it will attenuate electromagnetic waves by this mechanism. At frequencies in the vicinity of 10^7 Hz and below, the bound charge effects in water discussed earlier are negligible, and losses in seawater arise principally from the salt-associated conductivity. We consider an incident wave of frequency 1 MHz. We wish to find the skin depth, wavelength, and phase velocity. In seawater, $\sigma = 4$ S/m, and $\epsilon'_r = 81$.

---

[6] This utility company operates at 60 Hz.

Solution. We first evaluate the loss tangent, using the given data:

$$\frac{\sigma}{\omega\epsilon'} = \frac{4}{(2\pi \times 10^6)(81)(8.85 \times 10^{-12})} = 8.9 \times 10^2 \gg 1$$

Seawater is therefore a good conductor at 1 MHz (and at frequencies lower than this). The skin depth is

$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} = \frac{1}{\sqrt{(\pi \times 10^6)(4\pi \times 10^{-7})(4)}} = 0.25 \text{ m} = 25 \text{ cm}$$

Now

$$\lambda = 2\pi\delta = 1.6 \text{ m}$$

and

$$v_p = \omega\delta = (2\pi \times 10^6)(0.25) = 1.6 \times 10^6 \text{ m/sec}$$

In free space, these values would have been $\lambda = 300$ m and of course $v = c$.

With a 25-cm skin depth, it is obvious that radio frequency communication in seawater is quite impractical. Notice, however, that δ varies as $1/\sqrt{f}$, so that things will improve at lower frequencies. For example, if we use a frequency of 10 Hz (in the ELF, or extremely low frequency range), the skin depth is increased over that at 1 MHz by a factor of $\sqrt{10^6/10}$, so that

$$\delta(10 \text{ Hz}) \doteq 80 \text{ m}$$

The corresponding wavelength is $\lambda = 2\pi\delta \doteq 500$ m. Frequencies in the ELF range were used for many years in submarine communications. Signals were transmitted from gigantic ground-based antennas (required because the free-space wavelength associated with 10 Hz is 3×10^7 m). The signals were then received by submarines, from which a suspended wire antenna of length shorter than 500 m is sufficient. The drawback is that signal data rates at ELF are slow enough that a single word can take several minutes to transmit. Typically, ELF signals would be used to tell the submarine to initiate emergency procedures, or to come near the surface in order to receive a more detailed message via satellite.

We next turn our attention to finding the magnetic field, H_y, associated with E_x. To do so, we need an expression for the intrinsic impedance of a good conductor. We begin with Eq. (48), Section 11.2, with $\epsilon'' = \sigma/\omega$,

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

Since $\sigma \gg \omega\epsilon'$, we have

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

which may be written as

$$\eta = \frac{\sqrt{2}\angle 45°}{\sigma \delta} = \frac{(1+j)}{\sigma \delta} \tag{85}$$

Thus, if we write (80) in terms of the skin depth,

$$E_x = E_{x0}e^{-z/\delta} \cos\left(\omega t - \frac{z}{\delta}\right) \tag{86}$$

then

$$H_y = \frac{\sigma \delta E_{x0}}{\sqrt{2}} e^{-z/\delta} \cos\left(\omega t - \frac{z}{\delta} - \frac{\pi}{4}\right) \tag{87}$$

and we see that the maximum amplitude of the magnetic field intensity occurs one-eighth of a cycle later than the maximum amplitude of the electric field intensity at every point.

From (86) and (87) we may obtain the time-average Poynting vector by applying (77),

$$\langle S_z \rangle = \frac{1}{2} \frac{\sigma \delta E_{x0}^2}{\sqrt{2}} e^{-2z/\delta} \cos\left(\frac{\pi}{4}\right)$$

or

$$\langle S_z \rangle = \frac{1}{4} \sigma \delta E_{x0}^2 e^{-2z/\delta}$$

We again note that in a distance of one skin depth the power density is only $e^{-2} = 0.135$ of its value at the surface.

The total average power loss in a width $0 < y < b$ and length $0 < x < L$ in the direction of the current, as shown in Figure 11.3, is obtained by finding the power
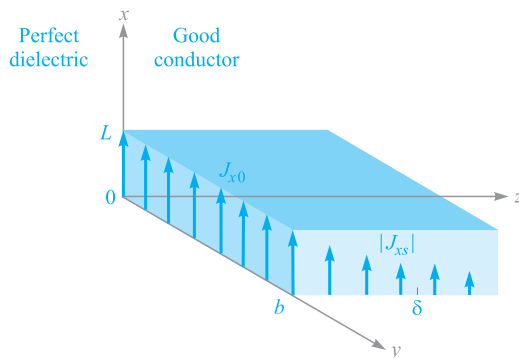
Figure 11.3   The current density $J_x = J_{x0}e^{-z/\delta}e^{-jz/\delta}$ decreases in magnitude as the wave propagates into the conductor. The average power loss in the region $0 < x < L$, $0 < y < b$, $z > 0$, is $\delta bL\, J_{x0}^2/4\sigma$ watts.

crossing the conductor surface within this area,

$$P_L = \int_{\text{area}} \langle S_z \rangle da = \int_0^b \int_0^L \frac{1}{4} \sigma \delta E_{x0}^2 e^{-2z/\delta} \Big|_{z=0} dx \, dy = \frac{1}{4} \sigma \delta b L E_{x0}^2$$

In terms of the current density J_{x0} at the surface,

$$J_{x0} = \sigma E_{x0}$$

we have

$$P_L = \frac{1}{4\sigma} \delta b L J_{x0}^2 \tag{88}$$

Now let us see what power loss would result if the *total* current in a width b were distributed *uniformly* in one skin depth. To find the total current, we integrate the current density over the infinite depth of the conductor,

$$I = \int_0^\infty \int_0^b J_x \, dy \, dz$$

where

$$J_x = J_{x0} e^{-z/\delta} \cos\left(\omega t - \frac{z}{\delta}\right)$$

or in complex exponential notation to simplify the integration,

$$J_{xs} = J_{x0} e^{-z/\delta} e^{-jz/\delta}$$
$$= J_{x0} e^{-(1+j)z/\delta}$$

Therefore,

$$I_s = \int_0^\infty \int_0^b J_{x0} e^{-(1+j)z/\delta} dy \, dz$$
$$= J_{x0} b e^{-(1+j)z/\delta} \frac{-\delta}{1+j} \Big|_0^\infty$$
$$= \frac{J_{x0} b \delta}{1+j}$$

and

$$I = \frac{J_{x0} b \delta}{\sqrt{2}} \cos\left(\omega t - \frac{\pi}{4}\right)$$

If this current is distributed with a uniform density J' throughout the cross section $0 < y < b, 0 < z < \delta$, then

$$J' = \frac{J_{x0}}{\sqrt{2}} \cos\left(\omega t - \frac{\pi}{4}\right)$$

The ohmic power loss per unit volume is $\mathbf{J} \cdot \mathbf{E}$, and thus the total instantaneous power dissipated in the volume under consideration is

$$P_{Li}(t) = \frac{1}{\sigma} (J')^2 b L \delta = \frac{J_{x0}^2}{2\sigma} b L \delta \cos^2\left(\omega t - \frac{\pi}{4}\right)$$

The time-average power loss is easily obtained, since the average value of the cosine-squared factor is one-half,

$$P_L = \frac{1}{4\sigma} J_{x0}^2 b L \delta \qquad (89)$$

Comparing (88) and (89), we see that they are identical. Thus the average power loss in a conductor with skin effect present may be calculated by assuming that the total current is distributed uniformly in one skin depth. In terms of resistance, we may say that the resistance of a width b and length L of an infinitely thick slab with skin effect is the same as the resistance of a rectangular slab of width b, length L, and thickness δ without skin effect, or with uniform current distribution.

We may apply this to a conductor of circular cross section with little error, provided that the radius a is much greater than the skin depth. The resistance at a high frequency where there is a well-developed skin effect is therefore found by considering a slab of width equal to the circumference $2\pi a$ and thickness δ. Hence

$$R = \frac{L}{\sigma S} = \frac{L}{2\pi a \sigma \delta} \qquad (90)$$

A round copper wire of 1 mm radius and 1 km length has a resistance at direct current of

$$R_{\text{dc}} = \frac{10^3}{\pi 10^{-6}(5.8 \times 10^7)} = 5.48 \ \Omega$$

At 1 MHz, the skin depth is 0.066 mm. Thus $\delta \ll a$, and the resistance at 1 MHz is found by (90),

$$R = \frac{10^3}{2\pi 10^{-3}(5.8 \times 10^7)(0.066 \times 10^{-3})} = 41.5 \ \Omega$$

D11.7. A steel pipe is constructed of a material for which $\mu_r = 180$ and $\sigma = 4 \times 10^6$ S/m. The two radii are 5 and 7 mm, and the length is 75 m. If the total current $I(t)$ carried by the pipe is 8 cos ωt A, where $\omega = 1200\pi$ rad/s, find: (a) the skin depth; (b) the effective resistance; (c) the dc resistance; (d) the time-average power loss.

Ans. 0.766 mm; 0.557 Ω; 0.249 Ω; 17.82 W

11.5 WAVE POLARIZATION

In the previous sections, we have treated uniform plane waves in which the electric and magnetic field vectors were assumed to lie in fixed directions. Specifically, with the wave propagating along the z axis, **E** was taken to lie along x, which then required **H** to lie along y. This orthogonal relationship between **E**, **H**, and **S** is always true for a uniform plane wave. The directions of **E** and **H** within the plane perpendicular to $\mathbf{a}_z$

may change, however, as functions of time and position, depending on how the wave was generated or on what type of medium it is propagating through. Thus a complete description of an electromagnetic wave would not only include parameters such as its wavelength, phase velocity, and power, but also a statement of the instantaneous orientation of its field vectors. We define the *wave polarization* as the time-dependent electric field vector orientation at a fixed point in space. A more complete character- ization of a wave's polarization would in fact include specifying the field orientation at *all* points because some waves demonstrate spatial variations in their polarization. Specifying only the electric field direction is sufficient, since magnetic field is readily found from **E** using Maxwell's equations.

In the waves we have previously studied, **E** was in a fixed straight orientation for all times and positions. Such a wave is said to be *linearly polarized*. We have taken **E** to lie along the x axis, but the field could be oriented in any fixed direction in the xy plane and be linearly polarized. For positive z propagation, the wave would in general have its electric field phasor expressed as

$$\mathbf{E}_s = (E_{x0}\mathbf{a}_x + E_{y0}\mathbf{a}_y)e^{-\alpha z}e^{-j\beta z} \tag{91}$$

where E_{x0} and E_{y0} are constant amplitudes along x and y. The magnetic field is readily found by determining its x and y components directly from those of $\mathbf{E}_s$. Specifically, $\mathbf{H}_s$ for the wave of Eq. (91) is

$$\mathbf{H}_s = \left[H_{x0}\mathbf{a}_x + H_{y0}\mathbf{a}_y\right]e^{-\alpha z}e^{-j\beta z} = \left[-\frac{E_{y0}}{\eta}\mathbf{a}_x + \frac{E_{x0}}{\eta}\mathbf{a}_y\right]e^{-\alpha z}e^{-j\beta z} \tag{92}$$

The two fields are sketched in Figure 11.4. The figure demonstrates the reason for the minus sign in the term involving E_{y0} in Eq. (92). The direction of power flow, given by $\mathbf{E} \times \mathbf{H}$, is in the positive z direction in this case. A component of **E** in the
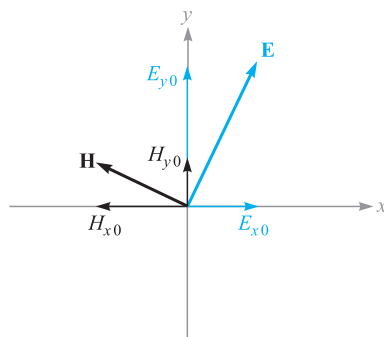
Figure 11.4 Electric and magnetic field configuration for a general linearly polarized plane wave propagating in the forward z direction (out of the page). Field components correspond to those in Eqs. (91) and (92).

positive y direction would require a component of **H** in the negative x direction—thus the minus sign. Using (91) and (92), the power density in the wave is found using (77):

$$\langle \mathbf{S}_z \rangle = \frac{1}{2} \text{Re}\{\mathbf{E}_s \times \mathbf{H}_s^*\} = \frac{1}{2} \text{Re}\{E_{x0} H_{y0}^* (\mathbf{a}_x \times \mathbf{a}_y) + E_{y0} H_{x0}^* (\mathbf{a}_y \times \mathbf{a}_x)\} e^{-2\alpha z}$$

$$= \frac{1}{2} \text{Re}\left\{ \frac{E_{x0} E_{x0}^*}{\eta^*} + \frac{E_{y0} E_{y0}^*}{\eta^*} \right\} e^{-2\alpha z} \mathbf{a}_z$$

$$= \frac{1}{2} \text{Re}\left\{ \frac{1}{\eta^*} \right\} (|E_{x0}|^2 + |E_{y0}|^2) e^{-2\alpha z} \mathbf{a}_z \ \text{W/m}^2$$

This result demonstrates the idea that our linearly polarized plane wave can be considered as two distinct plane waves having x and y polarizations, whose electric fields are combining *in phase* to produce the total **E**. The same is true for the magnetic field components. This is a critical point in understanding wave polarization, in that *any polarization state can be described in terms of mutually perpendicular components of the electric field and their relative phasing.*

We next consider the effect of a phase difference, ϕ, between E_{x0} and E_{y0}, where $\phi < \pi/2$. For simplicity, we will consider propagation in a lossless medium. The total field in phasor form is

$$\mathbf{E}_s = (E_{x0} \mathbf{a}_x + E_{y0} e^{j\phi} \mathbf{a}_y) e^{-j\beta z} \tag{93}$$

Again, to aid in visualization, we convert this wave to real instantaneous form by multiplying by $e^{j\omega t}$ and taking the real part:

$$\mathbf{E}(z, t) = E_{x0} \cos(\omega t - \beta z) \mathbf{a}_x + E_{y0} \cos(\omega t - \beta z + \phi) \mathbf{a}_y \tag{94}$$

where we have assumed that E_{x0} and E_{y0} are real. Suppose we set $t = 0$, in which case (94) becomes [using $\cos(-x) = \cos(x)$]

$$\mathbf{E}(z, 0) = E_{x0} \cos(\beta z) \mathbf{a}_x + E_{y0} \cos(\beta z - \phi) \mathbf{a}_y \tag{95}$$

The component magnitudes of $\mathbf{E}(z, 0)$ are plotted as functions of z in Figure 11.5. Since time is fixed at zero, the wave is frozen in position. An observer can move along the z axis, measuring the component magnitudes and thus the orientation of the total electric field at each point. Let's consider a crest of E_x, indicated as point a in Figure 11.5. If ϕ were zero, E_y would have a crest at the same location. Since ϕ is not zero (and positive), the crest of E_y that would otherwise occur at point a is now displaced to point b farther down z. The two points are separated by distance ϕ/β. E_y thus *lags behind* E_x when we consider the *spatial* dimension.

Now suppose the observer stops at some location on the z axis, and time is allowed to move forward. Both fields now move in the positive z direction, as (94) indicates. But point b reaches the observer first, followed by point a. So we see that E_y *leads* E_x when we consider the *time* dimension. In either case (fixed t and varying z, or vice versa) the observer notes that the net field rotates about the z axis while its magnitude changes. Considering a starting point in z and t, at which the field has a given orientation and magnitude, the wave will return to the same orientation and
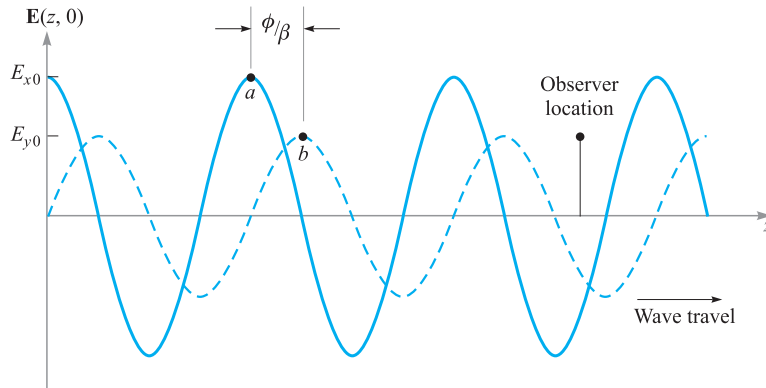
Figure 11.5 Plots of the electric field component magnitudes in Eq. (95) as functions of z. Note that the y component lags behind the x component in z. As time increases from zero, both waves travel to the right, as per Eq. (94). Thus, to an observer at a fixed location, the y component leads in time.

magnitude at a distance of one wavelength in z (for fixed t) or at a time $t = 2\pi/\omega$ later (at a fixed z).

For illustration purposes, if we take the length of the field vector as a measure of its magnitude, we find that at a fixed position, the tip of the vector traces out the shape of an ellipse over time $t = 2\pi/\omega$. The wave is said to be *elliptically polarized*. Elliptical polarization is in fact the most general polarization state of a wave, since it encompasses any magnitude and phase difference between E_x and E_y. Linear polarization is a special case of elliptical polarization in which the phase difference is zero.

Another special case of elliptical polarization occurs when $E_{x0} = E_{y0} = E_0$ and when $\phi = \pm\pi/2$. The wave in this case exhibits *circular polarization*. To see this, we incorporate these restrictions into Eq. (94) to obtain

$$\mathbf{E}(z, t) = E_0[\cos(\omega t - \beta z)\mathbf{a}_x + \cos(\omega t - \beta z \pm \pi/2)\mathbf{a}_y]$$
$$= E_0[\cos(\omega t - \beta z)\mathbf{a}_x \mp \sin(\omega t - \beta z)\mathbf{a}_y] \tag{96}$$

If we consider a fixed position along z (such as $z = 0$) and allow time to vary, (96), with $\phi = +\pi/2$, becomes

$$\mathbf{E}(0, t) = E_0[\cos(\omega t)\mathbf{a}_x - \sin(\omega t)\mathbf{a}_y] \tag{97}$$

If we choose $-\pi/2$ in (96), we obtain

$$\mathbf{E}(0, t) = E_0[\cos(\omega t)\mathbf{a}_x + \sin(\omega t)\mathbf{a}_y] \tag{98}$$

The field vector of Eq. (98) rotates in the counterclockwise direction in the xy plane, while maintaining constant amplitude E_0, and so the tip of the vector traces out a circle. Figure 11.6 shows this behavior.
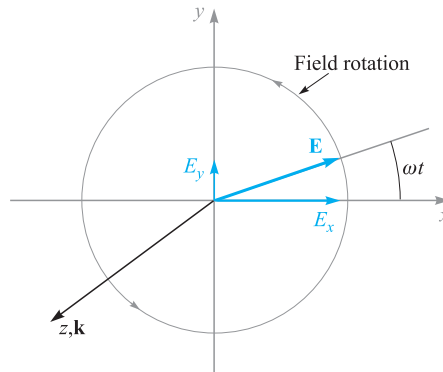
Figure 11.6   Electric field in the *xy* plane of a right circularly polarized plane wave, as described by Eq. (98). As the wave propagates in the forward *z* direction, the field vector rotates counterclockwise in the *xy* plane.

Choosing $+\pi/2$ leads to (97), whose field vector rotates in the clockwise direction. The *handedness* of the circular polarization is associated with the rotation and propagation directions in the following manner: The wave exhibits *left circular polarization* (l.c.p.) if, when orienting the left hand with the thumb in the direction of propagation, the fingers curl in the rotation direction of the field with time. The wave exhibits *right circular polarization* (r.c.p.) if, with the right-hand thumb in the propagation direction, the fingers curl in the field rotation direction.[7] Thus, with forward *z* propagation, (97) describes a left circularly polarized wave, and (98) describes a right circularly polarized wave. The same convention is applied to elliptical polarization, in which the descriptions *left elliptical polarization* and *right elliptical polarization* are used.

Using (96), the instantaneous angle of the field from the *x* direction can be found for any position along *z* through

$$\theta(z, t) = \tan^{-1}\left(\frac{E_y}{E_x}\right) = \tan^{-1}\left(\frac{\mp\sin(\omega t - \beta z)}{\cos(\omega t - \beta z)}\right) = \mp(\omega t - \beta z) \qquad (99)$$

where again the minus sign (yielding l.c.p. for positive *z* travel) applies for the choice of $\phi = +\pi/2$ in (96); the plus sign (yielding r.c.p. for positive *z* travel) is used if

---

[7] This convention is reversed by some workers (most notably in optics) who emphasize the importance of the *spatial* field configuration. Note that r.c.p. by our definition is formed by propagating a spatial field that is in the shape of a *left-handed* screw, and for that reason it is sometimes called left circular polarization (see Figure 11.7). Left circular polarization as we define it results from propagating a spatial field in the shape of a right-handed screw, and it is called right circular polarization by the spatial enthusiasts. Caution is obviously necessary in interpreting what is meant when polarization handedness is stated in an unfamiliar text.
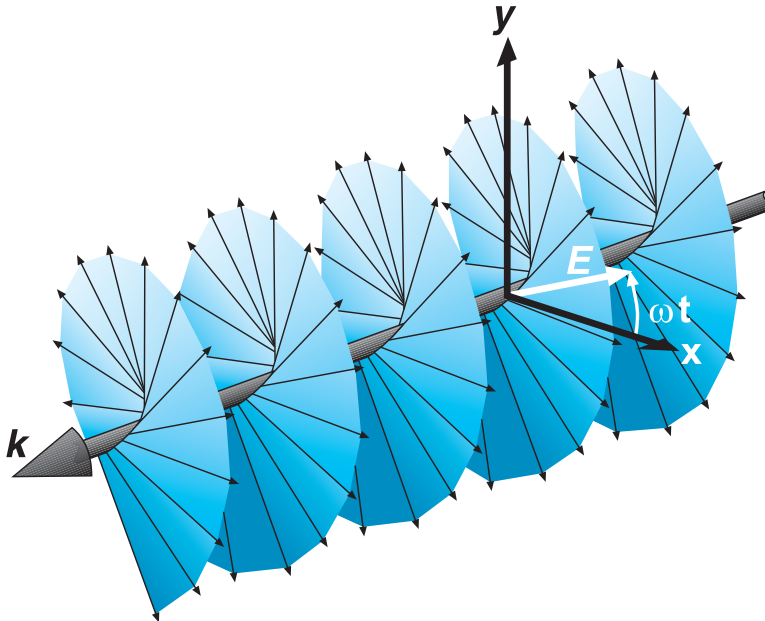
Figure 11.7 Representation of a right circularly polarized wave. The electric field vector (in white) will rotate toward the *y* axis as the entire wave moves through the *xy* plane in the direction of **k**. This counterclockwise rotation (when looking toward the wave source) satisfies the temporal right-handed rotation convention as described in the text. The wave, however, appears as a left-handed screw, and for this reason it is called left circular polarization in the other convention.

$\phi = -\pi/2$. If we choose $z = 0$, the angle becomes simply ωt, which reaches 2π (one complete rotation) at time $t = 2\pi/\omega$. If we choose $t = 0$ and allow z to vary, we form a corkscrew-like field pattern. One way to visualize this is to consider a spiral staircase–shaped pattern, in which the field lines (stairsteps) are perpendicular to the z (or staircase) axis. The relationship between this spatial field pattern and the resulting time behavior at fixed z as the wave propagates is shown in an artist's conception in Figure 11.7.

The handedness of the polarization is changed by reversing the pitch of the corkscrew pattern. The spiral staircase model is only a visualization aid. It must be remembered that the wave is still a uniform plane wave whose fields at any position along z are infinite in extent over the transverse plane.

There are many uses of circularly polarized waves. Perhaps the most obvious advantage is that reception of a wave having circular polarization does not depend on the antenna orientation in the plane normal to the propagation direction. Dipole antennas, for example, are required to be oriented along the electric field direction of the signal they receive. If circularly polarized signals are transmitted, the receiver orientation requirements are relaxed considerably. In optics, circularly polarized light

can be passed through a polarizer of any orientation, thus yielding linearly polarized light in any direction (although one loses half the original power this way). Other uses involve treating linearly polarized light as a superposition of circularly polarized waves, to be described next.

Circularly polarized light can be generated using an *anisotropic* medium—a material whose permittivity is a function of electric field direction. Many crystals have this property. A crystal orientation can be found such that along one direction (say, the x axis), the permittivity is lowest, while along the orthogonal direction (y axis), the permittivity is highest. The strategy is to input a linearly polarized wave with its field vector at 45 degrees to the x and y axes of the crystal. It will thus have equal-amplitude x and y components in the crystal, and these will now propagate in the z direction at different speeds. A phase difference (or *retardation*) accumulates between the components as they propagate, which can reach $\pi/2$ if the crystal is long enough. The wave at the output thus becomes circularly polarized. Such a crystal, cut to the right length and used in this manner, is called a *quarter-wave plate,* since it introduces a relative phase shift of $\pi/2$ between E_x and E_y, which is equivalent to $\lambda/4$.

It is useful to express circularly polarized waves in phasor form. To do this, we note that (96) can be expressed as

$$\mathbf{E}(z, t) = \text{Re}\{E_0 e^{j\omega t} e^{-j\beta z}[\mathbf{a}_x + e^{\pm j\pi/2}\mathbf{a}_y]\}$$

Using the fact that $e^{\pm j\pi/2} = \pm j$, we identify the phasor form as:

$$\boxed{\mathbf{E}_s = E_0(\mathbf{a}_x \pm j\mathbf{a}_y)e^{-j\beta z}} \tag{100}$$

where the plus sign is used for left circular polarization and the minus sign for right circular polarization. If the wave propagates in the negative z direction, we have

$$\boxed{\mathbf{E}_s = E_0(\mathbf{a}_x \pm j\mathbf{a}_y)e^{+j\beta z}} \tag{101}$$

where in this case the positive sign applies to right circular polarization and the minus sign to left circular polarization. The student is encouraged to verify this.

EXAMPLE 11.7

Let us consider the result of superimposing left and right circularly polarized fields of the same amplitude, frequency, and propagation direction, but where a phase shift of δ radians exists between the two.

Solution. Taking the waves to propagate in the $+z$ direction, and introducing a relative phase, δ, the total phasor field is found, using (100):

$$\mathbf{E}_{sT} = \mathbf{E}_{sR} + \mathbf{E}_{sL} = E_0[\mathbf{a}_x - j\mathbf{a}_y]e^{-j\beta z} + E_0[\mathbf{a}_x + j\mathbf{a}_y]e^{-j\beta z}e^{j\delta}$$

Grouping components together, this becomes

$$\mathbf{E}_{sT} = E_0[(1 + e^{j\delta})\mathbf{a}_x - j(1 - e^{j\delta})\mathbf{a}_y]e^{-j\beta z}$$

Factoring out an overall phase term, $e^{j\delta/2}$, we obtain

$$\mathbf{E}_{sT} = E_0 e^{j\delta/2}\left[(e^{-j\delta/2} + e^{j\delta/2})\mathbf{a}_x - j(e^{-j\delta/2} - e^{j\delta/2})\mathbf{a}_y\right]e^{-j\beta z}$$

From Euler's identity, we find that $e^{j\delta/2} + e^{-j\delta/2} = 2\cos\delta/2$, and $e^{j\delta/2} - e^{-j\delta/2} = 2j\sin\delta/2$. Using these relations, we obtain

$$\mathbf{E}_{sT} = 2E_0[\cos(\delta/2)\mathbf{a}_x + \sin(\delta/2)\mathbf{a}_y]e^{-j(\beta z - \delta/2)} \tag{102}$$

We recognize (102) as the electric field of a *linearly polarized* wave, whose field vector is oriented at angle $\delta/2$ from the x axis.

Example 11.7 shows that any linearly polarized wave can be expressed as the sum of two circularly polarized waves of opposite handedness, where the linear polarization direction is determined by the relative phase difference between the two waves. Such a representation is convenient (and necessary) when considering, for example, the propagation of linearly polarized light through media which contain organic molecules. These often exhibit spiral structures having left- or right-handed pitch, and they will thus interact differently with left- or right-hand circular polarization. As a result, the left circular component can propagate at a different speed than the right circular component, and so the two waves will accumulate a phase difference as they propagate. As a result, the direction of the linearly polarized field vector at the output of the material will differ from the direction that it had at the input. The extent of this rotation can be used as a measurement tool to aid in material studies.

Polarization issues will become extremely important when we consider wave reflection in Chapter 12.

REFERENCES

1. Balanis, C. A. *Advanced Engineering Electromagnetics*. New York: John Wiley & Sons, 1989.
2. International Telephone and Telegraph Co., Inc. *Reference Data for Radio Engineers.* 7th ed. Indianapolis, Ind.: Howard W. Sams & Co., 1985. This handbook has some excellent data on the properties of dielectric and insulating materials.
3. Jackson, J. D. *Classical Electrodynamics*. 3d ed. New York: John Wiley & Sons, 1999.
4. Ramo, S., J. R. Whinnery, and T. Van Duzer. *Fields and Waves in Communication Electronics*. 3d ed. New York: John Wiley & Sons, 1994.

CHAPTER 11   PROBLEMS

11.1 Show that $E_{xs} = Ae^{j(k_0 z + \phi)}$ is a solution of the vector Helmholtz equation, Eq. (30), for $k_0 = \omega\sqrt{\mu_0\epsilon_0}$ and any ϕ and A.

11.2 A 10 GHz uniform plane wave propagates in a lossless medium for which $\epsilon_r = 8$ and $\mu_r = 2$. Find (*a*) v_p; (*b*) β; (*c*) λ; (*d*) $\mathbf{E}_s$; (*e*) $\mathbf{H}_s$; (*f*) $\langle\mathbf{S}\rangle$.

11.3 An **H** field in free space is given as $\mathcal{H}(x, t) = 10\cos(10^8 t - \beta x)\mathbf{a}_y$ A/m. Find (*a*) β; (*b*) λ; (*c*) $\mathcal{E}(x, t)$ at $P(0.1, 0.2, 0.3)$ at $t = 1$ ns.

11.4 Small antennas have low efficiencies (as will be seen in Chapter 14), and the efficiency increases with size up to the point at which a critical dimension of

the antenna is an appreciable fraction of a wavelength, say $\lambda/8$. (*a*) An antenna that is 12 cm long is operated in air at 1 MHz. What fraction of a wavelength long is it? (*b*) The same antenna is embedded in a ferrite material for which $\epsilon_r = 20$ and $\mu_r = 2,000$. What fraction of a wavelength is it now?

11.5 A 150 MHz uniform plane wave in free space is described by $\mathbf{H}_s = (4 + j10)(2\mathbf{a}_x + j\mathbf{a}_y)e^{-j\beta z}$ A/m. (*a*) Find numerical values for ω, λ, and β. (*b*) Find $\mathcal{H}(z, t)$ at $t = 1.5$ ns, $z = 20$ cm. (*c*) What is $|E|_{\max}$?

11.6 A uniform plane wave has electric field $\mathbf{E}_s = (E_{y0}\,\mathbf{a}_y - E_{z0}\,\mathbf{a}_z)\,e^{-\alpha x}\,e^{-j\beta x}$ V/m. The intrinsic impedance of the medium is given as $\eta = |\eta|\,e^{j\phi}$, where ϕ is a constant phase. (*a*) Describe the wave polarization and state the direction of propagation. (*b*) Find $\mathbf{H}_s$. (*c*) Find $\mathcal{E}(x, t)$ and $\mathcal{H}(x, t)$. (*d*) Find $< \mathbf{S} >$ in W/m². (*e*) Find the time-average power in watts that is intercepted by an antenna of rectangular cross-section, having width w and height h, suspended parallel to the yz plane, and at a distance d from the wave source.

11.7 The phasor magnetic field intensity for a 400 MHz uniform plane wave propagating in a certain lossless material is $(2\mathbf{a}_y - j5\mathbf{a}_z)e^{-j25x}$ A/m. Knowing that the maximum amplitude of $\mathbf{E}$ is 1500 V/m, find β, η, λ, v_p, ϵ_r, μ_r, and $\mathcal{H}(x, y, z, t)$.

11.8 An electric field in free space is given in spherical coordinates as $\mathbf{E}_s(r) = E_0(r)e^{-jkr}\,\mathbf{a}_\theta$ V/m. (*a*) Find $\mathbf{H}_s(r)$ assuming uniform plane wave behavior. (*b*) Find $< \mathbf{S} >$. (*c*) Express the average outward power in watts through a closed spherical shell of radius r, centered at the origin. (*d*) Establish the required functional form of $E_0(r)$ that will enable the power flow in part c to be independent of radius. With this condition met, the given field becomes that of an *isotropic radiator* in a lossless medium (radiating equal power density in all directions).

11.9 A certain lossless material has $\mu_r = 4$ and $\epsilon_r = 9$. A 10-MHz uniform plane wave is propagating in the $\mathbf{a}_y$ direction with $E_{x0} = 400$ V/m and $E_{y0} = E_{z0} = 0$ at $P(0.6, 0.6, 0.6)$ at $t = 60$ ns. Find (*a*) β, λ, v_p, and η; (*b*) $\mathcal{E}(y, t)$; (*c*) $\mathcal{H}(y, t)$.

11.10 In a medium characterized by intrinsic impedance $\eta = |\eta|e^{j\phi}$, a linearly polarized plane wave propagates, with magnetic field given as $\mathbf{H}_s = (H_{0y}\mathbf{a}_y + H_{0z}\mathbf{a}_z)e^{-\alpha x}e^{-j\beta x}$. Find (*a*) $\mathbf{E}_s$; (*b*) $\mathcal{E}(x, t)$; (*c*) $\mathcal{H}(x, t)$; (*d*) $\langle \mathbf{S} \rangle$.

11.11 A 2 GHz uniform plane wave has an amplitude $E_{y0} = 1.4$ kV/m at $(0, 0, 0, t = 0)$ and is propagating in the $\mathbf{a}_z$ direction in a medium where $\epsilon'' = 1.6 \times 10^{-11}$ F/m, $\epsilon' = 3.0 \times 10^{-11}$ F/m, and $\mu = 2.5\ \mu$H/m. Find (*a*) E_y at $P(0, 0, 1.8$ cm) at 0.2 ns; (*b*) H_x at P at 0.2 ns.

11.12 Describe how the attenuation coefficient of a liquid medium, assumed to be a good conductor, could be determined through measurement of wavelength

in the liquid at a known frequency. What restrictions apply? Could this method be used to find the conductivity as well?

11.13 Let $jk = 0.2 + j1.5 \text{ m}^{-1}$ and $\eta = 450 + j60 \ \Omega$ for a uniform plane propagating in the $\mathbf{a}_z$ direction. If $\omega = 300$ Mrad/s, find μ, ϵ', and ϵ'' for the medium.

11.14 A certain nonmagnetic material has the material constants $\epsilon'_r = 2$ and $\epsilon''/\epsilon' = 4 \times 10^{-4}$ at $\omega = 1.5$ Grad/s. Find the distance a uniform plane wave can propagate through the material before (a) it is attenuated by 1 Np; (b) the power level is reduced by one-half; (c) the phase shifts $360°$.

11.15 A 10 GHz radar signal may be represented as a uniform plane wave in a sufficiently small region. Calculate the wavelength in centimeters and the attenuation in nepers per meter if the wave is propagating in a nonmagnetic material for which (a) $\epsilon'_r = 1$ and $\epsilon''_r = 0$; (b) $\epsilon'_r = 1.04$ and $\epsilon''_r = 9.00 \times 10^{-4}$; (c) $\epsilon'_r = 2.5$ and $\epsilon''_r = 7.2$.

11.16 Consider the power dissipation term, $\int \mathbf{E} \cdot \mathbf{J} dv$, in Poynting's theorem (Eq. (70)). This gives the power lost to heat within a volume into which electromagnetic waves enter. The term $p_d = \mathbf{E} \cdot \mathbf{J}$ is thus the power dissipation per unit volume in W/m^3. Following the same reasoning that resulted in Eq. (77), the time-average power dissipation per volume will be $< p_d > = (1/2)\mathcal{R}e \left\{ \mathbf{E}_s \cdot \mathbf{J}_s^* \right\}$. (a) Show that in a conducting medium, through which a uniform plane wave of amplitude E_0 propagates in the forward z direction, $< p_d > = (\sigma/2)|E_0|^2 e^{-2\alpha z}$. (b) Confirm this result for the special case of a good conductor by using the left hand side of Eq. (70), and consider a very small volume.

11.17 Let $\eta = 250 + j30 \ \Omega$ and $jk = 0.2 + j2 \text{m}^{-1}$ for a uniform plane wave propagating in the $\mathbf{a}_z$ direction in a dielectric having some finite conductivity. If $|E_s| = 400$ V/m at $z = 0$, find (a) $\langle \mathbf{S} \rangle$ at $z = 0$ and $z = 60$ cm; (b) the average ohmic power dissipation in watts per cubic meter at $z = 60$ cm.

11.18 Given a 100-MHz uniform plane wave in a medium known to be a good dielectric, the phasor electric field is $\mathcal{E}_s = 4e^{-0.5z}e^{-j20z}\mathbf{a}_x$ V/m. Determine (a) ϵ'; (b) ϵ''; (c) η; (d) $\mathbf{H}_s$; (e) $\langle \mathbf{S} \rangle$; (f) the power in watts that is incident on a rectangular surface measuring 20 m $\times$ 30 m at $z = 10$ m.

11.19 Perfectly conducting cylinders with radii of 8 mm and 20 mm are coaxial. The region between the cylinders is filled with a perfect dielectric for which $\epsilon = 10^{-9}/4\pi$ F/m and $\mu_r = 1$. If $\mathcal{E}$ in this region is $(500/\rho)\cos(\omega t - 4z)\mathbf{a}_\rho$ V/m, find (a) ω, with the help of Maxwell's equations in cylindrical coordinates; (b) $\mathcal{H}(\rho, z, t)$; (c) $\langle \mathbf{S}(\rho, z, t) \rangle$; (d) the average power passing through every cross section $8 < \rho < 20$ mm, $0 < \phi < 2\pi$.

11.20 Voltage breakdown in air at standard temperature and pressure occurs at an electric field strength of approximately 3×10^6 V/m. This becomes an issue

in some high-power optical experiments, in which tight focusing of light may be necessary. Estimate the lightwave power in watts that can be focused into a cylindrical beam of 10μm radius before breakdown occurs. Assume uniform plane wave behavior (although this assumption will produce an answer that is higher than the actual number by as much as a factor of 2, depending on the actual beam shape).

11.21 The cylindrical shell, 1 cm$< \rho < 1.2$ cm, is composed of a conducting material for which $\sigma = 10^6$ S/m. The external and internal regions are nonconducting. Let $H_\phi = 2000$ A/m at $\rho = 1.2$ cm. Find (*a*) **H** everywhere; (*b*) **E** everywhere; (*c*) $\langle \mathbf{S} \rangle$ everywhere.

11.22 The inner and outer dimensions of a coaxial copper transmission line are 2 and 7 mm, respectively. Both conductors have thicknesses much greater than δ. The dielectric is lossless and the operating frequency is 400 MHz. Calculate the resistance per meter length of the (*a*) inner conductor; (*b*) outer conductor; (*c*) transmission line.

11.23 A hollow tubular conductor is constructed from a type of brass having a conductivity of 1.2×10^7 S/m. The inner and outer radii are 9 and 10 mm, respectively. Calculate the resistance per meter length at a frequency of (*a*) dc; (*b*) 20 MHz; (*c*) 2 GHz.

11.24 (*a*) Most microwave ovens operate at 2.45 GHz. Assume that $\sigma = 1.2 \times 10^6$ S/m and $\mu_r = 500$ for the stainless steel interior, and find the depth of penetration. (*b*) Let $E_s = 50\angle 0°$ V/m at the surface of the conductor, and plot a curve of the amplitude of E_s versus the angle of E_s as the field propagates into the stainless steel.

11.25 A good conductor is planar in form, and it carries a uniform plane wave that has a wavelength of 0.3 mm and a velocity of 3×10^5 m/s. Assuming the conductor is nonmagnetic, determine the frequency and the conductivity.

11.26 The dimensions of a certain coaxial transmission line are $a = 0.8$ mm and $b = 4$ mm. The outer conductor thickness is 0.6 mm, and all conductors have $\sigma = 1.6 \times 10^7$ S/m. (*a*) Find R, the resistance per unit length at an operating frequency of 2.4 GHz. (*b*) Use information from Sections 6.3 and 8.10 to find C and L, the capacitance and inductance per unit length, respectively. The coax is air-filled. (*c*) Find α and β if $\alpha + j\beta = \sqrt{j\omega C(R + j\omega L)}$.

11.27 The planar surface $z = 0$ is a brass-Teflon interface. Use data available in Appendix C to evaluate the following ratios for a uniform plane wave having $\omega = 4 \times 10^{10}$ rad/s: (*a*) $\alpha_{\text{Tef}}/\alpha_{\text{brass}}$; (*b*) $\lambda_{\text{Tef}}/\lambda_{\text{brass}}$; (*c*) $v_{\text{Tef}}/v_{\text{brass}}$.

11.28 A uniform plane wave in free space has electric field vector given by $\mathbf{E}_s = 10e^{-j\beta x}\mathbf{a}_z + 15e^{-j\beta x}\mathbf{a}_y$ V/m. (*a*) Describe the wave polarization. (*b*) Find $\mathbf{H}_s$. (*c*) Determine the average power density in the wave in W/m^2.

11.29 Consider a left circularly polarized wave in free space that propagates in the forward z direction. The electric field is given by the appropriate form of Eq. (100). Determine (a) the magnetic field phasor, $\mathbf{H}_s$; (b) an expression for the average power density in the wave in W/m^2 by direct application of Eq. (77).

11.30 In an *anisotropic* medium, permittivity varies with electric field *direction*, and is a property seen in most crystals. Consider a uniform plane wave propagating in the z direction in such a medium, and which enters the material with equal field components along the x and y axes. The field phasor will take the form:

$$\mathbf{E}_s(z) = E_0(\mathbf{a}_x + \mathbf{a}_y \, e^{j\Delta\beta z}) e^{-j\beta z}$$

where $\Delta\beta = \beta_x - \beta_y$ is the difference in phase constants for waves that are linearly polarized in the x and y directions. Find distances into the material (in terms of $\Delta\beta$) at which the field is (a) linearly polarized and (b) circularly polarized. (c) Assume intrinsic impedance η that is approximately constant with field orientation and find $\mathbf{H}_s$ and $< \mathbf{S} >$.

11.31 A linearly polarized uniform plane wave, propagating in the forward z direction, is input to a lossless anisotropic material, in which the dielectric constant encountered by waves polarized along $y(\epsilon_{ry})$ differs from that seen by waves polarized along $x(\epsilon_{rx})$. Suppose $\epsilon_{rx} = 2.15$, $\epsilon_{ry} = 2.10$, and the wave electric field at input is polarized at $45°$ to the positive x and y axes. (a) Determine, in terms of the free space wavelength, λ, the shortest length of the material, such that the wave, as it emerges from the output, is circularly polarized. (b) Will the output wave be right or left circularly polarized? Problem 11.30 is good background.

11.32 Suppose that the length of the medium of Problem 11.31 is made to be *twice* that determined in the problem. Describe the polarization of the output wave in this case.

11.33 Given a wave for which $\mathbf{E}_s = 15e^{-j\beta z}\mathbf{a}_x + 18e^{-j\beta z}e^{j\phi}\mathbf{a}_y$ V/m in a medium characterized by complex intrinsic impedance, η (a) find $\mathbf{H}_s$; (b) determine the average power density in W/m^2.

11.34 Given a general elliptically polarized wave as per Eq. (93):

$$\mathbf{E}_s = [E_{x0}\mathbf{a}_x + E_{y0}e^{j\phi}\mathbf{a}_y]e^{-j\beta z}$$

(a) Show, using methods similar to those of Example 11.7, that a linearly polarized wave results when superimposing the given field and a phase-shifted field of the form:

$$\mathbf{E}_s = [E_{x0}\mathbf{a}_x + E_{y0}e^{-j\phi}\mathbf{a}_y]e^{-j\beta z}e^{j\delta}$$

where δ is a constant. (b) Find δ in terms of ϕ such that the resultant wave is linearly polarized along x.

12 CHAPTER

Plane Wave Reflection and Dispersion

I n Chapter 11, we learned how to mathematically represent uniform plane waves as functions of frequency, medium properties, and electric field orientation. We also learned how to calculate the wave velocity, attenuation, and power. In this chapter we consider wave reflection and transmission at planar boundaries between different media. Our study will allow any orientation between the wave and boundary and will also include the important cases of multiple boundaries. We will also study the practical case of waves that carry power over a finite band of frequencies, as would occur, for example, in a modulated carrier. We will consider such waves in *dispersive* media, in which some parameter that affects propagation (permittivity for example) varies with frequency. The effect of a dispersive medium on a signal is of great importance because the signal envelope will change its shape as it propagates. As a result, detection and faithful representation of the original signal at the receiving end become problematic. Consequently, dispersion and attenuation must both be evaluated when establishing maximum allowable transmission distances. ∎

12.1 REFLECTION OF UNIFORM PLANE WAVES AT NORMAL INCIDENCE

We first consider the phenomenon of reflection which occurs when a uniform plane wave is incident on the boundary between regions composed of two different materials. The treatment is specialized to the case of *normal incidence*—in which the wave propagation direction is perpendicular to the boundary. In later sections, we remove this restriction. Expressions will be found for the wave that is reflected from the interface and for that which is transmitted from one region into the other. These results are directly related to impedance-matching problems in ordinary transmission lines, as we have already encountered in Chapter 10. They are also applicable to waveguides, which we will study in Chapter 13.
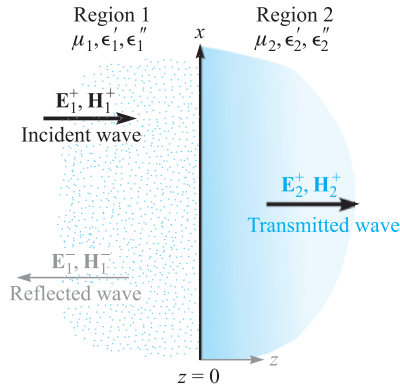
Figure 12.1 A plane wave incident on a boundary establishes reflected and transmitted waves having the indicated propagation directions. All fields are parallel to the boundary, with electric fields along x and magnetic fields along y.

We again assume that we have only a single vector component of the electric field intensity. Referring to Figure 12.1, we define region 1 (ϵ_1, μ_1) as the half-space for which $z < 0$; region 2 (ϵ_2, μ_2) is the half-space for which $z > 0$. Initially we establish a wave in region 1, traveling in the $+z$ direction, and linearly polarized along x.

$$\mathcal{E}_{x1}^+(z, t) = E_{x10}^+ e^{-\alpha_1 z} \cos(\omega t - \beta_1 z)$$

In phasor form, this is

$$E_{xs1}^+(z) = E_{x10}^+ e^{-jkz} \tag{1}$$

where we take E_{x10}^+ as real. The subscript 1 identifies the region, and the superscript $+$ indicates a positively traveling wave. Associated with $E_{xs1}^+(z)$ is a magnetic field in the y direction,

$$H_{ys1}^+(z) = \frac{1}{\eta_1} E_{x10}^+ e^{-jk_1 z} \tag{2}$$

where k_1 and η_1 are complex unless ϵ_1'' (or σ_1) is zero. This uniform plane wave in region 1 that is traveling toward the boundary surface at $z = 0$ is called the *incident* wave. Since the direction of propagation of the incident wave is perpendicular to the boundary plane, we describe it as normal incidence.

We now recognize that energy may be transmitted across the boundary surface at $z = 0$ into region 2 by providing a wave moving in the $+z$ direction in that medium. The phasor electric and magnetic fields for this wave are

$$E_{xs2}^+(z) = E_{x20}^+ e^{-jk_2 z} \tag{3}$$

$$H_{ys2}^+(z) = \frac{1}{\eta_2} E_{x20}^+ e^{-jk_2 z} \tag{4}$$

This wave, which moves away from the boundary surface into region 2, is called the *transmitted* wave. Note the use of the different propagation constant k_2 and intrinsic impedance η_2.

Now we must satisfy the boundary conditions at $z = 0$ with these assumed fields. With $\mathbf{E}$ polarized along x, the field is tangent to the interface, and therefore the $\mathbf{E}$ fields in regions 1 and 2 must be equal at $z = 0$. Setting $z = 0$ in (1) and (3) would require that $E_{x10}^+ = E_{x20}^+$. $\mathbf{H}$, being y-directed, is also a tangential field, and must be continuous across the boundary (no current sheets are present in real media). When we let $z = 0$ in (2) and (4), we find that we must have $E_{x10}^+/\eta_1 = E_{x20}^+/\eta_2$. Since $E_{x10}^+ = E_{x20}^+$, then $\eta_1 = \eta_2$. But this is a very special condition that does not fit the facts in general, and we are therefore unable to satisfy the boundary conditions with only an incident and a transmitted wave. We require a wave traveling away from the boundary in region 1, as shown in Figure 12.1; this is the *reflected* wave,

$$E_{xs1}^-(z) = E_{x10}^- e^{jk_1 z} \tag{5}$$

$$H_{xs1}^-(z) = -\frac{E_{x10}^-}{\eta_1} e^{jk_1 z} \tag{6}$$

where E_{x10}^- may be a complex quantity. Because this field is traveling in the $-z$ direction, $E_{xs1}^- = -\eta_1 H_{ys1}^-$ for the Poynting vector shows that $\mathbf{E}_1^- \times \mathbf{H}_1^-$ must be in the $-\mathbf{a}_z$ direction.

The boundary conditions are now easily satisfied, and in the process the amplitudes of the transmitted and reflected waves may be found in terms of E_{x10}^+. The total electric field intensity is continuous at $z = 0$,

$$E_{xs1} = E_{xs2} \qquad (z = 0)$$

or

$$E_{xs1}^+ + E_{xs1}^- = E_{xs2}^+ \qquad (z = 0)$$

Therefore

$$\boxed{E_{x10}^+ + E_{x10}^- = E_{x20}^+} \tag{7}$$

Furthermore,

$$H_{ys1} = H_{ys2} \qquad (z = 0)$$

or

$$H_{ys1}^+ + H_{ys1}^- = H_{ys2}^+ \qquad (z = 0)$$

and therefore

$$\boxed{\frac{E_{x10}^+}{\eta_1} - \frac{E_{x10}^-}{\eta_1} = \frac{E_{x20}^+}{\eta_2}} \tag{8}$$

Solving (8) for E_{x20}^+ and substituting into (7), we find

$$E_{x10}^+ + E_{x10}^- = \frac{\eta_2}{\eta_1} E_{x10}^+ - \frac{\eta_2}{\eta_1} E_{x10}^-$$

or

$$E_{x10}^{-} = E_{x10}^{+} \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1}$$

The ratio of the amplitudes of the reflected and incident electric fields defines the *reflection coefficient,* designated by Γ,

$$\Gamma = \frac{E_{x10}^{-}}{E_{x10}^{+}} = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} = |\Gamma| e^{j\phi} \tag{9}$$

It is evident that as η_1 or η_2 may be complex, Γ will also be complex, and so we include a reflective phase shift, ϕ. The interpretation of Eq. (9) is identical to that used with transmission lines [Eq. (73), Chapter 10].

The relative amplitude of the transmitted electric field intensity is found by combining (9) and (7) to yield the *transmission coefficient, τ*,

$$\tau = \frac{E_{x20}^{+}}{E_{x10}^{+}} = \frac{2\eta_2}{\eta_1 + \eta_2} = 1 + \Gamma = |\tau| e^{j\phi_i} \tag{10}$$

whose form and interpretation are consistent with the usage in transmission lines [Eq. (75), Chapter 10].

Let us see how these results may be applied to several special cases. We first let region 1 be a perfect dielectric and region 2 be a perfect conductor. Then we apply Eq. (48), Chapter 11, with $\epsilon_2'' = \sigma_2/\omega$, obtaining

$$\eta_2 = \sqrt{\frac{j\omega\mu_2}{\sigma_2 + j\omega\epsilon_2'}} = 0$$

in which zero is obtained since $\sigma_2 \to \infty$. Therefore, from (10),

$$E_{x20}^{+} = 0$$

No time-varying fields can exist in the perfect conductor. An alternate way of looking at this is to note that the skin depth is zero.

Because $\eta_2 = 0$, Eq. (9) shows that

$$\Gamma = -1$$

and

$$E_{x10}^{+} = -E_{x10}^{-}$$

The incident and reflected fields are of equal amplitude, and so all the incident energy is reflected by the perfect conductor. The fact that the two fields are of opposite sign indicates that at the boundary (or at the moment of reflection), the reflected field is shifted in phase by $180°$ relative to the incident field. The total **E** field in region 1 is

$$E_{xs1} = E_{xs1}^{+} + E_{xs1}^{-}$$
$$= E_{x10}^{+} e^{-j\beta_1 z} - E_{x10}^{+} e^{j\beta_1 z}$$

where we have let $jk_1 = 0 + j\beta_1$ in the perfect dielectric. These terms may be combined and simplified,

$$E_{xs1} = (e^{-j\beta_1 z} - e^{j\beta_1 z}) E_{x10}^+$$
$$= -j2 \sin(\beta_1 z) E_{x10}^+ \tag{11}$$

Multiplying (11) by $e^{j\omega t}$ and taking the real part, we obtain the real instantaneous form:

$$\mathcal{E}_{x1}(z, t) = 2E_{x10}^+ \sin(\beta_1 z) \sin(\omega t) \tag{12}$$

We recognize this total field in region 1 as a standing wave, obtained by combining two waves of equal amplitude traveling in opposite directions. We first encountered standing waves in transmission lines, but in the form of counterpropagating voltage waves (see Example 10.1).

Again, we compare the form of (12) to that of the incident wave,

$$\mathcal{E}_{x1}(z, t) = E_{x10}^+ \cos(\omega t - \beta_1 z) \tag{13}$$

Here we see the term $\omega t - \beta_1 z$ or $\omega(t - z/v_{p1})$, which characterizes a wave traveling in the $+z$ direction at a velocity $v_{p1} = \omega/\beta_1$. In (12), however, the factors involving time and distance are separate trigonometric terms. Whenever $\omega t = m\pi$, $\mathcal{E}_{x1}$ is zero at all positions. On the other hand, spatial nulls in the standing wave pattern occur for all times wherever $\beta_1 z = m\pi$, which in turn occurs when $m = (0, \pm 1, \pm 2, \ldots)$. In such cases,

$$\frac{2\pi}{\lambda_1} z = m\pi$$

and the null locations occur at

$$z = m\frac{\lambda_1}{2}$$

Thus $E_{x1} = 0$ at the boundary $z = 0$ and at every half-wavelength from the boundary in region 1, $z < 0$, as illustrated in Figure 12.2.

Because $E_{xs1}^+ = \eta_1 H_{ys1}^+$ and $E_{xs1}^- = -\eta_1 H_{ys1}^-$, the magnetic field is

$$H_{ys1} = \frac{E_{x10}^+}{\eta_1} (e^{-j\beta_1 z} + e^{j\beta_1 z})$$

or

$$H_{y1}(z, t) = 2\frac{E_{x10}^+}{\eta_1} \cos(\beta_1 z) \cos(\omega t) \tag{14}$$

This is also a standing wave, but it shows a maximum amplitude at the positions where $E_{x1} = 0$. It is also 90° out of time phase with E_{x1} everywhere. As a result, the average power as determined through the Poynting vector [Eq. (77), Chapter 11] is zero in the forward and backward directions.

Let us now consider perfect dielectrics in both regions 1 and 2; η_1 and η_2 are both real positive quantities and $\alpha_1 = \alpha_2 = 0$. Equation (9) enables us to calculate
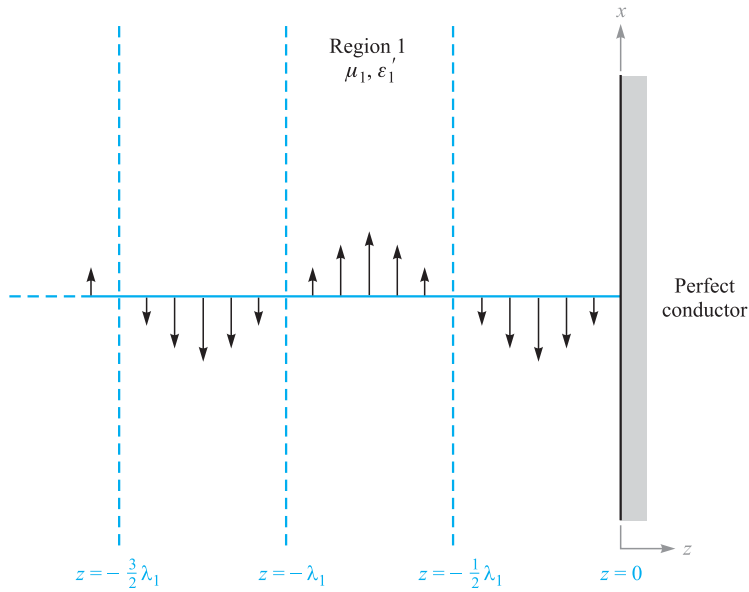
Figure 12.2 The instantaneous values of the total field E_{x1} are shown at $t = \pi/2$. $E_{x1} = 0$ for all time at multiples of one half-wavelength from the conducting surface.

the reflection coefficient and find E_{x1}^- in terms of the incident field E_{x1}^+. Knowing E_{x1}^+ and E_{x1}^-, we then find H_{y1}^+ and H_{y1}^-. In region 2, E_{x2}^+ is found from (10), and this then determines H_{y2}^+.

As a numerical example we select

$$\eta_1 = 100 \ \Omega$$

$$\eta_2 = 300 \ \Omega$$

$$E_{x10}^+ = 100 \ \text{V/m}$$

and calculate values for the incident, reflected, and transmitted waves.

Solution. The reflection coefficient is

$$\Gamma = \frac{300 - 100}{300 + 100} = 0.5$$

and thus

$$E_{x10}^- = 50 \ \text{V/m}$$

The magnetic field intensities are

$$H_{y10}^{+} = \frac{100}{100} = 1.00 \text{ A/m}$$

$$H_{y10}^{-} = -\frac{50}{100} = -0.50 \text{ A/m}$$

Using Eq. (77) from Chapter 11, we find that the magnitude of the average incident power density is

$$\langle S_{1i} \rangle = \left| \frac{1}{2} \text{Re}\{\mathbf{E}_s \times \mathbf{H}_s^*\} \right| = \frac{1}{2} E_{x10}^{+} H_{y10}^{+} = 50 \text{ W/m}^2$$

The average reflected power density is

$$\langle S_{1r} \rangle = -\frac{1}{2} E_{x10}^{-} H_{y10}^{-} = 12.5 \text{ W/m}^2$$

In region 2, using (10),

$$E_{x20}^{+} = \tau E_{x10}^{+} = 150 \text{ V/m}$$

and

$$H_{y20}^{+} = \frac{150}{300} = 0.500 \text{ A/m}$$

Therefore, the average power density that is transmitted through the boundary into region 2 is

$$\langle S_2 \rangle = \frac{1}{2} E_{x20}^{+} H_{y20}^{+} = 37.5 \text{ W/m}^2$$

We may check and confirm the power conservation requirement:

$$\langle S_{1i} \rangle = \langle S_{1r} \rangle + \langle S_2 \rangle$$

A general rule on the transfer of power through reflection and transmission can be formulated. We consider the same field vector and interface orientations as before, but allow for the case of complex impedances. For the incident power density, we have

$$\langle S_{1i} \rangle = \frac{1}{2} \text{Re}\{E_{xs1}^{+} H_{ys1}^{+*}\} = \frac{1}{2} \text{Re}\left\{ E_{x10}^{+} \frac{1}{\eta_1^*} E_{x10}^{+*} \right\} = \frac{1}{2} \text{Re}\left\{ \frac{1}{\eta_1^*} \right\} \left| E_{x10}^{+} \right|^2$$

The reflected power density is then

$$\langle S_{1r} \rangle = -\frac{1}{2} \text{Re}\{E_{xs1}^{-} H_{ys1}^{-*}\} = \frac{1}{2} \text{Re}\left\{ \Gamma E_{x10}^{+} \frac{1}{\eta_1^*} \Gamma^* E_{x10}^{+*} \right\} = \frac{1}{2} \text{Re}\left\{ \frac{1}{\eta_1^*} \right\} \left| E_{x10}^{+} \right|^2 |\Gamma|^2$$

We thus find the general relation between the reflected and incident power:

$$\boxed{\langle S_{1r} \rangle = |\Gamma|^2 \langle S_{1i} \rangle} \tag{15}$$

In a similar way, we find the transmitted power density:

$$\langle S_2 \rangle = \frac{1}{2} \text{Re}\{E_{xs2}^{+} H_{ys2}^{+*}\} = \frac{1}{2} \text{Re}\left\{ \tau E_{x10}^{+} \frac{1}{\eta_2^*} \tau^* E_{x10}^{+*} \right\} = \frac{1}{2} \text{Re}\left\{ \frac{1}{\eta_2^*} \right\} \left| E_{x10}^{+} \right|^2 |\tau|^2$$

and so we see that the incident and transmitted power densities are related through

$$\langle S_2 \rangle = \frac{\text{Re}\{1/\eta_2^*\}}{\text{Re}\{1/\eta_1^*\}} |\tau|^2 \langle S_{1i} \rangle = \left|\frac{\eta_1}{\eta_2}\right|^2 \left(\frac{\eta_2 + \eta_2^*}{\eta_1 + \eta_1^*}\right) |\tau|^2 \langle S_{1i} \rangle \quad (16)$$

Equation (16) is a relatively complicated way to calculate the transmitted power, unless the impedances are real. It is easier to take advantage of energy conservation by noting that whatever power is not reflected must be transmitted. Eq. (15) can be used to find

$$\boxed{\langle S_2 \rangle = (1 - |\Gamma|^2)\langle S_{1i} \rangle} \quad (17)$$

As would be expected (and which must be true), Eq. (17) can also be derived from Eq. (16).

D12.1. A 1-MHz uniform plane wave is normally incident onto a freshwater lake ($\epsilon_r' = 78$, $\epsilon_r'' = 0$, $\mu_r = 1$). Determine the fraction of the incident power that is (*a*) reflected and (*b*) transmitted. (*c*) Determine the amplitude of the electric field that is transmitted into the lake.

Ans. 0.63; 0.37; 0.20 V/m

12.2 STANDING WAVE RATIO

In cases where $|\Gamma| < 1$, some energy is transmitted into the second region and some is reflected. Region 1 therefore supports a field that is composed of both a traveling wave and a standing wave. We encountered this situation previously in transmission lines, in which partial reflection occurs at the load. Measurements of the voltage standing wave ratio and the locations of voltage minima or maxima enabled the determination of an unknown load impedance or established the extent to which the load impedance was matched to that of the line (Section 10.10). Similar measurements can be performed on the field amplitudes in plane wave reflection.

Using the same fields investigated in the previous section, we combine the incident and reflected electric field intensities. Medium 1 is assumed to be a perfect dielectric ($\alpha_1 = 0$), but region 2 may be any material. The total electric field phasor in region 1 will be

$$E_{x1T} = E_{x1}^+ + E_{x1}^- = E_{x10}^+ e^{-j\beta_1 z} + \Gamma E_{x10}^+ e^{j\beta_1 z} \quad (18)$$

where the reflection coefficient is as expressed in (9):

$$\Gamma = \frac{\eta_2 - \eta_1}{\eta_2 + \eta_1} = |\Gamma| e^{j\phi}$$

We allow for the possibility of a complex reflection coefficient by including its phase, ϕ. This is necessary because although η_1 is real and positive for a lossless medium,

η_2 will in general be complex. Additionally, if region 2 is a perfect conductor, η_2 is zero, and so ϕ is equal to π; if η_2 is real and less than η_1, ϕ is also equal to π; and if η_2 is real and greater than η_1, ϕ is zero.

Incorporating the phase of Γ into (18), the total field in region 1 becomes

$$E_{x1T} = \left(e^{-j\beta_1 z} + |\Gamma|e^{j(\beta_1 z + \phi)}\right)E_{x10}^+ \tag{19}$$

The maximum and minimum field amplitudes in (19) are z-dependent and are subject to measurement. Their ratio, as found for voltage amplitudes in transmission lines (Section 10.10), is the *standing wave ratio,* denoted by s. We have a maximum when each term in the larger parentheses in (19) has the same phase angle; so, for E_{x10}^+ positive and real,

$$|E_{x1T}|_{max} = (1 + |\Gamma|)E_{x10}^+ \tag{20}$$

and this occurs where

$$-\beta_1 z = \beta_1 z + \phi + 2m\pi \qquad (m = 0, \pm 1, \pm 2, \dots) \tag{21}$$

Therefore

$$z_{max} = -\frac{1}{2\beta_1}(\phi + 2m\pi) \tag{22}$$

Note that an electric field maximum is located at the boundary plane ($z = 0$) if $\phi = 0$; moreover, $\phi = 0$ when Γ is real and positive. This occurs for real η_1 and η_2 when $\eta_2 > \eta_1$. Thus there is a field maximum at the boundary surface when the intrinsic impedance of region 2 is greater than that of region 1 and both impedances are real. With $\phi = 0$, maxima also occur at $z_{max} = -m\pi/\beta_1 = -m\lambda_1/2$.

For the perfect conductor $\phi = \pi$, and these maxima are found at $z_{max} = -\pi/(2\beta_1)$, $-3\pi/(2\beta_1)$, or $z_{max} = -\lambda_1/4$, $-3\lambda_1/4$, and so forth.

The minima must occur where the phase angles of the two terms in the larger parentheses in (19) differ by 180°, thus

$$|E_{x1T}|_{min} = (1 - |\Gamma|)E_{x10}^+ \tag{23}$$

and this occurs where

$$-\beta_1 z = \beta_1 z + \phi + \pi + 2m\pi \qquad (m = 0, \pm 1, \pm 2, \dots) \tag{24}$$

or

$$z_{min} = -\frac{1}{2\beta_1}(\phi + (2m + 1)\pi) \tag{25}$$

The minima are separated by multiples of one half-wavelength (as are the maxima), and for the perfect conductor the first minimum occurs when $-\beta_1 z = 0$, or at the conducting surface. In general, an electric field minimum is found at $z = 0$ whenever $\phi = \pi$; this occurs if $\eta_2 < \eta_1$ and both are real. The results are mathematically identical to those found for the transmission line study in Section 10.10. Figure 10.6 in that chapter provides a visualization.

Further insights can be obtained by working with Eq. (19) and rewriting it in real instantaneous form. The steps are identical to those taken in Chapter 10, Eqs. (81) through (84). We find the total field in region 1 to be

$$
\mathcal{E}_{x1T}(z, t) = \underbrace{(1 - |\Gamma|)E_{x10}^{+} \cos(\omega t - \beta_1 z)}_{\text{traveling wave}}
$$
$$
+ \underbrace{2|\Gamma|E_{x10}^{+} \cos(\beta_1 z + \phi/2)\cos(\omega t + \phi/2)}_{\text{standing wave}} \qquad (26)
$$

The field expressed in Eq. (26) is the sum of a traveling wave of amplitude $(1 - |\Gamma|)E_{x10}^{+}$ and a standing wave having amplitude $2|\Gamma|E_{x10}^{+}$. The portion of the incident wave that reflects and back-propagates in region 1 interferes with an equivalent portion of the incident wave to form a standing wave. The rest of the incident wave (that does not interfere) is the traveling wave part of (26). The maximum amplitude observed in region 1 is found where the amplitudes of the two terms in (26) add directly to give $(1 + |\Gamma|)E_{x10}^{+}$. The minimum amplitude is found where the standing wave achieves a null, leaving only the traveling wave amplitude of $(1 - |\Gamma|)E_{x10}^{+}$. The fact that the two terms in (26) combine in this way with the proper phasing can be confirmed by substituting $z_{\max}$ and $z_{\min}$, as given by (22) and (25).

<div style="text-align:right">**EXAMPLE 12.2**</div>

To illustrate some of these results, let us consider a 100-V/m, 3-GHz wave that is propagating in a material having $\epsilon'_{r1} = 4$, $\mu_{r1} = 1$, and $\epsilon''_r = 0$. The wave is normally incident on another perfect dielectric in region 2, $z > 0$, where $\epsilon'_{r2} = 9$ and $\mu_{r2} = 1$ (Figure 12.3). We seek the locations of the maxima and minima of **E**.

Dielectric 1
$\epsilon'_{r1} = 4, \mu_{r1} = 1, \epsilon''_{r1} = 0$

Dielectric 2
$\epsilon'_{r2} = 9, \mu_{r2} = 1, \epsilon''_{r2} = 0$

$E_{xs1}^{+} = 100e^{-j40\pi z}$

$E_{xs1}^{-} = -20e^{j40\pi z}$

z
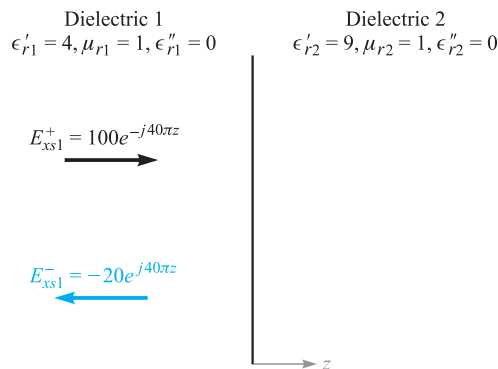
Figure 12.3 An incident wave, $E_{xs1}^{+} = 100e^{-j40\pi z}$ V/m, is reflected with a reflection coefficient $\Gamma = -0.2$. Dielectric 2 is infinitely thick.

Solution. We calculate $\omega = 6\pi \times 10^9$ rad/s, $\beta_1 = \omega\sqrt{\mu_1\epsilon_1} = 40\pi$ rad/m, and $\beta_2 = \omega\sqrt{\mu_2\epsilon_2} = 60\pi$ rad/m. Although the wavelength would be 10 cm in air, we find here that $\lambda_1 = 2\pi/\beta_1 = 5$ cm, $\lambda_2 = 2\pi/\beta_2 = 3.33$ cm, $\eta_1 = 60\pi$ Ω, $\eta_2 = 40\pi$ Ω, and $\Gamma = (\eta_2 - \eta_1)/(\eta_2 + \eta_1) = -0.2$. Because Γ is real and negative ($\eta_2 < \eta_1$), there will be a minimum of the electric field at the boundary, and it will be repeated at half-wavelength (2.5 cm) intervals in dielectric l. From (23), we see that $|E_{x1T}|_{\min} = 80$ V/m.

Maxima of **E** are found at distances of 1.25, 3.75, 6.25, ... cm from $z = 0$. These maxima all have amplitudes of 120 V/m, as predicted by (20).

There are no maxima or minima in region 2 because there is no reflected wave there.

The ratio of the maximum to minimum amplitudes is the standing wave ratio:

$$s = \frac{|E_{x1T}|_{\max}}{|E_{x1T}|_{\min}} = \frac{1 + |\Gamma|}{1 - |\Gamma|} \qquad (27)$$

Because $|\Gamma| < 1$, s is always positive and greater than or equal to unity. For the preceding example,

$$s = \frac{1 + |-0.2|}{1 - |-0.2|} = \frac{1.2}{0.8} = 1.5$$

If $|\Gamma| = 1$, the reflected and incident amplitudes are equal, all the incident energy is reflected, and s is infinite. Planes separated by multiples of $\lambda_1/2$ can be found on which E_{x1} is zero at all times. Midway between these planes, E_{x1} has a maximum amplitude twice that of the incident wave.

If $\eta_2 = \eta_1$, then $\Gamma = 0$, no energy is reflected, and $s = 1$; the maximum and minimum amplitudes are equal.

If one-half the incident power is reflected, $|\Gamma|^2 = 0.5$, $|\Gamma| = 0.707$, and $s = 5.83$.

D12.2. What value of s results when $\Gamma = \pm 1/2$?

Ans. 3

Because the standing wave ratio is a ratio of amplitudes, the relative amplitudes, as measured by a probe, permit its use to determine s experimentally.

EXAMPLE 12.3

A uniform plane wave in air partially reflects from the surface of a material whose properties are unknown. Measurements of the electric field in the region in front of the interface yield a 1.5-m spacing between maxima, with the first maximum occurring 0.75 m from the interface. A standing wave ratio of 5 is measured. Determine the intrinsic impedance, η_u, of the unknown material.

Solution. The 1.5 m spacing between maxima is $\lambda/2$, which implies that a wavelength is 3.0 m, or $f = 100$ MHz. The first maximum at 0.75 m is thus at a distance of $\lambda/4$ from the interface, which means that a field minimum occurs at the boundary. Thus Γ will be real and negative. We use (27) to write

$$|\Gamma| = \frac{s-1}{s+1} = \frac{5-1}{5+1} = \frac{2}{3}$$

So

$$\Gamma = -\frac{2}{3} = \frac{\eta_u - \eta_0}{\eta_u + \eta_0}$$

which we solve for η_u to obtain

$$\eta_u = \frac{1}{5}\eta_0 = \frac{377}{5} = 75.4\ \Omega$$

12.3 WAVE REFLECTION FROM MULTIPLE INTERFACES

So far we have treated the reflection of waves at the single boundary that occurs between semi-infinite media. In this section, we consider wave reflection from materials that are finite in extent, such that we must consider the effect of the front and back surfaces. Such a two-interface problem would occur, for example, for light incident on a flat piece of glass. Additional interfaces are present if the glass is coated with one or more layers of dielectric material for the purpose (as we will see) of reducing reflections. Such problems in which more than one interface is involved are frequently encountered; single-interface problems are in fact more the exception than the rule.

Consider the general situation shown in Figure 12.4, in which a uniform plane wave propagating in the forward z direction is normally incident from the left onto the interface between regions 1 and 2; these have intrinsic impedances η_1 and η_2. A third region of impedance η_3 lies beyond region 2, and so a second interface exists between regions 2 and 3. We let the second interface location occur at $z = 0$, and so all positions to the left will be described by values of z that are negative. The width of the second region is l, so the first interface will occur at position $z = -l$.

When the incident wave reaches the first interface, events occur as follows: A portion of the wave reflects, while the remainder is transmitted, to propagate toward the second interface. There, a portion is transmitted into region 3, while the rest reflects and returns to the first interface; there it is again partially reflected. This reflected wave then combines with additional transmitted energy from region 1, and the process repeats. We thus have a complicated sequence of multiple reflections that occur within region 2, with partial transmission at each bounce. To analyze the situation in this way would involve keeping track of a very large number of reflections; this would be necessary when studying the *transient* phase of the process, where the incident wave first encounters the interfaces.

If the incident wave is left on for all time, however, a *steady-state* situation is eventually reached, in which (1) an overall fraction of the incident wave is reflected
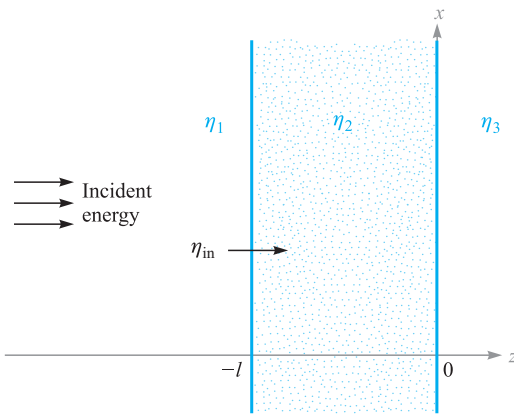
Figure 12.4 Basic two-interface problem, in which the impedances of regions 2 and 3, along with the finite thickness of region 2, are accounted for in the input impedance at the front surface, η_{in}.

from the two-interface configuration and back-propagates in region 1 with a definite amplitude and phase; (2) an overall fraction of the incident wave is transmitted through the two interfaces and forward-propagates in the third region; (3) a net backward wave exists in region 2, consisting of all reflected waves from the second interface; and (4) a net forward wave exists in region 2, which is the superposition of the transmitted wave through the first interface and all waves in region 2 that have reflected from the first interface and are now forward-propagating. The effect of combining many co-propagating waves in this way is to establish a single wave which has a definite amplitude and phase, determined through the sums of the amplitudes and phases of all the component waves. In steady state, we thus have a total of five waves to consider. These are the incident and net reflected waves in region 1, the net transmitted wave in region 3, and the two counterpropagating waves in region 2.

The situation is analyzed in the same manner as that used in the analysis of finite-length transmission lines (Section 10.11). Let us assume that all regions are composed of lossless media, and consider the two waves in region 2. If we take these as x-polarized, their electric fields combine to yield

$$E_{xs2} = E_{x20}^{+} e^{-j\beta_2 z} + E_{x20}^{-} e^{j\beta_2 z} \tag{28a}$$

where $\beta_2 = \omega\sqrt{\epsilon_{r2}}/c$, and where the amplitudes, E_{x20}^{+} and E_{x20}^{-}, are complex. The y-polarized magnetic field is similarly written, using complex amplitudes:

$$H_{ys2} = H_{y20}^{+} e^{-j\beta_2 z} + H_{y20}^{-} e^{j\beta_2 z} \tag{28b}$$

We now note that the forward and backward electric field amplitudes in region 2 are related through the reflection coefficient at the second interface, Γ_{23}, where

$$\Gamma_{23} = \frac{\eta_3 - \eta_2}{\eta_3 + \eta_2} \tag{29}$$

We thus have

$$E_{x20}^- = \Gamma_{23} E_{x20}^+ \qquad (30)$$

We then write the magnetic field amplitudes in terms of electric field amplitudes through

$$H_{y20}^+ = \frac{1}{\eta_2} E_{x20}^+ \qquad (31a)$$

and

$$H_{y20}^- = -\frac{1}{\eta_2} E_{x20}^- = -\frac{1}{\eta_2} \Gamma_{23} E_{x20}^+ \qquad (31b)$$

We now define the *wave impedance, η_w*, as the z-dependent ratio of the total electric field to the total magnetic field. In region 2, this becomes, using (28a) and (28b),

$$\eta_w(z) = \frac{E_{xs2}}{H_{ys2}} = \frac{E_{x20}^+ e^{-j\beta_2 z} + E_{x20}^- e^{j\beta_2 z}}{H_{y20}^+ e^{-j\beta_2 z} + H_{y20}^- e^{j\beta_2 z}}$$

Then, using (30), (31a), and (31b), we obtain

$$\eta_w(z) = \eta_2 \left[\frac{e^{-j\beta_2 z} + \Gamma_{23} e^{j\beta_2 z}}{e^{-j\beta_2 z} - \Gamma_{23} e^{j\beta_2 z}} \right]$$

Now, using (29) and Euler's identity, we have

$$\eta_w(z) = \eta_2 \times \frac{(\eta_3 + \eta_2)(\cos \beta_2 z - j \sin \beta_2 z) + (\eta_3 - \eta_2)(\cos \beta_2 z + j \sin \beta_2 z)}{(\eta_3 + \eta_2)(\cos \beta_2 z - j \sin \beta_2 z) - (\eta_3 - \eta_2)(\cos \beta_2 z + j \sin \beta_2 z)}$$

This is easily simplified to yield

$$\eta_w(z) = \eta_2 \frac{\eta_3 \cos \beta_2 z - j \eta_2 \sin \beta_2 z}{\eta_2 \cos \beta_2 z - j \eta_3 \sin \beta_2 z} \qquad (32)$$

We now use the wave impedance in region 2 to solve our reflection problem. Of interest to us is the net reflected wave amplitude at the first interface. Since tangential **E** and **H** are continuous across the boundary, we have

$$E_{xs1}^+ + E_{xs1}^- = E_{xs2} \qquad (z = -l) \qquad (33a)$$

and

$$H_{ys1}^+ + H_{ys1}^- = H_{ys2} \qquad (z = -l) \qquad (33b)$$

Then, in analogy to (7) and (8), we may write

$$E_{x10}^+ + E_{x10}^- = E_{xs2}(z = -l) \qquad (34a)$$

and

$$\frac{E_{x10}^+}{\eta_1} - \frac{E_{x10}^-}{\eta_1} = \frac{E_{xs2}(z = -l)}{\eta_w(-l)} \qquad (34b)$$

where E_{x10}^+ and E_{x10}^- are the amplitudes of the incident and reflected fields. We call $\eta_w(-l)$ the *input impedance, η_{in}*, to the two-interface combination. We now solve

(34a) and (34b) together, eliminating E_{xs2}, to obtain

$$\frac{E_{x10}^{-}}{E_{x10}^{+}} = \Gamma = \frac{\eta_{in} - \eta_1}{\eta_{in} + \eta_1} \tag{35}$$

To find the input impedance, we evaluate (32) at $z = -l$, resulting in

$$\eta_{in} = \eta_2 \frac{\eta_3 \cos \beta_2 l + j\eta_2 \sin \beta_2 l}{\eta_2 \cos \beta_2 l + j\eta_3 \sin \beta_2 l} \tag{36}$$

Equations (35) and (36) are general results that enable us to calculate the net reflected wave amplitude and phase from two parallel interfaces between lossless media.[1] Note the dependence on the interface spacing, l, and on the wavelength as measured in region 2, characterized by β_2. Of immediate importance to us is the fraction of the incident power that reflects from the dual interface and back-propagates in region 1. As we found earlier, this fraction will be $|\Gamma|^2$. Also of interest is the transmitted power, which propagates away from the second interface in region 3. It is simply the remaining power fraction, which is $1 - |\Gamma|^2$. The power in region 2 stays constant in steady state; power leaves that region to form the reflected and transmitted waves, but is immediately replenished by the incident wave. We have already encountered an analogous situation involving cascaded transmission lines, which culminated in Eq. (101) in Chapter 10.

   An important result of situations involving two interfaces is that it is possible to achieve total transmission in certain cases. From (35), we see that total transmission occurs when $\Gamma = 0$, or when $\eta_{in} = \eta_1$. In this case, as in transmission lines, we say that the input impedance is *matched* to that of the incident medium. There are a few methods of accomplishing this.

   As a start, suppose that $\eta_3 = \eta_1$, and region 2 is of such thickness that $\beta_2 l = m\pi$, where m is an integer. Now $\beta_2 = 2\pi/\lambda_2$, where λ_2 is the wavelength *as measured in region 2*. Therefore

$$\frac{2\pi}{\lambda_2}l = m\pi$$

or

$$l = m\frac{\lambda_2}{2} \tag{37}$$

With $\beta_2 l = m\pi$, the second region thickness is an integer multiple of half-wavelengths as measured in that medium. Equation (36) now reduces to $\eta_{in} = \eta_3$. Thus the general effect of a multiple half-wave thickness is to render the second region immaterial to

---

[1] For convenience, (34a) and (34b) have been written for a specific time at which the incident wave amplitude, E_{x10}^{+}, occurs at $z = -l$. This establishes a zero-phase reference at the front interface for the incident wave, and so it is from this reference that the reflected wave phase is determined. Equivalently, we have repositioned the $z = 0$ point at the front interface. Eq. (36) allows this because it is only a function of the interface spacing, l.

the results on reflection and transmission. Equivalently, we have a single-interface problem involving η_1 and η_3. Now, with $\eta_3 = \eta_1$, we have a matched input impedance, and there is no net reflected wave. This method of choosing the region 2 thickness is known as *half-wave matching*. Its applications include, for example, antenna housings on airplanes known as *radomes,* which form a part of the fuselage. The antenna, inside the aircraft, can transmit and receive through this layer, which can be shaped to enable good aerodynamic characteristics. Note that the half-wave matching condition no longer applies as we deviate from the wavelength that satisfies it. When this is done, the device reflectivity increases (with increased wavelength deviation), so it ultimately acts as a bandpass filter.

Often, it is convenient to express the dielectric constant of the medium through the *refractive index* (or just index), n, defined as

$$n = \sqrt{\epsilon_r} \tag{38}$$

Characterizing materials by their refractive indices is primarily done at optical frequencies (on the order of 10^{14} Hz), whereas at much lower frequencies, a dielectric constant is traditionally specified. Since ϵ_r is complex in lossy media, the index will also be complex. Rather than complicate the situation in this way, we will restrict our use of the refractive index to cases involving lossless media, having $\epsilon_r'' = 0$, and $\mu_r = 1$. Under lossless conditions, we may write the plane wave phase constant and the material intrinsic impedance in terms of the index through

$$\beta = k = \omega\sqrt{\mu_0\epsilon_0}\sqrt{\epsilon_r} = \frac{n\omega}{c} \tag{39}$$

and

$$\eta = \frac{1}{\sqrt{\epsilon_r}}\sqrt{\frac{\mu_0}{\epsilon_0}} = \frac{\eta_0}{n} \tag{40}$$

Finally, the phase velocity and wavelength in a material of index n are

$$v_p = \frac{c}{n} \tag{41}$$

and

$$\lambda = \frac{v_p}{f} = \frac{\lambda_0}{n} \tag{42}$$

where λ_0 is the wavelength in free space. It is obviously important not to confuse the index n with the similar-appearing Greek η (intrinsic impedance), which has an entirely different meaning.

Another application, typically seen in optics, is the *Fabry-Perot interferometer*. This, in its simplest form, consists of a single block of glass or other transparent

material of index n, whose thickness, l, is set to transmit wavelengths which satisfy the condition $\lambda = \lambda_0/n = 2l/m$. Often we want to transmit only one wavelength, not several, as (37) would allow. We would therefore like to assure that adjacent wavelengths that are passed through the device are separated as far as possible, so that only one will lie within the input power spectrum. In terms of wavelength as measured in the material, this separation is in general given by

$$\lambda_{m-1} - \lambda_m = \Delta\lambda_f = \frac{2l}{m-1} - \frac{2l}{m} = \frac{2l}{m(m-1)} \doteq \frac{2l}{m^2}$$

Note that m is the number of half-wavelengths in region 2, or $m = 2l/\lambda = 2nl/\lambda_0$, where λ_0 is the desired free-space wavelength for transmission. Thus

$$\Delta\lambda_f \doteq \frac{\lambda_2^2}{2l} \tag{43a}$$

In terms of wavelength measured in free space, this becomes

$$\Delta\lambda_{f0} = n\Delta\lambda_f \doteq \frac{\lambda_0^2}{2nl} \tag{43b}$$

$\Delta\lambda_{f0}$ is known as the *free spectral range* of the Fabry-Perot interferometer in terms of free-space wavelength separation. The interferometer can be used as a narrow-band filter (transmitting a desired wavelength and a narrow spectrum around this wavelength) if the spectrum to be filtered is narrower than the free spectral range.

EXAMPLE 12.4

Suppose we wish to filter an optical spectrum of full width $\Delta\lambda_{s0} = 50$ nm (measured in free space), whose center wavelength, λ_0, is in the red part of the visible spectrum at 600 nm, where one nm (nanometer) is 10^{-9} m. A Fabry-Perot filter is to be used, consisting of a lossless glass plate in air, having refractive index $n = 1.45$. We need to find the required range of glass thicknesses such that multiple wavelength orders will not be transmitted.

Solution. We require that the free spectral range be greater than the optical spectral width, or $\Delta\lambda_{f0} > \Delta\lambda_s$. Using (43b)

$$l < \frac{\lambda_0^2}{2n\,\Delta\lambda_{s0}}$$

So

$$l < \frac{600^2}{2(1.45)(50)} = 2.5 \times 10^3 \text{nm} = 2.5\,\mu\text{m}$$

where 1μm (micrometer) $= 10^{-6}$ m. Fabricating a glass plate of this thickness or less is somewhat ridiculous to contemplate. Instead, what is often used is an airspace of thickness on this order, between two thick plates whose surfaces on the sides opposite the airspace are antireflection coated. This is in fact a more versatile configuration because the wavelength to be transmitted (and the free spectral range) can be adjusted by varying the plate separation.

Next, we remove the restriction $\eta_1 = \eta_3$ and look for a way to produce zero reflection. Returning to Eq. (36), suppose we set $\beta_2 l = (2m - 1)\pi/2$, or an odd multiple of $\pi/2$. This means that

$$\frac{2\pi}{\lambda_2}l = (2m - 1)\frac{\pi}{2} \qquad (m = 1, 2, 3, \ldots)$$

or

$$l = (2m - 1)\frac{\lambda_2}{4} \tag{44}$$

The thickness is an odd multiple of a quarter-wavelength as measured in region 2. Under this condition, (36) reduces to

$$\eta_{\text{in}} = \frac{\eta_2^2}{\eta_3} \tag{45}$$

Typically, we choose the second region impedance to allow matching between given impedances η_1 and η_3. To achieve total transmission, we require that $\eta_{\text{in}} = \eta_1$, so that the required second region impedance becomes

$$\eta_2 = \sqrt{\eta_1 \eta_3} \tag{46}$$

With the conditions given by (44) and (46) satisfied, we have performed *quarter-wave matching*. The design of antireflective coatings for optical devices is based on this principle.

We wish to coat a glass surface with an appropriate dielectric layer to provide total transmission from air to the glass at a free-space wavelength of 570 nm. The glass has refractive index $n_3 = 1.45$. Determine the required index for the coating and its minimum thickness.

Solution. The known impedances are $\eta_1 = 377 \, \Omega$ and $\eta_3 = 377/1.45 = 260 \, \Omega$. Using (46) we have

$$\eta_2 = \sqrt{(377)(260)} = 313 \, \Omega$$

The index of region 2 will then be

$$n_2 = \left(\frac{377}{313}\right) = 1.20$$

The wavelength in region 2 will be

$$\lambda_2 = \frac{570}{1.20} = 475 \, \text{nm}$$

The minimum thickness of the dielectric layer is then

$$l = \frac{\lambda_2}{4} = 119 \, \text{nm} = 0.119 \, \mu\text{m}$$
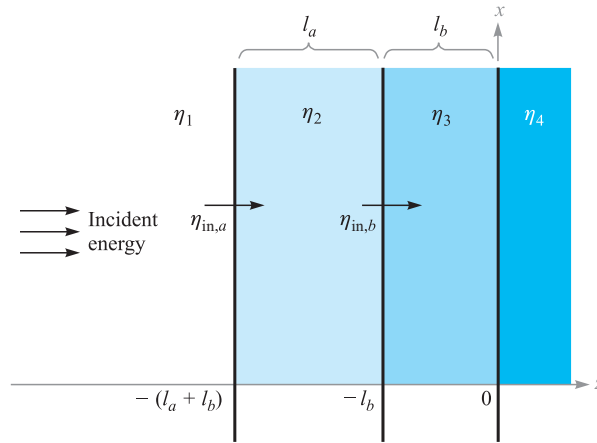
Figure 12.5 A three-interface problem in which input impedance $\eta_{\text{in},a}$ is transformed back to the front interface to form input impedance $\eta_{\text{in},b}$.

The procedure in this section for evaluating wave reflection has involved calculating an effective impedance at the first interface, η_{in}, which is expressed in terms of the impedances that lie beyond the front surface. This process of *impedance transformation* is more apparent when we consider problems involving more than two interfaces.

For example, consider the three-interface situation shown in Figure 12.5, where a wave is incident from the left in region 1. We wish to determine the fraction of the incident power that is reflected and back-propagates in region 1 and the fraction of the incident power that is transmitted into region 4. To do this, we need to find the input impedance at the front surface (the interface between regions 1 and 2). We start by transforming the impedance of region 4 to form the input impedance at the boundary between regions 2 and 3. This is shown as $\eta_{\text{in},b}$ in Figure 12.5. Using (36), we have

$$\eta_{\text{in},b} = \eta_3 \frac{\eta_4 \cos \beta_3 l_b + j\eta_3 \sin \beta_3 l_b}{\eta_3 \cos \beta_3 l_b + j\eta_4 \sin \beta_3 l_b} \tag{47}$$

We have now effectively reduced the situation to a two-interface problem in which $\eta_{\text{in},b}$ is the impedance of all that lies beyond the second interface. The input impedance at the front interface, $\eta_{\text{in},a}$, is now found by transforming $\eta_{\text{in},b}$ as follows:

$$\eta_{\text{in},a} = \eta_2 \frac{\eta_{\text{in},b} \cos \beta_2 l_a + j\eta_2 \sin \beta_2 l_a}{\eta_2 \cos \beta_2 l_a + j\eta_{\text{in},b} \sin \beta_2 l_a} \tag{48}$$

The reflected power fraction is now $|\Gamma|^2$, where

$$\Gamma = \frac{\eta_{\text{in},a} - \eta_1}{\eta_{\text{in},a} + \eta_1}$$

The fraction of the power transmitted into region 4 is, as before, $1 - |\Gamma|^2$. The method of impedance transformation can be applied in this manner to any number of inter- faces. The process, although tedious, is easily handled by a computer.

The motivation for using multiple layers to reduce reflection is that the resulting structure is less sensitive to deviations from the design wavelength if the impedances (or refractive indices) are arranged to progressively increase or decrease from layer to layer. For multiple layers to antireflection coat a camera lens, for example, the layer on the lens surface would be of impedance very close to that of the glass. Subsequent layers are given progressively higher impedances. With a large number of layers fabricated in this way, the situation begins to approach (but never reaches) the ideal case, in which the top layer impedance matches that of air, while the impedances of deeper layers continuously decrease until reaching the value of the glass surface. With this continuously varying impedance, there is no surface from which to reflect, and so light of any wavelength is totally transmitted. Multilayer coatings designed in this way produce excellent broadband transmission characteristics.

> **D12.3.** A uniform plane wave in air is normally incident on a dielectric slab of thickness $\lambda_2/4$, and intrinsic impedance $\eta_2 = 260\ \Omega$. Determine the magnitude and phase of the reflection coefficient.
>
> **Ans.** $0.356;\ 180°$

12.4 PLANE WAVE PROPAGATION IN GENERAL DIRECTIONS

In this section, we will learn how to mathematically describe uniform plane waves that propagate in any direction. Our motivation for doing this is our need to address the problem of incident waves on boundaries that are not perpendicular to the prop- agation direction. Such problems of *oblique incidence* generally occur, with normal incidence being a special case. Addressing such problems requires (as always) that we establish an appropriate coordinate system. With the boundary positioned in the x, y plane, for example, the incident wave will propagate in a direction that could involve all three coordinate axes, whereas with normal incidence, we were only concerned with propagation along z. We need a mathematical formalism that will allow for the general direction case.

Let us consider a wave that propagates in a lossless medium, with propagation constant $\beta = k = \omega\sqrt{\mu\epsilon}$. For simplicity, we consider a two-dimensional case, where the wave travels in a direction between the x and z axes. The first step is to consider the propagation constant as a *vector,* $\mathbf{k}$, indicated in Figure 12.6. The direction of $\mathbf{k}$ is the propagation direction, which is the same as the direction of the Poynting vector in our case.[2] The magnitude of $\mathbf{k}$ is the phase shift per unit distance *along that direction*.

---

[2] We assume here that the wave is in an isotropic medium, where the permittivity and permeability do not change with field orientation. In anisotropic media (where ϵ and/or μ depend on field orientation), the directions of the Poynting vector and $\mathbf{k}$ may differ.
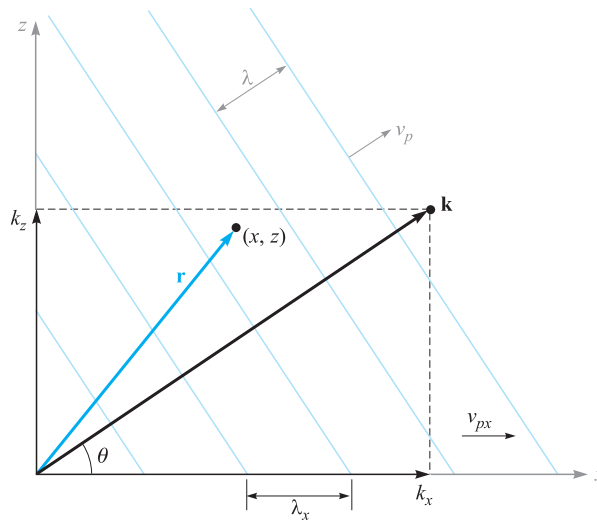
Figure 12.6   Representation of a uniform plane wave with wavevector **k** at angle θ to the x axis. The phase at point (x, z) is given by **k** · **r**. Planes of constant phase (shown as lines perpendicular to **k**) are spaced by wavelength λ but have wider spacing when measured along the x or z axis.

Part of the process of characterizing a wave involves specifying its phase at any spatial location. For the waves we have considered that propagate along the z axis, this was accomplished by the factor $e^{\pm jkz}$ in the phasor form. To specify the phase in our two-dimensional problem, we make use of the vector nature of **k** and consider the phase at a general location (x, z) described through the position vector **r**. The phase at that location, referenced to the origin, is given by the projection of **k** along **r** times the magnitude of **r**, or just **k** · **r**. If the electric field is of magnitude E_0, we can thus write down the phasor form of the wave in Figure 12.6 as

$$\mathbf{E}_s = \mathbf{E}_0 e^{-j\mathbf{k} \cdot \mathbf{r}} \tag{49}$$

The minus sign in the exponent indicates that the phase along **r** moves in time in the direction of increasing **r**. Again, the wave power flow in an isotropic medium occurs in the direction along which the phase shift per unit distance is maximum—or along **k**. The vector **r** serves as a means to measure phase at any point using **k**. This construction is easily extended to three dimensions by allowing **k** and **r** to each have three components.

In our two-dimensional case of Figure 12.6, we can express **k** in terms of its x and z components:

$$\mathbf{k} = k_x \mathbf{a}_x + k_z \mathbf{a}_z$$

The position vector, **r**, can be similarly expressed:

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

so that

$$\mathbf{k} \cdot \mathbf{r} = k_x x + k_z z$$

Equation (49) now becomes

$$\mathbf{E}_s = \mathbf{E}_0 e^{-j(k_x x + k_z z)} \tag{50}$$

Whereas Eq. (49) provided the general form of the wave, Eq. (50) is the form that is specific to the situation. Given a wave expressed by (50), the angle of propagation from the x axis is readily found through

$$\theta = \tan^{-1}\left(\frac{k_z}{k_x}\right)$$

The wavelength and phase velocity depend on the direction one is considering. In the direction of **k**, these will be

$$\lambda = \frac{2\pi}{k} = \frac{2\pi}{\left(k_x^2 + k_z^2\right)^{1/2}}$$

and

$$v_p = \frac{\omega}{k} = \frac{\omega}{\left(k_x^2 + k_z^2\right)^{1/2}}$$

If, for example, we consider the x direction, these quantities will be

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

and

$$v_{px} = \frac{\omega}{k_x}$$

Note that both λ_x and v_{px} are greater than their counterparts along the direction of **k**. This result, at first surprising, can be understood through the geometry of Figure 12.6. The diagram shows a series of phase fronts (planes of constant phase) which intersect **k** at right angles. The phase shift between adjacent fronts is set at 2π in the figure; this corresponds to a spatial separation along the **k** direction of one wavelength, as shown. The phase fronts intersect the x axis, and we see that *along x* the front separation is greater than it was along **k**. λ_x is the spacing between fronts along x and is indicated

on the figure. The phase velocity along x is the velocity of the intersection points between the phase fronts and the x axis. Again, from the geometry, we see that this velocity must be faster than the velocity along $\mathbf{k}$ and will, of course, exceed the speed of light in the medium. This does not constitute a violation of special relativity, however, since the energy in the wave flows in the direction of $\mathbf{k}$ and not along x or z. The wave frequency is $f = \omega/2\pi$ and is invariant with direction. Note, for example, that in the directions we have considered,

$$f = \frac{v_p}{\lambda} = \frac{v_{px}}{\lambda_x} = \frac{\omega}{2\pi}$$

EXAMPLE 12.6

Consider a 50-MHz uniform plane wave having electric field amplitude 10 V/m. The medium is lossless, having $\epsilon_r = \epsilon'_r = 9.0$ and $\mu_r = 1.0$. The wave propagates in the x, y plane at a 30° angle to the x axis and is linearly polarized along z. Write down the phasor expression for the electric field.

Solution. The propagation constant magnitude is

$$k = \omega\sqrt{\mu\epsilon} = \frac{\omega\sqrt{\epsilon_r}}{c} = \frac{2\pi \times 50 \times 10^6(3)}{3 \times 10^8} = 3.2 \text{ m}^{-1}$$

The vector $\mathbf{k}$ is now

$$\mathbf{k} = 3.2(\cos 30\mathbf{a}_x + \sin 30\mathbf{a}_y) = 2.8\mathbf{a}_x + 1.6\mathbf{a}_y \text{ m}^{-1}$$

Then

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

With the electric field directed along z, the phasor form becomes

$$\mathbf{E}_s = E_0 e^{-j\mathbf{k}\cdot\mathbf{r}}\,\mathbf{a}_z = 10e^{-j(2.8x+1.6y)}\,\mathbf{a}_z$$

D12.4. For Example 12.6, calculate λ_x, λ_y, v_{px}, and v_{py}.

Ans. 2.2 m; 3.9 m; 1.1×10^8 m/s; 2.0×10^8 m/s

12.5 PLANE WAVE REFLECTION AT OBLIQUE INCIDENCE ANGLES

We now consider the problem of wave reflection from plane interfaces, in which the incident wave propagates at some angle to the surface. Our objectives are (1) to determine the relation between incident, reflected, and transmitted angles, and (2) to derive reflection and transmission coefficients that are functions of the incident angle and wave polarization. We will also show that cases exist in which total reflection or total transmission may occur at the interface between two dielectrics if the angle of incidence and the polarization are appropriately chosen.
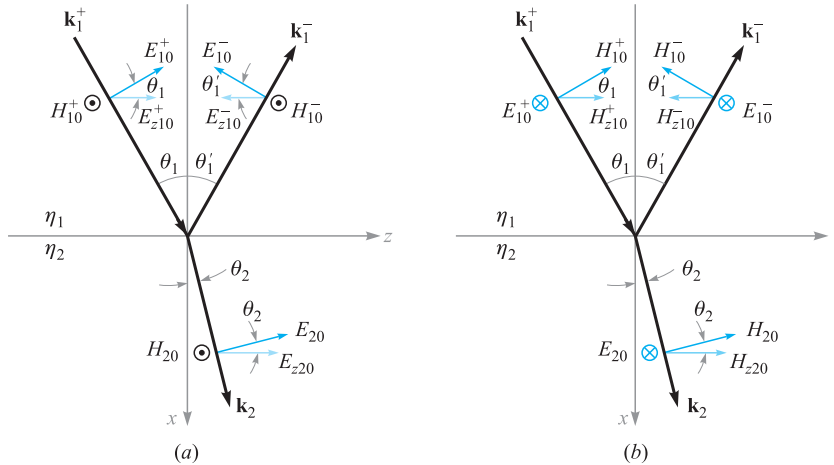
Figure 12.7 Geometries for plane wave incidence at angle θ_1 onto an interface between dielectrics having intrinsic impedances η_1 and η_2. The two polarization cases are shown: (a) p-polarization (or TM), with **E** in the plane of incidence; (b) s-polarization (or TE), with **E** perpendicular to the plane of incidence.

The situation is illustrated in Figure 12.7, in which the incident wave direction and position-dependent phase are characterized by wavevector $\mathbf{k}_1^+$. The angle of incidence is the angle between $\mathbf{k}_1^+$ and a line that is normal to the surface (the x axis in this case). The incidence angle is shown as θ_1. The reflected wave, characterized by wavevector $\mathbf{k}_1^-$, will propagate away from the interface at angle θ_1'. Finally, the transmitted wave, characterized by $\mathbf{k}_2$, will propagate into the second region at angle θ_2 as shown. One would suspect (from previous experience) that the incident and reflected angles are equal ($\theta_1 = \theta_1'$), which is correct. We need to show this, however, to be complete.

The two media are lossless dielectrics, characterized by intrinsic impedances η_1 and η_2. We will assume, as before, that the materials are nonmagnetic, and thus have permeability μ_0. Consequently, the materials are adequately described by specifying their dielectric constants, ϵ_{r1} and ϵ_{r2}, or their refractive indices, $n_1 = \sqrt{\epsilon_{r1}}$ and $n_2 = \sqrt{\epsilon_{r2}}$.

In Figure 12.7, two cases are shown that differ by the choice of electric field orientation. In Figure 12.7a, the **E** field is polarized in the plane of the page, with **H** therefore perpendicular to the page and pointing outward. In this illustration, the plane of the page is also the *plane of incidence,* which is more precisely defined as the plane spanned by the incident **k** vector and the normal to the surface. With **E** lying in the plane of incidence, the wave is said to have *parallel polarization* or to be *p-polarized* (**E** is parallel to the incidence plane). Note that although **H** is perpendicular to the incidence plane, it lies parallel (or transverse) to the interface. Consequently, another name for this type of polarization is *transverse magnetic,* or TM polarization.

Figure 12.7b shows the situation in which the field directions have been rotated by 90°. Now **H** lies in the plane of incidence, whereas **E** is perpendicular to the plane. Because **E** is used to define polarization, the configuration is called *perpendicular*

polarization, or is said to be *s-polarized*.[3] $\mathbf{E}$ is also parallel to the interface, and so the case is also called *transverse electric,* or TE polarization. We will find that the reflection and transmission coefficients will differ for the two polarization types, but that reflection and transmission angles will not depend on polarization. We only need to consider s- and p-polarizations because any other field direction can be constructed as some combination of s and p waves.

Our desired knowledge of reflection and transmission coefficients, as well as how the angles relate, can be found through the field boundary conditions at the interface. Specifically, we require that the transverse components of $\mathbf{E}$ and $\mathbf{H}$ be continuous across the interface. These were the conditions we used to find Γ and τ for normal incidence ($\theta_1 = 0$), which is in fact a special case of our current problem. We will consider the case of p-polarization (Figure 12.7a) first. To begin, we write down the incident, reflected, and transmitted fields in phasor form, using the notation developed in Section 12.4:

$$\mathbf{E}_{s1}^+ = \mathbf{E}_{10}^+ e^{-j\mathbf{k}_1^+ \cdot \mathbf{r}} \tag{51}$$

$$\mathbf{E}_{s1}^- = \mathbf{E}_{10}^- e^{-j\mathbf{k}_1^- \cdot \mathbf{r}} \tag{52}$$

$$\mathbf{E}_{s2} = \mathbf{E}_{20} e^{-j\mathbf{k}_2 \cdot \mathbf{r}} \tag{53}$$

where

$$\mathbf{k}_1^+ = k_1(\cos\theta_1\, \mathbf{a}_x + \sin\theta_1\, \mathbf{a}_z) \tag{54}$$

$$\mathbf{k}_1^- = k_1(-\cos\theta_1'\, \mathbf{a}_x + \sin\theta_1'\, \mathbf{a}_z) \tag{55}$$

$$\mathbf{k}_2 = k_2(\cos\theta_2\, \mathbf{a}_x + \sin\theta_2\, \mathbf{a}_z) \tag{56}$$

and where

$$\mathbf{r} = x\, \mathbf{a}_x + z\, \mathbf{a}_z \tag{57}$$

The wavevector magnitudes are $k_1 = \omega\sqrt{\epsilon_{r1}}/c = n_1\omega/c$ and $k_2 = \omega\sqrt{\epsilon_{r2}}/c = n_2\omega/c$.

Now, to evaluate the boundary condition that requires continuous tangential electric field, we need to find the components of the electric fields (z components) that are parallel to the interface. Projecting all $\mathbf{E}$ fields in the z direction, and using (51) through (57), we find

$$E_{zs1}^+ = E_{z10}^+ e^{-j\mathbf{k}_1^+ \cdot \mathbf{r}} = E_{10}^+ \cos\theta_1 e^{-jk_1(x\cos\theta_1 + z\sin\theta_1)} \tag{58}$$

$$E_{zs1}^- = E_{z10}^- e^{-j\mathbf{k}_1^- \cdot \mathbf{r}} = E_{10}^- \cos\theta_1' e^{jk_1(x\cos\theta_1' - z\sin\theta_1')} \tag{59}$$

$$E_{zs2} = E_{z20} e^{-j\mathbf{k}_2 \cdot \mathbf{r}} = E_{20} \cos\theta_2 e^{-jk_2(x\cos\theta_2 + z\sin\theta_2)} \tag{60}$$

---

[3] The *s* designation is an abbreviation for the German *senkrecht,* meaning *perpendicular*. The *p* in *p-polarized* is an abbreviation for the German word for parallel, which is *parallel*.

The boundary condition for a continuous tangential electric field now reads:

$$E_{zs1}^+ + E_{zs1}^- = E_{zs2} \quad (\text{at } x = 0)$$

We now substitute Eqs. (58) through (60) into (61) and evaluate the result at $x = 0$ to obtain

$$E_{10}^+ \cos\theta_1 \, e^{-jk_1 z \sin\theta_1} + E_{10}^- \cos\theta_1' \, e^{-jk_1 z \sin\theta_1'} = E_{20} \cos\theta_2 \, e^{-jk_2 z \sin\theta_2} \quad (61)$$

Note that E_{10}^+, E_{10}^-, and E_{20} are all constants (independent of z). Further, we require that (61) hold for all values of z (everywhere on the interface). For this to occur, it must follow that all the phase terms appearing in (61) are equal. Specifically,

$$k_1 z \sin\theta_1 = k_1 z \sin\theta_1' = k_2 z \sin\theta_2$$

From this, we see immediately that $\theta_1' = \theta_1$, or the angle of reflection is equal to the angle of incidence. We also find that

$$\boxed{k_1 \sin\theta_1 = k_2 \sin\theta_2} \quad (62)$$

Equation (62) is known as *Snell's law of refraction*. Because, in general, $k = n\omega/c$, we can rewrite (62) in terms of the refractive indices:

$$\boxed{n_1 \sin\theta_1 = n_2 \sin\theta_2} \quad (63)$$

Equation (63) is the form of Snell's law that is most readily used for our present case of nonmagnetic dielectrics. Equation (62) is a more general form which would apply, for example, to cases involving materials with different permeabilities as well as different permittivities. In general, we would have $k_1 = (\omega/c)\sqrt{\mu_{r1}\epsilon_{r1}}$ and $k_2 = (\omega/c)\sqrt{\mu_{r2}\epsilon_{r2}}$.

Having found the relations between angles, we next turn to our second objective, which is to determine the relations between the amplitudes, E_{10}^+, E_{10}^-, and E_{20}. To accomplish this, we need to consider the other boundary condition, requiring tangential continuity of **H** at $x = 0$. The magnetic field vectors for the p-polarized wave are all negative y-directed. At the boundary, the field amplitudes are related through

$$H_{10}^+ + H_{10}^- = H_{20} \quad (64)$$

Then, when we use the fact that $\theta_1' = \theta_1$ and invoke Snell's law, (61) becomes

$$E_{10}^+ \cos\theta_1 + E_{10}^- \cos\theta_1 = E_{20} \cos\theta_2 \quad (65)$$

Using the medium intrinsic impedances, we know, for example, that $E_{10}^+/H_{10}^+ = \eta_1$ and $E_{20}^+/H_{20}^+ = \eta_2$. Eq. (64) can be written as follows:

$$\frac{E_{10}^+ \cos\theta_1}{\eta_{1p}} - \frac{E_{10}^- \cos\theta_1}{\eta_{1p}} = \frac{E_{20}^+ \cos\theta_2}{\eta_{2p}} \quad (66)$$

Note the minus sign in front of the second term in (66), which results from the fact that $E_{10}^- \cos\theta_1$ is negative (from Figure 12.7a), whereas H_{10}^- is positive (again from the figure). When we write Eq. (66), *effective impedances,* valid for p-polarization,

are defined through

$$\eta_{1p} = \eta_1 \cos \theta_1 \qquad (67)$$

and

$$\eta_{2p} = \eta_2 \cos \theta_2 \qquad (68)$$

Using this representation, Eqs. (65) and (66) are now in a form that enables them to be solved together for the ratios E_{10}^-/E_{10}^+ and E_{20}/E_{10}^+. Performing analogous procedures to those used in solving (7) and (8), we find the reflection and transmission coefficients:

$$\Gamma_p = \frac{E_{10}^-}{E_{10}^+} = \frac{\eta_{2p} - \eta_{1p}}{\eta_{2p} + \eta_{1p}} \qquad (69)$$

$$\tau_p = \frac{E_{20}}{E_{10}^+} = \frac{2\eta_{2p}}{\eta_{2p} + \eta_{1p}} \left(\frac{\cos \theta_1}{\cos \theta_2} \right) \qquad (70)$$

A similar procedure can be carried out for s-polarization, referring to Figure 12.7b. The details are left as an exercise; the results are

$$\Gamma_s = \frac{E_{y10}^-}{E_{y10}^+} = \frac{\eta_{2s} - \eta_{1s}}{\eta_{2s} + \eta_{1s}} \qquad (71)$$

$$\tau_s = \frac{E_{y20}}{E_{y10}^+} = \frac{2\eta_{2s}}{\eta_{2s} + \eta_{1s}} \qquad (72)$$

where the effective impedances for s-polarization are

$$\eta_{1s} = \eta_1 \sec \theta_1 \qquad (73)$$

and

$$\eta_{2s} = \eta_2 \sec \theta_2 \qquad (74)$$

Equations (67) through (74) are what we need to calculate wave reflection and transmission for either polarization, and at any incident angle.

EXAMPLE 12.7

A uniform plane wave is incident from air onto glass at an angle from the normal of 30°. Determine the fraction of the incident power that is reflected and transmitted for (a) p-polarization and (b) s-polarization. Glass has refractive index $n_2 = 1.45$.

Solution. First, we apply Snell's law to find the transmission angle. Using $n_1 = 1$ for air, we use (63) to find

$$\theta_2 = \sin^{-1}\left(\frac{\sin 30}{1.45}\right) = 20.2°$$

Now, for p-polarization:

$$\eta_{1p} = \eta_1 \cos 30 = (377)(.866) = 326\ \Omega$$

$$\eta_{2p} = \eta_2 \cos 20.2 = \frac{377}{1.45}(.938) = 244\ \Omega$$

Then, using (69), we find

$$\Gamma_p = \frac{244 - 326}{244 + 326} = -0.144$$

The fraction of the incident power that is reflected is

$$\frac{P_r}{P_{inc}} = |\Gamma_p|^2 = .021$$

The transmitted fraction is then

$$\frac{P_t}{P_{inc}} = 1 - |\Gamma_p|^2 = .979$$

For s-polarization, we have

$$\eta_{1s} = \eta_1 \sec 30 = 377/.866 = 435\ \Omega$$

$$\eta_{2s} = \eta_2 \sec 20.2 = \frac{377}{1.45(.938)} = 277\ \Omega$$

Then, using (71):

$$\Gamma_s = \frac{277 - 435}{277 + 435} = -.222$$

The reflected power fraction is thus

$$|\Gamma_s|^2 = .049$$

The fraction of the incident power that is transmitted is

$$1 - |\Gamma_s|^2 = .951$$

In Example 12.7, reflection coefficient values for the two polarizations were found to be negative. The meaning of a negative reflection coefficient is that the component of the reflected electric field that is parallel to the interface will be directed opposite the incident field component when both are evaluated at the boundary.

This effect is also observed when the second medium is a perfect conductor. In this case, we know that the electric field inside the conductor must be zero. Consequently, $\eta_2 = E_{20}/H_{20} = 0$, and the reflection coefficients will be $\Gamma_p = \Gamma_s = -1$. Total reflection occurs, regardless of the incident angle or polarization.

12.6 TOTAL REFLECTION AND TOTAL TRANSMISSION OF OBLIQUELY INCIDENT WAVES

Now that we have methods available to us for solving problems involving oblique incidence reflection and transmission, we can explore the special cases of *total reflection* and *total transmission*. We look for special combinations of media, incidence angles, and polarizations that produce these properties. To begin, we identify the necessary condition for total reflection. We want total *power* reflection, so that $|\Gamma|^2 = \Gamma\Gamma^* = 1$, where Γ is either Γ_p or Γ_s. The fact that this condition involves the possibility of a complex Γ allows some flexibility. For the incident medium, we note that η_{1p} and η_{1s} will always be real and positive. On the other hand, when we consider the second medium, η_{2p} and η_{2s} involve factors of $\cos\theta_2$ or $1/\cos\theta_2$, where

$$\cos\theta_2 = \left[1 - \sin^2\theta_2\right]^{1/2} = \left[1 - \left(\frac{n_1}{n_2}\right)^2 \sin^2\theta_1\right]^{1/2} \tag{75}$$

where Snell's law has been used. We observe that $\cos\theta_2$, and hence η_{2p} and η_{2s}, become imaginary whenever $\sin\theta_1 > n_2/n_1$. Let us consider parallel polarization, for example. Under conditions of imaginary η_{2p}, (69) becomes

$$\Gamma_p = \frac{j|\eta_{2p}| - \eta_{1p}}{j|\eta_{2p}| + \eta_{1p}} = -\frac{\eta_{1p} - j|\eta_{2p}|}{\eta_{1p} + j|\eta_{2p}|} = -\frac{Z}{Z^*}$$

where $Z = \eta_{1p} - j|\eta_{2p}|$. *We can therefore see that $\Gamma_p\Gamma_p^* = 1$, meaning total power reflection, whenever η_{2p} is imaginary.* The same will be true whenever η_{2p} is zero, which will occur when $\sin\theta_1 = n_2/n_1$. We thus have our condition for total reflection, which is

$$\sin\theta_1 \geq \frac{n_2}{n_1} \tag{76}$$

From this condition arises the *critical angle* of total reflection, θ_c, defined through

$$\sin\theta_c = \frac{n_2}{n_1} \tag{77}$$

The total reflection condition can thus be more succinctly written as

$$\theta_1 \geq \theta_c \quad \text{(for total reflection)} \tag{78}$$

Note that for (76) and (77) to make sense, it must be true that $n_2 < n_1$, or the wave must be incident from a medium of higher refractive index than that of the medium beyond the boundary. For this reason, the total reflection condition is sometimes called total *internal* reflection; it is often seen (and applied) in optical devices such
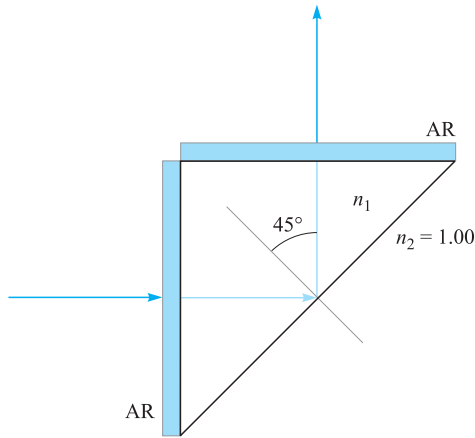
Figure 12.8 Beam-steering prism for Example 12.8.

as beam-steering prisms, where light within the glass structure totally reflects from glass-air interfaces.

<div style="text-align:right">**EXAMPLE 12.8**</div>

A prism is to be used to turn a beam of light by $90°$, as shown in Figure 12.8. Light enters and exits the prism through two antireflective (AR-coated) surfaces. Total reflection is to occur at the back surface, where the incident angle is $45°$ to the normal. Determine the minimum required refractive index of the prism material if the surrounding region is air.

Solution. Considering the back surface, the medium beyond the interface is air, with $n_2 = 1.00$. Because $\theta_1 = 45°$, (76) is used to obtain

$$n_1 \geq \frac{n_2}{\sin 45} = \sqrt{2} = 1.41$$

Because fused silica glass has refractive index $n_g = 1.45$, it is a suitable material for this application and is in fact widely used.

Another important application of total reflection is in *optical waveguides*. These, in their simplest form, are constructed of three layers of glass, in which the middle layer has a slightly higher refractive index than the outer two. Figure 12.9 shows the basic structure. Light, propagating from left to right, is confined to the middle layer by total reflection at the two interfaces, as shown. Optical fiber waveguides are constructed on this principle, in which a cylindrical glass core region of small radius is surrounded coaxially by a lower-index cladding glass material of larger radius. Basic waveguiding principles as applied to metallic and dielectric structures will be presented in Chapter 13.
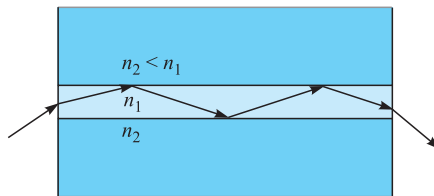
Figure 12.9 A dielectric slab waveguide (symmetric case), showing light confinement to the center material by total reflection.

We next consider the possibility of *total transmission*. In this case, the requirement is simply that $\Gamma = 0$. We investigate this possibility for the two polarizations. First, we consider s-polarization. If $\Gamma_s = 0$, then from (71) we require that $\eta_{2s} = \eta_{1s}$, or

$$\eta_2 \sec \theta_2 = \eta_1 \sec \theta_1$$

Using Snell's law to write θ_2 in terms of θ_1, the preceding equation becomes

$$\eta_2 \left[1 - \left(\frac{n_1}{n_2} \right)^2 \sin^2 \theta_1 \right]^{-1/2} = \eta_1 \left[1 - \sin^2 \theta_1 \right]^{-1/2}$$

There is no value of θ_1 that will satisfy this, so we turn instead to p-polarization. Using (67), (68), and (69), with Snell's law, we find that the condition for $\Gamma_p = 0$ is

$$\eta_2 \left[1 - \left(\frac{n_1}{n_2} \right)^2 \sin^2 \theta_1 \right]^{1/2} = \eta_1 \left[1 - \sin^2 \theta_1 \right]^{1/2}$$

This equation does have a solution, which is

$$\sin \theta_1 = \sin \theta_B = \frac{n_2}{\sqrt{n_1^2 + n_2^2}} \qquad (79)$$

where we have used $\eta_1 = \eta_0 / n_1$ and $\eta_2 = \eta_0 / n_2$. We call this special angle θ_B, where total transmission occurs, the *Brewster angle* or *polarization angle*. The latter name comes from the fact that if light having both s- and p-polarization components is incident at $\theta_1 = \theta_B$, the p component will be totally transmitted, leaving the partially reflected light entirely s-polarized. At angles that are slightly off the Brewster angle, the reflected light is still predominantly s-polarized. Most reflected light that we see originates from horizontal surfaces (such as the surface of the ocean), and so the light has mostly horizontal polarization. Polaroid sunglasses take advantage of this fact to reduce glare, for they are made to block the transmission of horizontally polarized light while passing light that is vertically polarized.

EXAMPLE 12.9

Light is incident from air to glass at Brewster's angle. Determine the incident and transmitted angles.

Solution. Because glass has refractive index $n_2 = 1.45$, the incident angle will be

$$\theta_1 = \theta_B = \sin^{-1}\left(\frac{n_2}{\sqrt{n_1^2 + n_2^2}}\right) = \sin^{-1}\left(\frac{1.45}{\sqrt{1.45^2 + 1}}\right) = 55.4°$$

The transmitted angle is found from Snell's law, through

$$\theta_2 = \sin^{-1}\left(\frac{n_1}{n_2}\sin\theta_B\right) = \sin^{-1}\left(\frac{n_1}{\sqrt{n_1^2 + n_2^2}}\right) = 34.6°$$

Note from this exercise that $\sin\theta_2 = \cos\theta_B$, which means that the sum of the incident and refracted angles at the Brewster condition is always $90°$.

Many of the results we have seen in this section are summarized in Figure 12.10, in which Γ_p and Γ_s, from (69) and (71), are plotted as functions of the incident angle, θ_1. Curves are shown for selected values of the refractive index ratio, n_1/n_2. For all plots in which $n_1/n_2 > 1$, Γ_s and Γ_p achieve values of ± 1 at the critical angle. At larger angles, the reflection coefficients become imaginary (and are not shown) but nevertheless retain magnitudes of unity. The occurrence of the Brewster angle is evident in the curves for Γ_p (Figure 12.10a) because all curves cross the θ_1 axis. This behavior is not seen in the Γ_s functions because Γ_s is positive for all values of θ_1 when $n_1/n_2 > 1$.

> **D12.5.** In Example 12.9, calculate the reflection coefficient for s-polarized light.
>
> **Ans.** -0.355

12.7 WAVE PROPAGATION IN DISPERSIVE MEDIA

In Chapter 11, we encountered situations in which the complex permittivity of the medium depends on frequency. This is true in all materials through a number of possible mechanisms. One of these, mentioned earlier, is that oscillating bound charges in a material are in fact harmonic oscillators that have resonant frequencies associated with them (see Appendix D). When the frequency of an incoming electromagnetic wave is at or near a bound charge resonance, the wave will induce strong oscillations; these in turn have the effect of depleting energy from the wave in its original form. The wave thus experiences absorption, and it does so to a greater extent than it would at a frequency that is detuned from resonance. A related effect is that the
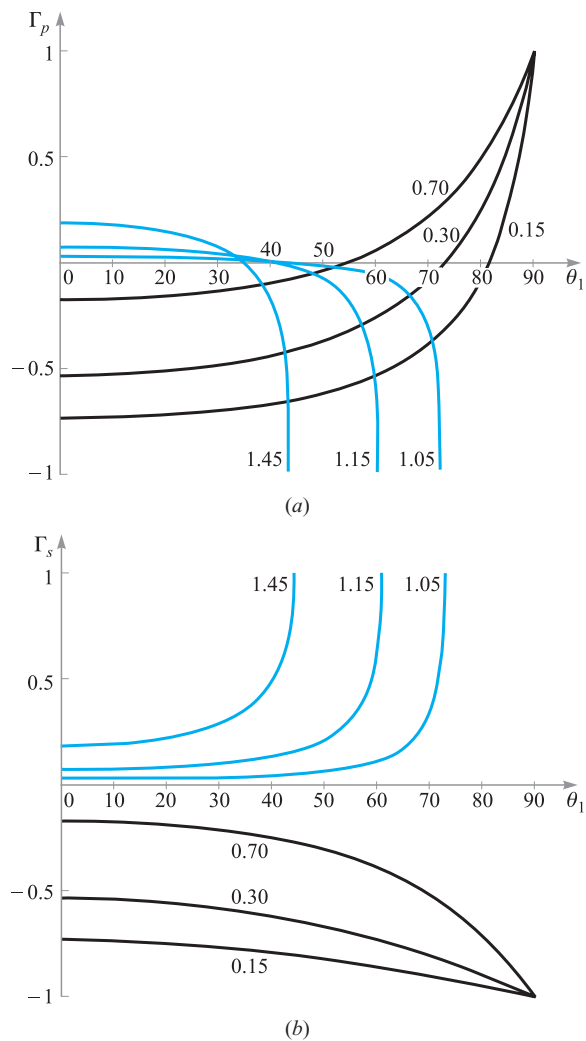
Figure 12.10 (a) Plots of Γ_p [Eq. (69)] as functions of the incident angle, θ_1, as shown in Figure 12.7a. Curves are shown for selected values of the refractive index ratio, n_1/n_2. Both media are lossless and have $\mu_r = 1$. Thus $\eta_1 = \eta_0/n_1$ and $\eta_2 = \eta_0/n_2$. (b) Plots of Γ_s [Eq. (71)] as functions of the incident angle, θ_1, as shown in Figure 12.7b. As in Figure 12.10a, the media are lossless, and curves are shown for selected n_1/n_2.
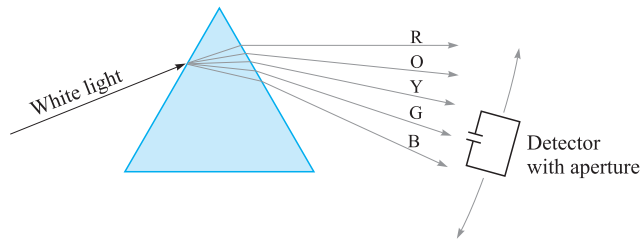
Figure 12.11 The angular dispersion of a prism can be measured using a movable device which measures both wavelength and power. The device senses light through a small aperture, thus improving wavelength resolution.

real part of the dielectric constant will be different at frequencies near resonance than at frequencies far from resonance. In short, resonance effects give rise to values of ϵ' and ϵ'' that will vary continuously with frequency. These in turn will produce a fairly complicated frequency dependence in the attenuation and phase constants as expressed in Eqs. (44) and (45) in Chapter 11.

This section concerns the effect of a frequency-varying dielectric constant (or refractive index) on a wave as it propagates in an otherwise lossless medium. This situation arises quite often because significant refractive index variation can occur at frequencies far away from resonance, where absorptive losses are negligible. A classic example of this is the separation of white light into its component colors by a glass prism. In this case, the frequency-dependent refractive index results in different angles of refraction for the different colors—hence the separation. The color separation effect produced by the prism is known as *angular dispersion,* or more specifically, *chromatic* angular dispersion.

The term *dispersion* implies a *separation* of distinguishable components of a wave. In the case of the prism, the components are the various colors that have been spatially separated. An important point here is that the spectral *power* has been dispersed by the prism. We can illustrate this idea by considering what it would take to measure the difference in refracted angles between, for example, blue and red light. One would need to use a power detector with a very narrow aperture, as shown in Figure 12.11. The detector would be positioned at the locations of the blue and red light from the prism, with the narrow aperture allowing essentially one color at a time (or light over a very narrow spectral range) to pass through to the detector. The detector would then measure the power in what we could call a "spectral packet," or a very narrow slice of the total power spectrum. The smaller the aperture, the narrower the spectral width of the packet, and the greater the precision in the measurement.[4] It

---

[4] To perform this experiment, one would need to measure the wavelength as well. To do this, the detector would likely be located at the output of a spectrometer or monochromer whose input slit performs the function of the bandwidth-limiting aperture.
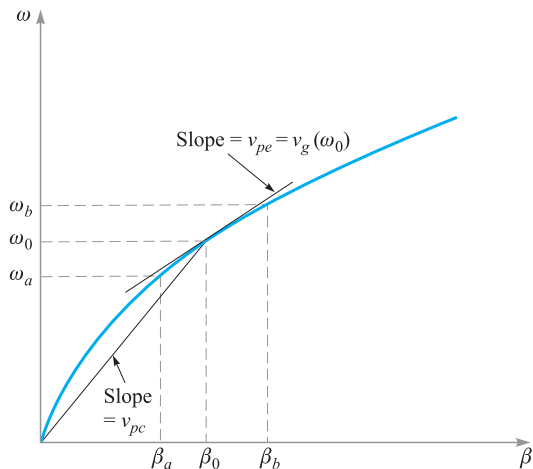
Figure 12.12 ω-β diagram for a material in which the refractive index increases with frequency. The slope of a line tangent to the curve at ω_0 is the group velocity at that frequency. The slope of a line joining the origin to the point on the curve at ω_0 is the phase velocity at ω_0.

is important for us to think of wave power as subdivided into spectral packets in this way because it will figure prominently in our interpretation of the main topic of this section, which is wave dispersion *in time*.

We now consider a lossless nonmagnetic medium in which the refractive index varies with frequency. The phase constant of a uniform plane wave in this medium will assume the form

$$\beta(\omega) = k = \omega\sqrt{\mu_0\epsilon(\omega)} = n(\omega)\frac{\omega}{c} \tag{80}$$

If we take $n(\omega)$ to be a monotonically increasing function of frequency (as is usually the case), a plot of ω versus β would look something like the curve shown in Figure 12.12. Such a plot is known as an ω-β *diagram* for the medium. Much can be learned about how waves propagate in the material by considering the shape of the ω-β curve.

Suppose we have two waves at two frequencies, ω_a and ω_b, which are co-propagating in the material and whose amplitudes are equal. The two frequencies are labeled on the curve in Figure 12.12, along with the frequency midway between the two, ω_0. The corresponding phase constants, β_a, β_b, and β_0, are also labeled. The electric fields of the two waves are linearly polarized in the same direction (along x, for example), while both waves propagate in the forward z direction. The waves will thus interfere with each other, producing a resultant wave whose field function can be found simply by adding the **E** fields of the two waves. This addition is done using

the complex fields:

$$E_{c,\text{net}}(z, t) = E_0[e^{-j\beta_a z}e^{j\omega_a t} + e^{-j\beta_b z}e^{j\omega_b t}]$$

Note that we must use the full complex forms (with frequency dependence retained) as opposed to the phasor forms, since the waves are at different frequencies. Next, we factor out the term $e^{-j\beta_0 z}e^{j\omega_0 t}$:

$$E_{c,\text{net}}(z, t) = E_0 e^{-j\beta_0 z}e^{j\omega_0 t}[e^{j\Delta\beta z}e^{-j\Delta\omega t} + e^{-j\Delta\beta z}e^{j\Delta\omega t}]$$

$$= 2E_0 e^{-j\beta_0 z}e^{j\omega_0 t}\cos(\Delta\omega t - \Delta\beta z) \tag{81}$$

where

$$\Delta\omega = \omega_0 - \omega_a = \omega_b - \omega_0$$

and

$$\Delta\beta = \beta_0 - \beta_a = \beta_b - \beta_0$$

The preceding expression for $\Delta\beta$ is approximately true as long as $\Delta\omega$ is small. This can be seen from Figure 12.12 by observing how the shape of the curve affects $\Delta\beta$, given uniform frequency spacings.

The real instantaneous form of (81) is found through

$$\mathcal{E}_{\text{net}}(z, t) = \text{Re}\{E_{c,\text{net}}\} = 2E_0 \cos(\Delta\omega t - \Delta\beta z)\cos(\omega_0 t - \beta_0 z) \tag{82}$$

If $\Delta\omega$ is fairly small compared to ω_0, we recognize (82) as a carrier wave at frequency ω_0 that is sinusoidally modulated at frequency $\Delta\omega$. The two original waves are thus "beating" together to form a slow modulation, as one would hear when the same note is played by two slightly out-of-tune musical instruments. The resultant wave is shown in Figure 12.13.
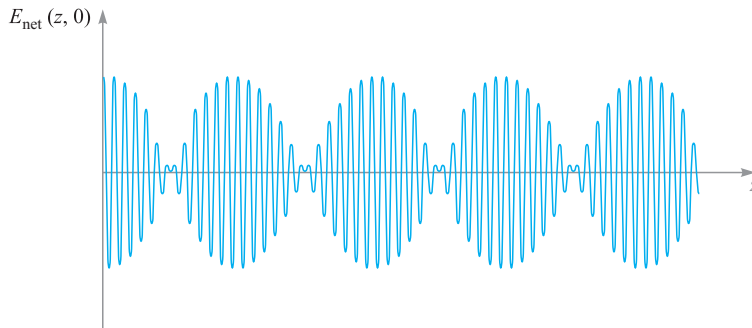
Figure 12.13 Plot of the total electric field strength as a function of z (with $t = 0$) of two co-propagating waves having different frequencies, ω_a and ω_b, as per Eq. (81). The rapid oscillations are associated with the carrier frequency, $\omega_0 = (\omega_a + \omega_b)/2$. The slower modulation is associated with the envelope or "beat" frequency, $\Delta\omega = (\omega_b - \omega_a)/2$.

Of interest to us are the phase velocities of the carrier wave and the modulation envelope. From (82), we can immediately write these down as:

$$v_{pc} = \frac{\omega_0}{\beta_0} \qquad \text{(carrier velocity)} \tag{83}$$

$$v_{pe} = \frac{\Delta\omega}{\Delta\beta} \qquad \text{(envelope velocity)} \tag{84}$$

Referring to the ω-β diagram, Figure 12.12, we recognize the carrier phase velocity as the slope of the straight line that joins the origin to the point on the curve whose coordinates are ω_0 and β_0. We recognize the envelope velocity as a quantity that approximates the slope of the ω-β curve at the location of an operation point specified by (ω_0, β_0). The envelope velocity in this case is thus somewhat less than the carrier velocity. As $\Delta\omega$ becomes vanishingly small, the envelope velocity is exactly the slope of the curve at ω_0. We can therefore state the following for our example:

$$\lim_{\Delta\omega \to 0} \frac{\Delta\omega}{\Delta\beta} = \frac{d\omega}{d\beta}\bigg|_{\omega_0} = v_g(\omega_0) \tag{85}$$

The quantity $d\omega/d\beta$ is called the *group velocity* function for the material, $v_g(\omega)$. When evaluated at a specified frequency ω_0, it represents the velocity of a group of frequencies within a spectral packet of vanishingly small width, centered at frequency ω_0. In stating this, we have extended our two-frequency example to include waves that have a continuous frequency spectrum. Each frequency component (or packet) is associated with a group velocity at which the energy in that packet propagates. Since the slope of the ω-β curve changes with frequency, group velocity will obviously be a function of frequency. The *group velocity dispersion* of the medium is, to the first order, the rate at which the slope of the ω-β curve changes with frequency. It is this behavior that is of critical practical importance to the propagation of modulated waves within dispersive media and to understanding the extent to which the modulation envelope may degrade with propagation distance.

EXAMPLE 12.10

Consider a medium in which the refractive index varies linearly with frequency over a certain range:

$$n(\omega) = n_0 \frac{\omega}{\omega_0}$$

Determine the group velocity and the phase velocity of a wave at frequency ω_0.

Solution. First, the phase constant will be

$$\beta(\omega) = n(\omega)\frac{\omega}{c} = \frac{n_0\omega^2}{\omega_0 c}$$

Now

$$\frac{d\beta}{d\omega} = \frac{2n_0\omega}{\omega_0 c}$$

so that

$$v_g = \frac{d\omega}{d\beta} = \frac{\omega_0 c}{2 n_0 \omega}$$

The group velocity at ω_0 is

$$v_g(\omega_0) = \frac{c}{2 n_0}$$

The phase velocity at ω_0 will be

$$v_p(\omega_0) = \frac{\omega}{\beta(\omega_0)} = \frac{c}{n_0}$$

12.8 PULSE BROADENING IN DISPERSIVE MEDIA

To see how a dispersive medium affects a modulated wave, let us consider the propagation of an electromagnetic pulse. Pulses are used in digital signals, where the presence or absence of a pulse in a given time slot corresponds to a digital "one" or "zero." The effect of the dispersive medium on a pulse is to broaden it in time. To see how this happens, we consider the pulse *spectrum,* which is found through the Fourier transform of the pulse in time domain. In particular, suppose the pulse shape in time is Gaussian, and has electric field given at position $z = 0$ by

$$E(0, t) = E_0 e^{-\frac{1}{2}(t/T)^2} e^{j\omega_0 t} \tag{86}$$

where E_0 is a constant, ω_0 is the carrier frequency, and T is the characteristic half-width of the pulse envelope; this is the time at which the pulse *intensity,* or magnitude of the Poynting vector, falls to $1/e$ of its maximum value (note that intensity is proportional to the square of the electric field). The frequency spectrum of the pulse is the Fourier transform of (86), which is

$$E(0, \omega) = \frac{E_0 T}{\sqrt{2\pi}} e^{-\frac{1}{2}T^2(\omega - \omega_0)^2} \tag{87}$$

Note from (87) that the frequency displacement from ω_0 at which the spectral *intensity* (proportional to $|E(0, \omega)|^2$) falls to $1/e$ of its maximum is $\Delta\omega = \omega - \omega_0 = 1/T$.

Figure 12.14a shows the Gaussian intensity spectrum of the pulse, centered at ω_0, where the frequencies corresponding to the $1/e$ spectral intensity positions, ω_a and ω_b, are indicated. Figure 12.14b shows the same three frequencies marked on the ω-β curve for the medium. Three lines are drawn that are tangent to the curve at the three frequency locations. The slopes of the lines indicate the group velocities at ω_a, ω_b, and ω_0, indicated as v_{ga}, v_{gb}, and v_{g0}. We can think of the pulse spreading in time as resulting from the differences in propagation times of the spectral energy packets that make up the pulse spectrum. Since the pulse spectral energy is highest
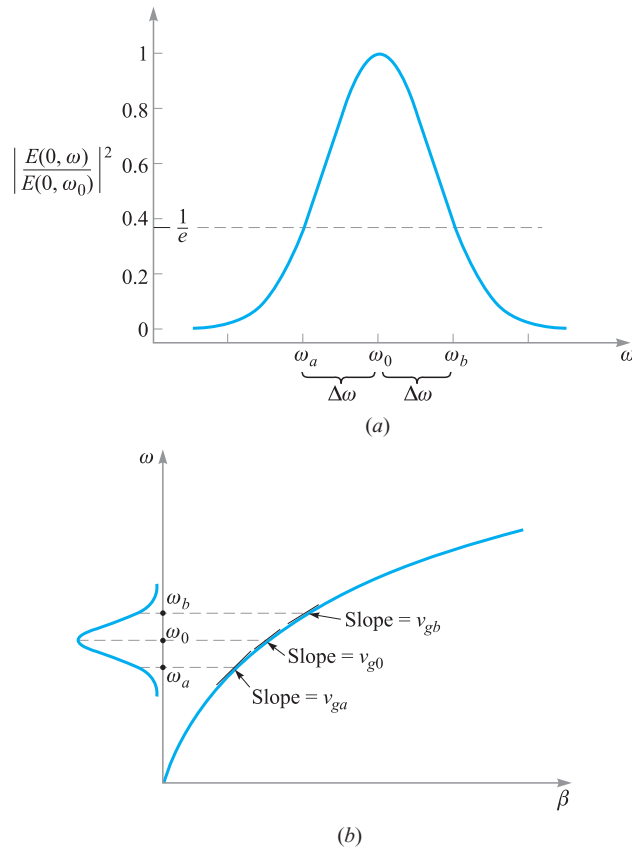
Figure 12.14 (a) Normalized power spectrum of a Gaussian pulse, as determined from Eq. (86). The spectrum is centered at carrier frequency ω_0 and has $1/e$ half-width, $\Delta\omega$. Frequencies ω_a and ω_b correspond to the $1/e$ positions on the spectrum. (b) The spectrum of Figure 12.14a as shown on the ω-β diagram for the medium. The three frequencies specified in Figure 12.14a are associated with three different slopes on the curve, resulting in different group delays for the spectral components.

at the center frequency, ω_0, we can use this as a reference point about which further spreading of the energy will occur. For example, let us consider the difference in arrival times (group delays) between the frequency components, ω_0 and ω_b, after propagating through a distance z of the medium:

$$\Delta\tau = z\left(\frac{1}{v_{gb}} - \frac{1}{v_{g0}}\right) = z\left(\left.\frac{d\beta}{d\omega}\right|_{\omega_b} - \left.\frac{d\beta}{d\omega}\right|_{\omega_0}\right) \tag{88}$$

The essential point is that the medium is acting as what could be called a *temporal prism*. Instead of spreading out the spectral energy packets spatially, it is spreading

them out in time. In this process, a new temporal pulse envelope is constructed whose width is based fundamentally on the spread of propagation delays of the different spectral components. By determining the delay difference between the peak spectral component and the component at the spectral half-width, we construct an expression for the new *temporal* half-width. This assumes, of course, that the initial pulse width is negligible in comparison, but if not, we can account for that also, as will be shown later on.

To evaluate (88), we need more information about the ω-β curve. If we assume that the curve is smooth and has fairly uniform curvature, we can express $\beta(\omega)$ as the first three terms of a Taylor series expansion about the carrier frequency, ω_0:

$$\beta(\omega) \doteq \beta(\omega_0) + (\omega - \omega_0)\beta_1 + \frac{1}{2}(\omega - \omega_0)^2\beta_2 \qquad (89)$$

where

$$\beta_0 = \beta(\omega_0)$$

$$\beta_1 = \left.\frac{d\beta}{d\omega}\right|_{\omega_0} \qquad (90)$$

and

$$\beta_2 = \left.\frac{d^2\beta}{d\omega^2}\right|_{\omega_0} \qquad (91)$$

Note that if the ω-β curve were a straight line, then the first two terms in (89) would precisely describe $\beta(\omega)$. It is the third term in (89), involving β_2, that describes the curvature and ultimately the dispersion.

Noting that β_0, β_1, and β_2 are constants, we take the first derivative of (89) with respect to ω to find

$$\frac{d\beta}{d\omega} = \beta_1 + (\omega - \omega_0)\beta_2 \qquad (92)$$

We now substitute (92) into (88) to obtain

$$\Delta\tau = [\beta_1 + (\omega_b - \omega_0)\beta_2]\,z - [\beta_1 + (\omega_0 - \omega_0)\beta_2]\,z = \Delta\omega\beta_2 z = \frac{\beta_2 z}{T} \qquad (93)$$

where $\Delta\omega = (\omega_b - \omega_0) = 1/T$. β_2, as defined in Eq. (91), is the *dispersion parameter*. Its units are in general time^2/distance, that is, pulse spread in time per unit spectral bandwidth, per unit distance. In optical fibers, for example, the units most commonly used are picoseconds^2/kilometer (psec^2/km). β_2 can be determined when we know how β varies with frequency, or it can be measured.

If the initial pulse width is very short compared to $\Delta\tau$, then the broadened pulse width at location z will be simply $\Delta\tau$. If the initial pulse width is comparable to $\Delta\tau$, then the pulse width at z can be found through the convolution of the initial Gaussian

pulse envelope of width T with a Gaussian envelope whose width is $\Delta\tau$. Thus, in general, the pulse width at location z will be

$$T' = \sqrt{T^2 + (\Delta\tau)^2} \tag{94}$$

EXAMPLE 12.11

An optical fiber link is known to have dispersion $\beta_2 = 20$ ps^2/km. A Gaussian light pulse at the input of the fiber is of initial width $T = 10$ ps. Determine the width of the pulse at the fiber output if the fiber is 15 km long.

Solution. The pulse spread will be

$$\Delta\tau = \frac{\beta_2 z}{T} = \frac{(20)(15)}{10} = 30 \text{ ps}$$

So the output pulse width is

$$T' = \sqrt{(10)^2 + (30)^2} = 32 \text{ ps}$$

An interesting by-product of pulse broadening through chromatic dispersion is that the broadened pulse is *chirped*. This means that the instantaneous frequency of the pulse varies monotonically (either increases or decreases) with time over the pulse envelope. This again is just a manifestation of the broadening mechanism, in which the spectral components at different frequencies are spread out in time as they propagate at different group velocities. We can quantify the effect by calculating the group delay, τ_g, as a function of frequency, using (92). We obtain:

$$\tau_g = \frac{z}{v_g} = z \frac{d\beta}{d\omega} = (\beta_1 + (\omega - \omega_0)\beta_2) z \tag{95}$$

This equation tells us that the group delay will be a linear function of frequency and that higher frequencies will arrive at later times if β_2 is positive. We refer to the chirp as positive if the lower frequencies lead the higher frequencies in time [requiring a positive β_2 in (95)]; chirp is negative if the higher frequencies lead in time (negative β_2). Figure 12.15 shows the broadening effect and illustrates the chirping phenomenon.

> **D12.6.** For the fiber link of Example 12.11, a 20-ps pulse is input instead of the 10-ps pulse in the example. Determine the output pulsewidth.
>
> **Ans.** 25 ps

As a final point, we note that the pulse bandwidth, $\Delta\omega$, was found to be $1/T$. This is true as long as the Fourier transform of the pulse *envelope* is taken, as was done with (86) to obtain (87). In that case, E_0 was taken to be a constant, and so the only time variation arose from the carrier wave and the Gaussian envelope. Such a
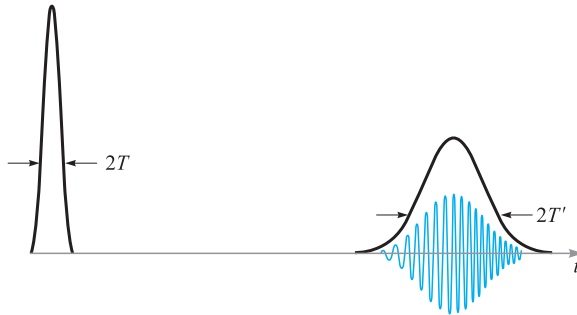
Figure 12.15 Gaussian pulse intensities as functions of time (smooth curves) before and after propagation through a dispersive medium, as exemplified by the ω-β diagram of Figure 12.14*b*. The electric field oscillations are shown under the second trace to demonstrate the chirping effect as the pulse broadens. Note the reduced amplitude of the broadened pulse, which occurs because the pulse energy (the area under the intensity envelope) is constant.

pulse, whose frequency spectrum is obtained only from the pulse envelope, is known as *transform-limited*. In general, however, additional frequency bandwidth may be present since E_0 may vary with time for one reason or another (such as phase noise that could be present on the carrier). In these cases, pulse broadening is found from the more general expression

$$\Delta\tau = \Delta\omega\beta_2 z \qquad (96)$$

where $\Delta\omega$ is the net spectral bandwidth arising from all sources. Clearly, transform-limited pulses are preferred in order to minimize broadening because these will have the smallest spectral width for a given pulse width.

REFERENCES

1. DuBroff, R. E., S. V. Marshall, and G. G. Skitek. *Electromagnetic Concepts and Applications*. 4th ed. Englewood Cliffs, N. J.: Prentice-Hall, 1996. Chapter 9 of this text develops the concepts presented here, with additional examples and applications.
2. Iskander, M. F. *Electromagnetic Fields and Waves*. Englewood Cliffs, N. J.: Prentice-Hall, 1992. The multiple interface treatment in Chapter 5 of this text is particularly good.
3. Harrington, R. F. *Time-Harmonic Electromagnetic Fields*. New York: McGraw-Hill, 1961. This advanced text provides a good overview of general wave reflection concepts in Chapter 2.
4. Marcuse, D. *Light Transmission Optics*. New York: Van Nostrand Reinhold, 1982. This intermediate-level text provides detailed coverage of optical waveguides and pulse propagation in dispersive media.

CHAPTER 12   PROBLEMS

12.1 A uniform plane wave in air, $E_{x1}^+ = E_{x10}^+ \cos(10^{10}t - \beta z)$ V/m, is normally incident on a copper surface at $z = 0$. What percentage of the incident power density is transmitted into the copper?

12.2 The plane $z = 0$ defines the boundary between two dielectrics. For $z < 0$, $\epsilon_{r1} = 9$, $\epsilon_{r1}'' = 0$, and $\mu_1 = \mu_0$. For $z > 0$, $\epsilon_{r2}' = 3$, $\epsilon_{r2}'' = 0$, and $\mu_2 = \mu_0$. Let $E_{x1}^+ = 10 \cos(\omega t - 15z)$ V/m and find (a) ω; (b) $\langle \mathbf{S}_1^+ \rangle$; (c) $\langle \mathbf{S}_1^- \rangle$; (d) $\langle \mathbf{S}_2^+ \rangle$.

12.3 A uniform plane wave in region 1 is normally incident on the planar boundary separating regions 1 and 2. If $\epsilon_1'' = \epsilon_2'' = 0$, while $\epsilon_{r1}' = \mu_{r1}^3$ and $\epsilon_{r2}' = \mu_{r2}^3$, find the ratio $\epsilon_{r2}'/\epsilon_{r1}'$ if 20% of the energy in the incident wave is reflected at the boundary. There are two possible answers.

12.4 A 10 MHz uniform plane wave having an initial average power density of 5 W/m² is normally incident from free space onto the surface of a lossy material in which $\epsilon_2''/\epsilon_2' = 0.05$, $\epsilon_{r2}' = 5$, and $\mu_2 = \mu_0$. Calculate the distance into the lossy medium at which the transmitted wave power density is down by 10 dB from the initial 5 W/m².

12.5 The region $z < 0$ is characterized by $\epsilon_r' = \mu_r = 1$ and $\epsilon_r'' = 0$. The total **E** field here is given as the sum of two uniform plane waves, $\mathbf{E}_s = 150 \, e^{-j10z}\mathbf{a}_x + (50\angle 20°) \, e^{j10z}\mathbf{a}_x$ V/m. (a) What is the operating frequency? (b) Specify the intrinsic impedance of the region $z > 0$ that would provide the appropriate reflected wave. (c) At what value of z, $-10\,\text{cm} < z < 0$, is the total electric field intensity a maximum amplitude?

12.6 In the beam-steering prism of Example 12.8, suppose the antireflective coatings are removed, leaving bare glass-to-air interfaces. Calcluate the ratio of the prism output power to the input power, assuming a single transit.

12.7 The semi-infinite regions $z < 0$ and $z > 1$ m are free space. For $0 < z < 1$ m, $\epsilon_r' = 4$, $\mu_r = 1$, and $\epsilon_r'' = 0$. A uniform plane wave with $\omega = 4 \times 10^8$ rad/s is traveling in the $\mathbf{a}_z$ direction toward the interface at $z = 0$. (a) Find the standing wave ratio in each of the three regions. (b) Find the location of the maximum $|\mathbf{E}|$ for $z < 0$ that is nearest to $z = 0$.

12.8 A wave starts at point a, propagates 1 m through a lossy dielectric rated at 0.1 dB/cm, reflects at normal incidence at a boundary at which $\Gamma = 0.3 + j0.4$, and then returns to point a. Calculate the ratio of the final power to the incident power after this round trip, and specify the overall loss in decibels.

12.9 Region 1, $z < 0$, and region 2, $z > 0$, are both perfect dielectrics ($\mu = \mu_0$, $\epsilon'' = 0$). A uniform plane wave traveling in the $\mathbf{a}_z$ direction has a radian frequency of 3×10^{10} rad/s. Its wavelengths in the two regions are $\lambda_1 = 5$ cm and $\lambda_2 = 3$ cm. What percentage of the energy incident on the

boundary is (*a*) reflected; (*b*) transmitted? (*c*) What is the standing wave ratio in region 1?

12.10 In Figure 12.1, let region 2 be free space, while $\mu_{r1} = 1$, $\epsilon''_{r1} = 0$, and ϵ'_{r1} is unknown. Find ϵ'_{r1} if (*a*) the amplitude of $\mathbf{E}_1^-$ is one-half that of $\mathbf{E}_1^+$; (*b*) $\langle \mathbf{S}_1^- \rangle$ is one-half of $\langle \mathbf{S}_1^+ \rangle$; (*c*) $|\mathbf{E}_1|_{\min}$ is one-half of $|\mathbf{E}_1|_{\max}$.

12.11 A 150-MHz uniform plane wave is normally incident from air onto a material whose intrinsic impedance is unknown. Measurements yield a standing wave ratio of 3 and the appearance of an electric field minimum at 0.3 wavelengths in front of the interface. Determine the impedance of the unknown material.

12.12 A 50-MHz uniform plane wave is normally incident from air onto the surface of a calm ocean. For seawater, $\sigma = 4$ S/m, and $\epsilon'_r = 78$. (*a*) Determine the fractions of the incident power that are reflected and transmitted. (*b*) Qualitatively, how (if at all) will these answers change as the frequency is increased?

12.13 A right-circularly polarized plane wave is normally incident from air onto a semi-infinite slab of plexiglas ($\epsilon'_r = 3.45$, $\epsilon''_r = 0$). Calculate the fractions of the incident power that are reflected and transmitted. Also, describe the polarizations of the reflected and transmitted waves.

12.14 A left-circularly polarized plane wave is normally incident onto the surface of a perfect conductor. (*a*) Construct the superposition of the incident and reflected waves in phasor form. (*b*) Determine the real instantaneous form of the result of part (*a*). (*c*) Describe the wave that is formed.

12.15 Sulfur hexafluoride (SF_6) is a high-density gas that has refractive index, $n_s = 1.8$ at a specified pressure, temperature, and wavelength. Consider the retro-reflecting prism shown in Fig. 12.16, that is immersed in SF_6. Light enters through a quarter-wave antireflective coating and then totally reflects from the back surfaces of the glass. In principle, the beam should experience zero loss at the design wavelength ($P_{out} = P_{in}$). (*a*) Determine the minimum required value of the glass refractive index, n_g, so that the interior beam will totally reflect. (*b*) Knowing n_g, find the required refractive index of the quarter-wave film, n_f. (*c*) With the SF_6 gas evacuated
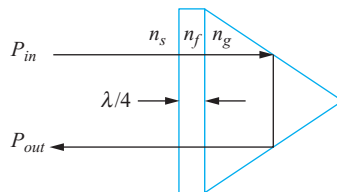
Figure 12.16   See Problem 12.15.

from the chamber, and with the glass and film values as previously found, find the ratio, P_{out}/P_{in}. Assume very slight misalignment, so that the long beam path through the prism is not retraced by reflected waves.

12.16 In Figure 12.5, let regions 2 and 3 both be of quarter-wave thickness. Region 4 is glass, having refractive index, $n_4 = 1.45$; region 1 is air. (*a*) Find $\eta_{in,b}$. (*b*) Find $\eta_{in,a}$. (*c*) Specify a relation between the four intrinsic impedances that will enable total transmission of waves incident from the left into region 4. (*d*) Specify refractive index values for regions 2 and 3 that will accomplish the condition of part (*c*). (*e*) Find the fraction of incident power transmitted if the two layers were of half-wave thickness instead of quarter wave.

12.17 A uniform plane wave in free space is normally incident onto a dense dielectric plate of thickness $\lambda/4$, having refractive index n. Find the required value of n such that exactly half the incident power is reflected (and half transmitted). Remember that $n > 1$.

12.18 A uniform plane wave is normally incident onto a slab of glass ($n = 1.45$) whose back surface is in contact with a perfect conductor. Determine the reflective phase shift at the front surface of the glass if the glass thickness is (*a*) $\lambda/2$; (*b*) $\lambda/4$; (*c*) $\lambda/8$.

12.19 You are given four slabs of lossless dielectric, all with the same intrinsic impedance, η, known to be different from that of free space. The thickness of each slab is $\lambda/4$, where λ is the wavelength as measured in the slab material. The slabs are to be positioned parallel to one another, and the combination lies in the path of a uniform plane wave, normally incident. The slabs are to be arranged such that the airspaces between them are either zero, one-quarter wavelength, or one-half wavelength in thickness. Specify an arrangement of slabs and airspaces such that (*a*) the wave is totally transmitted through the stack, and (*b*) the stack presents the highest reflectivity to the incident wave. Several answers may exist.

12.20 The 50-MHz plane wave of Problem 12.12 is incident onto the ocean surface at an angle to the normal of 60°. Determine the fractions of the incident power that are reflected and transmitted for (*a*) s-polarization, and (*b*) p-polarization.

12.21 A right-circularly polarized plane wave in air is incident at Brewster's angle onto a semi-infinite slab of plexiglas ($\epsilon'_r = 3.45$, $\epsilon''_r = 0$). (*a*) Determine the fractions of the incident power that are reflected and transmitted. (*b*) Describe the polarizations of the reflected and transmitted waves.

12.22 A dielectric waveguide is shown in Figure 12.17 with refractive indices as labeled. Incident light enters the guide at angle ϕ from the front surface normal as shown. Once inside, the light totally reflects at the upper $n_1 - n_2$ interface, where $n_1 > n_2$. All subsequent reflections from the upper and lower boundaries will be total as well, and so the light is confined to the
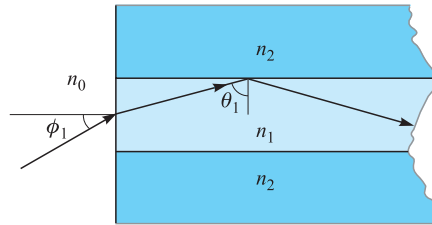
Figure 12.17 See Problems 12.22 and 12.23.

guide. Express, in terms of n_1 and n_2, the maximum value of ϕ such that total confinement will occur, with $n_0 = 1$. The quantity $\sin \phi$ is known as the *numerical aperture* of the guide.

12.23 Suppose that ϕ in Figure 12.17 is Brewster's angle, and that θ_1 is the critical angle. Find n_0 in terms of n_1 and n_2.

12.24 A *Brewster prism* is designed to pass p-polarized light without any reflective loss. The prism of Figure 12.18 is made of glass ($n = 1.45$) and is in air. Considering the light path shown, determine the vertex angle α.

12.25 In the Brewster prism of Figure 12.18, determine for s-polarized light the fraction of the incident power that is transmitted through the prism, and from this specify the dB *insertion loss*, defined as $10 \log_{10}$ of that number.

12.26 Show how a single block of glass can be used to turn a p-polarized beam of light through $180°$, with the light suffering (in principle) zero reflective loss. The light is incident from air, and the returning beam (also in air) may be displaced sideways from the incident beam. Specify all pertinent angles and use $n = 1.45$ for glass. More than one design is possible here.

12.27 Using Eq. (79) in Chapter 11 as a starting point, determine the ratio of the group and phase velocities of an electromagnetic wave in a good conductor. Assume conductivity does not vary with frequency.

12.28 Over a small wavelength range, the refractive index of a certain material varies approximately linearly with wavelength as $n(\lambda) \doteq n_a + n_b(\lambda - \lambda_a)$, where n_a, n_b and λ_a are constants, and where λ is the free-space wavelength. (*a*) Show that $d/d\omega = -(2\pi c/\omega^2)d/d\lambda$. (*b*) Using $\beta(\lambda) = 2\pi n/\lambda$,
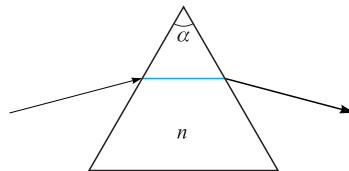
Figure 12.18 See Problems 12.24 and 12.25.

determine the wavelength-dependent (or independent) group delay over a unit distance. (c) Determine β_2 from your result of part (b). (d) Discuss the implications of these results, if any, on pulse broadening.

12.29 A $T = 5$ ps transform-limited pulse propagates in a dispersive medium for which $\beta_2 = 10\,\text{ps}^2/\text{km}$. Over what distance will the pulse spread to twice its initial width?

12.30 A $T = 20$ ps transform-limited pulse propagates through 10 km of a dispersive medium for which $\beta_2 = 12\,\text{ps}^2/\text{km}$. The pulse then propagates through a second 10 km medium for which $\beta_2 = -12\,\text{ps}^2/\text{km}$. Describe the pulse at the output of the second medium and give a physical explanation for what happened.

Guided Waves

I n this chapter, we investigate several structures for guiding electromagnetic waves, and we explore the principles by which these operate. Included are transmission lines, which we first explored from the viewpoint of their currents and voltages in Chapter 10, and which we now revisit from a fields point of view. We then broaden the discussion to include several waveguiding devices. Broadly defined, a waveguide is a structure through which electromagnetic waves can be transmitted from point to point and within which the fields are confined to a certain extent. A transmission line fits this description, but it is a special case that employs two conductors, and it propagates a purely TEM field configuration. Waveguides in general depart from these restrictions and may employ any number of conductors and dielectrics—or as we will see, dielectrics alone and no conductors.

The chapter begins with a presentation of several transmission line structures, with emphasis on obtaining expressions for the primary constants, L, C, G, and R, for high- and low-frequency operating regimes. Next, we begin our study of waveguides by first taking a broad view of waveguide devices to obtain a physical understanding of how they work and the conditions under which they are used. We then explore the simple parallel-plate structure and distinguish between its operation as a transmission line and as a waveguide. In this device, the concept of waveguide modes is developed, as are the conditions under which these will occur. We will study the electric and magnetic field configurations of the guided modes using simple plane wave models and the wave equation. We will then study more complicated structures, including rectangular waveguides, dielectric slab waveguides, and optical fibers. ■

13.1 TRANSMISSION LINE FIELDS AND PRIMARY CONSTANTS

We begin by establishing the equivalence between transmission line operations when considering voltage and current, from the point of view of the fields within the line. Consider, for example, the parallel-plate line shown in Figure 13.1. In the line, we
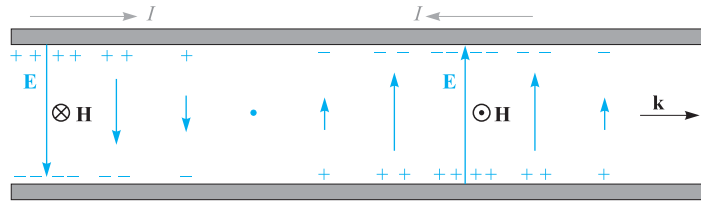
Figure 13.1   A transmission-line wave represented by voltage and current distributions along the length is associated with transverse electric and magnetic fields, forming a TEM wave.

assume that the plate spacing, d, is much less than the line width, b (into the page), so electric and magnetic fields can be assumed to be uniform within any transverse plane. Lossless propagation is also assumed. Figure 13.1 shows the side view, which includes the propagation axis z. The fields, along with the voltage and current, are shown at an instant in time.

The voltage and current in phasor form are:

$$V_s(z) = V_0 e^{-j\beta z} \tag{1a}$$

$$I_s(z) = \frac{V_0}{Z_0} e^{-j\beta z} \tag{1b}$$

where $Z_0 = \sqrt{L/C}$. The electric field in a given transverse plane at location z is just the parallel-plate capacitor field:

$$E_{sx}(z) = \frac{V_s}{d} = \frac{V_0}{d} e^{-j\beta z} \tag{2a}$$

The magnetic field is equal to the surface current density, assumed uniform, on either plate [Eq. (12), Chapter 7]:

$$H_{sy}(z) = K_{sz} = \frac{I_s}{b} = \frac{V_0}{bZ_0} e^{-j\beta z} \tag{2b}$$

The two fields, both uniform, orthogonal, and lying in the transverse plane, are identical in form to those of a uniform plane wave. As such, they are transverse electromagnetic (TEM) fields, also known simply as transmission-line fields. They differ from the fields of the uniform plane wave only in that they exist within the interior of the line, and nowhere else.

The power flow down the line is found through the time-average Poynting vector, integrated over the line cross section. Using ($2a$) and ($2b$), we find:

$$P_z = \int_0^b \int_0^d \frac{1}{2} \mathrm{Re}\{E_{xs} H_{ys}^*\}\, dxdy = \frac{1}{2} \frac{V_0}{d} \frac{V_0^*}{bZ_0^*}(bd) = \frac{|V_0|^2}{2Z_0^*} = \frac{1}{2}\mathrm{Re}\{V_s I_s^*\} \tag{3}$$

The power transmitted by the line is one of the most important quantities that we wish to know from a practical standpoint. Eq. (3) shows that this can be obtained consistently through the line fields, or through the voltage and current. As would be
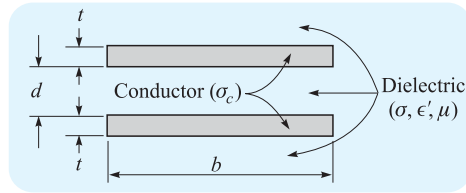
Figure 13.2 The geometry of the parallel-plate transmission line.

expected, this consistency is maintained when losses are included. The fields picture is in fact advantageous, and is generally preferred, since it is easy to incorporate dielectric loss mechanisms (other than conductivity) in addition to the dispersive properties of the dielectric. The transmission-line fields are also needed to produce the primary constants, as we now demonstrate for the parallel-plate line and other selected line geometries.

We assume the line is filled with dielectric having permittivity ϵ', conductivity σ, and permeability μ, usually μ_0 (Figure 13.2). The upper and lower plate thickness is t, which, along with the plate width b and plate conductivity σ_c, is used to evaluate the resistance per unit length parameter R under low-frequency conditions. We will, however, consider high-frequency operation, in which the skin effect gives an effective plate thickness or skin depth δ that is much less than t.

First, the capacitance and conductance per unit length are simply those of the parallel-plate structure, assuming static fields. Using Eq. (27) from Chapter 6, we find

$$C = \frac{\epsilon' b}{d} \tag{4}$$

The value of permittivity used should be appropriate for the range of operating frequencies considered.

The conductance per unit length may be determined easily from the capacitance expression by use of the simple relation between capacitance and resistance [Eq. (45), Chapter 6]:

$$G = \frac{\sigma}{\epsilon'} C = \frac{\sigma b}{d} \tag{5}$$

The evaluation of L and R involves the assumption of a well-developed skin effect such that $\delta \ll t$. Consequently, the inductance is primarily external because the magnetic flux within either conductor is negligible compared to that between conductors. Therefore,

$$L \doteq L_{\text{ext}} = \frac{\mu d}{b} \tag{6}$$

Note that $L_{\text{ext}} C = \mu \epsilon' = 1/v_p^2$, and we are therefore able to evaluate the external inductance for any transmission line for which we know the capacitance and insulator characteristics.

The last of the four parameters that we need is the resistance R per unit length. If the frequency is very high and the skin depth δ is very small, then we obtain an

appropriate expression for R by distributing the total current uniformly throughout a depth δ. The skin effect resistance (through both conductors in series over a unit length) is

$$R = \frac{2}{\sigma_c \delta b} \tag{7}$$

Finally, it is convenient to include the common expression for the characteristic impedance of the line here with the parameter formulas:

$$Z_0 = \sqrt{\frac{L_{\text{ext}}}{C}} = \sqrt{\frac{\mu}{\epsilon'}} \frac{d}{b} \tag{8}$$

If necessary, a more accurate value may be obtained from Eq. (47), Chapter 10. Note that when substituting (8) into (2b), and using (2a), we obtain the expected relation for a TEM wave, $E_{xs} = \eta H_{ys}$, where $\eta = \sqrt{\mu/\epsilon'}$.

> **D13.1.** Parameters for the planar transmission line shown in Figure 13.2 are $b = 6$ mm, $d = 0.25$ mm, $t = 25$ mm, $\sigma_c = 5.5 \times 10^7$ S/m, $\epsilon' = 25$ pF/m, $\mu = \mu_0$, and $\sigma/\omega\epsilon' = 0.03$. If the operating frequency is 750 MHz, calculate: (a) α; (b) β; (c) Z_0.
>
> **Ans.** 0.47 Np/m; 26 rad/m; 9.3∠0.7° Ω

13.1.1 Coaxial (High Frequencies)

We next consider a coaxial cable in which the dielectric has an inner radius a and outer radius b (Figure 13.3). The capacitance per unit length, obtained as Eq. (5) of Section 6.3, is

$$C = \frac{2\pi\epsilon'}{\ln(b/a)} \tag{9}$$
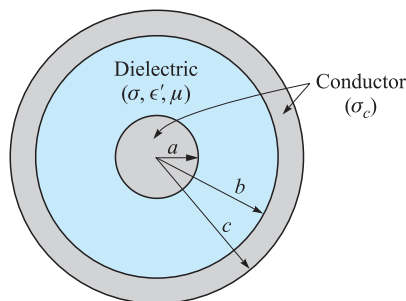
Figure 13.3 Coaxial transmission-line geometry.

Now, using the relation $RC = \epsilon/\sigma$ (see Problem 6.6), the conductance is

$$G = \frac{2\pi\sigma}{\ln(b/a)} \tag{10}$$

where σ is the conductivity of the dielectric between the conductors at the operating frequency.

The inductance per unit length was computed for the coaxial cable as Eq. (50) in Section 8.10,

$$L_{\text{ext}} = \frac{\mu}{2\pi}\ln(b/a) \tag{11}$$

Again, this is an external inductance, for the small skin depth precludes any appreciable magnetic flux within the conductors.

For a circular conductor of radius a and conductivity σ_c, we let Eq. (90) of Section 11.4 apply to a unit length, obtaining

$$R_{\text{inner}} = \frac{1}{2\pi a\delta\sigma_c}$$

There is also a resistance for the outer conductor, which has an inner radius b. We assume the same conductivity σ_c and the same value of skin depth δ, leading to

$$R_{\text{outer}} = \frac{1}{2\pi b\delta\sigma_c}$$

Because the line current flows through these two resistances in series, the total resistance is the sum:

$$R = \frac{1}{2\pi\delta\sigma_c}\left(\frac{1}{a} + \frac{1}{b}\right) \tag{12}$$

Finally, the characteristic impedance, assuming low losses, is

$$Z_0 = \sqrt{\frac{L_{\text{ext}}}{C}} = \frac{1}{2\pi}\sqrt{\frac{\mu}{\epsilon'}}\ln\frac{b}{a} \tag{13}$$

13.1.2 Coaxial (Low Frequencies)

We now obtain the coaxial line parameter values at very low frequencies where there is no appreciable skin effect and the current is assumed to be distributed uniformly throughout the conductor cross sections.

We first note that the current distribution in the conductor does not affect either the capacitance or conductance per unit length. Therefore,

$$C = \frac{2\pi\epsilon'}{\ln(b/a)} \tag{14}$$

and

$$G = \frac{2\pi\sigma}{\ln(b/a)} \tag{15}$$

The resistance per unit length may be calculated by dc methods, $R = l/(\sigma_c S)$, where $l = 1$ m and σ_c is the conductivity of the outer and inner conductors. The area of the

center conductor is πa^2 and that of the outer is $\pi(c^2 - b^2)$. Adding the two resistance values, we have

$$R = \frac{1}{\sigma_c \pi} \left(\frac{1}{a^2} + \frac{1}{c^2 - b^2} \right) \tag{16}$$

Only one of the four parameter values remains to be found, the inductance per unit length. The external inductance that we calculated at high frequencies is the greatest part of the total inductance. To it, however, must be added smaller terms representing the internal inductances of the inner and outer conductors.

At very low frequencies where the current distribution is uniform, the internal inductance of the center conductor is the subject of Problem 43 in Chapter 8; the relationship is also given as Eq. (62) in Section 8.10:

$$L_{a,\text{int}} = \frac{\mu}{8\pi} \tag{17}$$

The determination of the internal inductance of the outer shell is a more difficult problem, and most of the work was requested in Problem 36 in Chapter 8. There, we found that the energy stored per unit length in an outer cylindrical shell of inner radius b and outer radius c with uniform current distribution is

$$W_H = \frac{\mu I^2}{16\pi(c^2 - b^2)} \left(b^2 - 3c^2 + \frac{4c^2}{c^2 - b^2} \ln \frac{c}{b} \right)$$

Thus the internal inductance of the outer conductor at very low frequencies is

$$L_{bc,\text{int}} = \frac{\mu}{8\pi(c^2 - b^2)} \left(b^2 - 3c^2 + \frac{4c^2}{c^2 - b^2} \ln \frac{c}{b} \right) \tag{18}$$

At low frequencies the total inductance is obtained by adding (11), (17), and (18):

$$L = \frac{\mu}{2\pi} \left[\ln \frac{b}{a} + \frac{1}{4} + \frac{1}{4(c^2 - b^2)} \left(b^2 - 3c^2 + \frac{4c^2}{c^2 - b^2} \ln \frac{c}{b} \right) \right] \tag{19}$$

13.1.3 Coaxial (Intermediate Frequencies)

There still remains the frequency interval where the skin depth is neither very much larger than nor very much smaller than the radius. In this case, the current distribution is governed by Bessel functions, and both the resistance and internal inductance are complicated expressions. Values are tabulated in the handbooks, and it is necessary to use them for very small conductor sizes at high frequencies and for larger conductor sizes used in power transmission at low frequencies.[1]

---

[1] Bessel functions are discussed within the context of optical fiber in Section 13.7. The current distribution, internal inductance, and internal resistance of round wires is discussed (with numerical examples) in Weeks, pp. 35–44. See the References at the end of this chapter.

D13.2. The dimensions of a coaxial transmission line are $a = 4$ mm, $b = 17.5$ mm, and $c = 20$ mm. The conductivity of the inner and outer conductors is 2×10^7 S/m, and the dielectric properties are $\mu_r = 1$, $\epsilon_r' = 3$, and $\sigma/\omega\epsilon' = 0.025$. Assume that the loss tangent is constant with frequency. Determine: (a) L, C, R, G, and Z_0 at 150 MHz; (b) L and R at 60 Hz.

Ans. 0.30 μH/m, 113 pF/m, 0.27 Ω/m, 2.7 mS/m, 51 Ω; 0.36 μH/m, 1.16 mΩ/m

13.1.4 Two-Wire (High Frequencies)

For the two-wire transmission line of Figure 13.4 with conductors of radius a and conductivity σ_c with center-to-center separation d in a medium of permeability μ, permittivity ϵ', and conductivity σ_c, the capacitance per unit length is found using the results of Section 6.4:

$$C = \frac{\pi \epsilon'}{\cosh^{-1}(d/2a)} \tag{20}$$

or

$$C \doteq \frac{\pi \epsilon'}{\ln(d/a)} \qquad (a \ll d)$$

The external inductance may be found from $L_{\text{ext}}C = \mu\epsilon'$. It is

$$L_{\text{ext}} = \frac{\mu}{\pi} \cosh^{-1}(d/2a) \tag{21}$$

or

$$L_{\text{ext}} \doteq \frac{\mu}{\pi} \ln(d/a) \qquad (a \ll d)$$

The conductance per unit length may be written immediately from an inspection of the capacitance expression, and using the relation $RC = \epsilon/\sigma$:

$$G = \frac{\pi \sigma}{\cosh^{-1}(d/2a)} \tag{22}$$

The resistance per unit length is twice that of the center conductor of the coax,

$$R = \frac{1}{\pi a \delta \sigma_c} \tag{23}$$
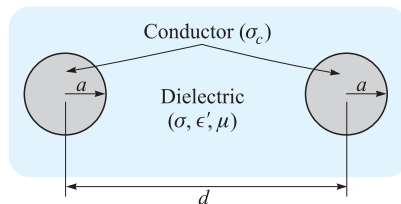
Figure 13.4 The geometry of the two-wire transmission line.

Finally, using the capacitance and the external inductance expressions, we obtain a value for the characteristic impedance,

$$Z_0 = \sqrt{\frac{L_{\text{ext}}}{C}} = \frac{1}{\pi}\sqrt{\frac{\mu}{\epsilon}} \cosh^{-1}(d/2a) \tag{24}$$

13.1.5 Two-Wire (Low Frequencies)

At low frequencies where a uniform current distribution may be assumed, we again must modify the L and R expressions, but not those for C and G. The latter two are again expressed by (20) and (22):

$$C = \frac{\pi\epsilon'}{\cosh^{-1}(d/2a)}$$

$$G = \frac{\pi\sigma}{\cosh^{-1}(d/2a)}$$

The inductance per unit length must be increased by twice the internal inductance of a straight round wire,

$$L = \frac{\mu}{\pi}\left[\frac{1}{4} + \cosh^{-1}(d/2a)\right] \tag{25}$$

The resistance becomes twice the dc resistance of a wire of radius a, conductivity σ_c, and unit length:

$$R = \frac{2}{\pi a^2 \sigma_c} \tag{26}$$

D13.3. The conductors of a two-wire transmission line each have a radius of 0.8 mm and a conductivity of 3×10^7 S/m. They are separated by a center-to-center distance of 0.8 cm in a medium for which $\epsilon'_r = 2.5$, $\mu_r = 1$, and $\sigma = 4 \times 10^{-9}$ S/m. If the line operates at 60 Hz, find: $(a)\,\delta$; $(b)\,C$; $(c)\,G$; $(d)\,L$; $(e)\,R$.

Ans. 1.2 cm; 30 pF/m; 5.5 nS/m; 1.02 μH/m; 0.033 Ω/m

13.1.6 Microstrip Line (Low Frequencies)

Microstrip line is one example of a class of configurations involving planar conductors of finite widths on or within dielectric substrates; they are usually employed as device interconnects for microelectronic circuitry. The microstrip configuration, shown in Figure 13.5, consists of a dielectric (assumed lossless) of thickness d and of permittivity $\epsilon' = \epsilon_r\epsilon_0$, sandwiched between a conducting ground plane and a narrow conducting strip of width w. The region above the top strip is air (assumed here) or a dielectric of lower permittivity.

The structure approaches the case of the parallel-plate line if $w \gg d$. In a microstrip, such an assumption is generally not valid, and so significant charge densities exist on both surfaces of the upper conductor. The resulting electric field, originating at the top conductor and terminating on the bottom conductor, will exist within
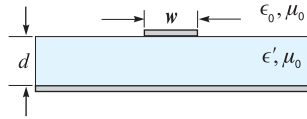
Figure 13.5 Microstrip line geometry.

both substrate *and* air regions. The same is true for the magnetic field, which circulates around the top conductor. This electromagnetic field configuration cannot propagate as a purely TEM wave because wave velocities within the two media will differ. Instead, waves having z components of **E** and **H** occur, with the z component magnitudes established so that the air and dielectric fields do achieve equal phase velocities (the reasoning behind this will be explained in Section 13.6). Analyzing the structure while allowing for the special fields is complicated, but it is usually permissible to approach the problem under the assumption of negligible z components. This is the *quasi TEM* approximation, in which the static fields (obtainable through numerical solution of Laplace's equation, for example) are used to evaluate the primary constants. Accurate results are obtained at low frequencies (below 1 or 2 GHz). At higher frequencies, results obtained through the static fields can still be used but in conjunction with appropriate modifying functions. We will consider the simple case of low-frequency operation and assume lossless propagation.[2]

To begin, it is useful to consider the microstrip line characteristics when the dielectric is *not* present. Assuming that both conductors have very small thicknesses, the internal inductance will be negligible, and so the phase velocity within the air-filled line, v_{p0}, will be

$$v_{p0} = \frac{1}{\sqrt{L_{ext}C_0}} = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = c \tag{27a}$$

where C_0 is the capacitance of the air-filled line (obtained from the electric field for that case), and c is the velocity of light. With the dielectric in place, the capacitance changes, *but the inductance does not,* provided the dielectric permeability is μ_0. Using (27a), the phase velocity now becomes

$$v_p = \frac{1}{\sqrt{L_{ext}C}} = c\sqrt{\frac{C_0}{C}} = \frac{c}{\sqrt{\epsilon_{r,eff}}} \tag{27b}$$

where the *effective dielectric constant* for the microstrip line is

$$\epsilon_{r,eff} = \frac{C}{C_0} = \left(\frac{c}{v_p}\right)^2 \tag{28}$$

It is implied from (28) that the microstrip capacitance C would result if both the air and substrate regions were filled homogeneously with material having dielectric constant $\epsilon_{r,eff}$. The effective dielectric constant is a convenient parameter to use

---

[2] The high-frequency case is treated in detail in Edwards (Reference 2).

because it provides a way of unifying the effects of the dielectric and the conductor geometry. To see this, consider the two extreme cases involving large and small width-to-height ratios, w/d. If w/d is very large, then the line resembles the parallel-plate line, in which nearly all of the electric field exists within the dielectric. In this case $\epsilon_{r,\text{eff}} \doteq \epsilon_r$. On the other hand, for a very narrow top strip, or small w/d, the dielectric and air regions contain roughly equal amounts of electric flux. In that case, the effective dielectric constant approaches its minimum, given by the average of the two dielectric constants. We therefore obtain the range of allowed values of $\epsilon_{r,\text{eff}}$:

$$\frac{1}{2}(\epsilon_r + 1) < \epsilon_{r,\text{eff}} < \epsilon_r \tag{29}$$

The physical interpretation of $\epsilon_{r,\text{eff}}$ is that it is a *weighted average* of the dielectric constants of the substrate and air regions, with the weighting determined by the extent to which the electric field fills either region. We may thus write the effective dielectric constant in terms of a *field filling factor, q*, for the substrate:

$$\epsilon_{r,\text{eff}} = 1 + q(\epsilon_r - 1) \tag{30}$$

where $0.5 < q < 1$. With large w/d, $q \to 1$; with small w/d, $q \to 0.5$.

Now, the characteristic impedances of the air-filled line and the line with dielectric substrate are, respectively, $Z_0^{\text{air}} = \sqrt{L_{\text{ext}}/C_0}$ and $Z_0 = \sqrt{L_{\text{ext}}C}$. Then, using (28), we find

$$Z_0 = \frac{Z_0^{\text{air}}}{\sqrt{\epsilon_{r,\text{eff}}}} \tag{31}$$

A procedure for obtaining the characteristic impedance would be to first evaluate the air-filled impedance for a given w/d. Then, knowing the effective dielectric constant, determine the actual impedance using (31). Another problem would be to determine the required ratio w/d for a given substrate material in order to achieve a desired characteristic impedance.

Detailed analyses have led to numerous approximation formulas for the evaluation of $\epsilon_{r,\text{eff}}$, Z_0^{air}, and Z_0 within different regimes (again, see Reference 2 and the references therein). For example, with dimensions restricted such that $1.3 < w/d < 3.3$, applicable formulas include:

$$Z_0^{\text{air}} \doteq 60 \ln \left[4\left(\frac{d}{w}\right) + \sqrt{16\left(\frac{d}{w}\right)^2 + 2} \right] \qquad \frac{w}{d} < 3.3 \tag{32}$$

and

$$\epsilon_{r,\text{eff}} \doteq \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2}\left(1 + 10\frac{d}{w}\right)^{-0.555} \qquad \frac{w}{d} > 1.3 \tag{33}$$

Or, if a line is to be fabricated having a desired value of Z_0, the effective dielectric constant (from which the required w/d can be obtained) is found through:

$$\epsilon_{r,\text{eff}} \doteq \epsilon_r[0.96 + \epsilon_r(0.109 - 0.004\epsilon_r)(\log_{10}(10 + Z_0) - 1)]^{-1} \qquad \frac{w}{d} > 1.3 \tag{34}$$

13.2 BASIC WAVEGUIDE OPERATION

Waveguides assume many different forms that depend on the purpose of the guide and on the frequency of the waves to be transmitted. The simplest form (in terms of analysis) is the parallel-plate guide shown in Figure 13.6. Other forms are the hollow-pipe guides, including the rectangular waveguide of Figure 13.7, and the cylindrical guide, shown in Figure 13.8. Dielectric waveguides, used primarily at optical frequencies, include the slab waveguide of Figure 13.9 and the optical fiber, shown in Figure 13.10. Each of these structures possesses certain advantages over the others, depending on the application and the frequency of the waves to be transmitted. All guides, however, exhibit the same basic operating principles, which we will explore in this section.

To develop an understanding of waveguide behavior, we consider the parallel-plate waveguide of Figure 13.6. At first, we recognize this as one of the transmission-line structures that we investigated in Section 13.1. So the first question that arises is: how does a waveguide differ from a transmission line to begin with? The difference lies in the form of the electric and magnetic fields within the line. To see this, consider again Figure 13.1, which shows the fields when the line operates as a transmission line. As we saw earlier, a sinusoidal voltage wave, with voltage applied between conductors, leads to an electric field that is directed vertically between the conductors as shown. Because current flows only in the z direction, magnetic field will be oriented in and out of the page (in the y direction). The interior fields comprise a plane electromagnetic wave which propagates in the z direction (as the Poynting vector will show), since both fields lie in the transverse plane. We refer to this as a transmission-line wave,
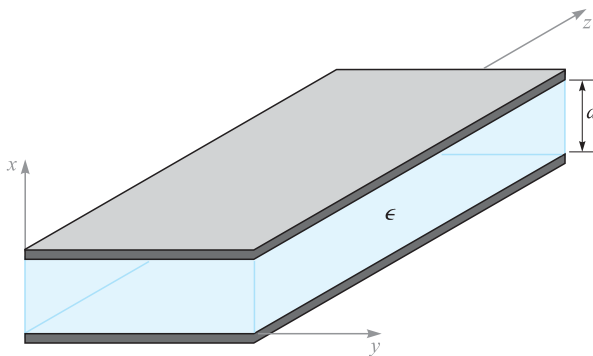
Figure 13.6 Parallel-plate waveguide, with metal plates at $x = 0, d$. Between the plates is a dielectric of permittivity ϵ.
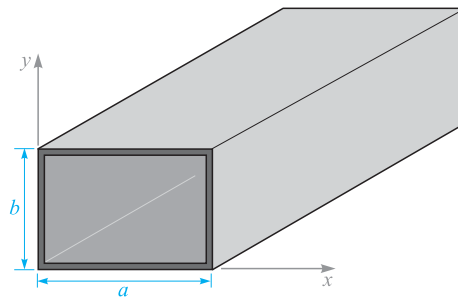
Figure 13.7   Rectangular waveguide.
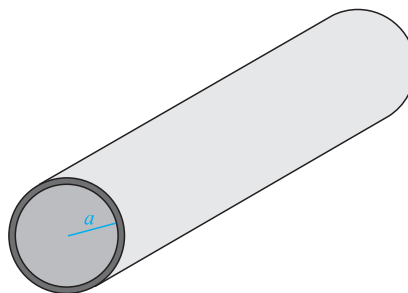
Figure 13.8   Cylindrical waveguide.
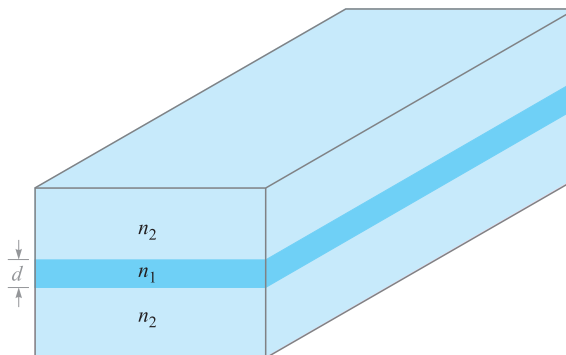
Figure 13.9   Symmetric dielectric slab waveguide, with slab region (refractive index n_1) surrounded by two dielectrics of index $n_2 < n_1$.
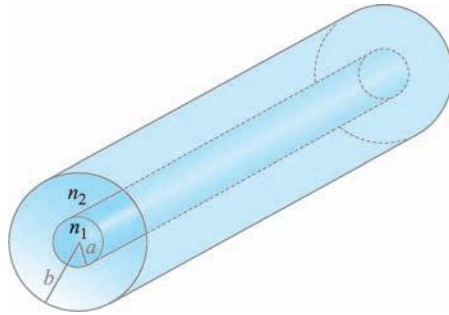
Figure 13.10 Optical fiber waveguide, with the core dielectric ($\rho < a$) of refractive index n_1. The cladding dielectric ($a < \rho < b$) is of index $n_2 < n_1$.

which, as discussed in Section 13.1, is a transverse electromagnetic, or TEM, wave. The wavevector **k**, shown in Figure 13.1, indicates the direction of wave travel as well as the direction of power flow.

As the frequency is increased, a remarkable change occurs in the way the fields progagate down the line. Although the original field configuration of Figure 13.1 may still be present, another possibility emerges, which is shown in Figure 13.11. Again, a plane wave is guided in the z direction, but by means of a progression of zig-zag reflections at the upper and lower plates. Wavevectors $\mathbf{k}_u$ and $\mathbf{k}_d$ are associated with the upward-and downward-propagating waves, respectively, and these have identical magnitudes,

$$|\mathbf{k}_u| = |\mathbf{k}_d| = k = \omega\sqrt{\mu\epsilon}$$

For such a wave to propagate, all upward-propagating waves must be *in phase* (as must be true of all downward-propagating waves). This condition can only be satisfied at certain discrete angles of incidence, shown as θ in Figure 13.11. An allowed value of θ, along with the resulting field configuration, comprise a *waveguide mode* of the structure. Associated with each guided mode is a *cutoff frequency*. If the operating frequency is below the cutoff frequency, the mode will not propagate. If above cutoff,
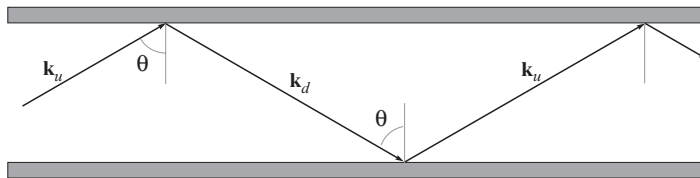
Figure 13.11 In a parallel-plate waveguide, plane waves can propagate by oblique reflection from the conducting walls. This produces a waveguide mode that is not TEM.
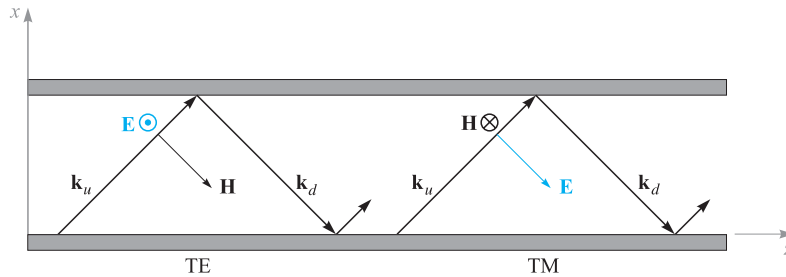
Figure 13.12   Plane wave representation of TE and TM modes in a parallel-plate guide.

the mode propagates. The TEM mode, however, has no cutoff; it will be supported at any frequency. At a given frequency, the guide may support several modes, the quantity of which depends on the plate separation and on the dielectric constant of the interior medium, as will be shown. The number of modes increases as the frequency is raised.

So to answer our initial question on the distinction between transmission lines and waveguides, we can state the following: Transmission lines consist of two or more conductors and as a rule will support TEM waves (or something which could approximate such a wave). A waveguide may consist of one or more conductors, or no conductors at all, and will support waveguide modes of forms similar to those just described. Waveguides may or may not support TEM waves, depending on the design.

In the parallel-plate guide, two types of waveguide modes can be supported. These are shown in Figure 13.12 as arising from the s and p orientations of the plane wave polarizations. In a manner consistent with our previous discussions on oblique reflection (Section 12.5), we identify a *transverse electric* or *TE* mode when **E** is perpendicular to the plane of incidence (s-polarized); this positions **E** parallel to the transverse plane of the waveguide, as well as to the boundaries. Similarly, a *transverse magnetic* or *TM* mode results with a p polarized wave; the entire **H** field is in the y direction and is thus within the transverse plane of the guide. Both possibilities are illustrated in Figure 13.12. Note, for example, that with **E** in the y direction (TE mode), **H** will have x and z components. Likewise, a TM mode will have x and z components of **E**.[3] In any event, the reader can verify from the geometry of Figure 13.12 that it is not possible to achieve a purely TEM mode for values of θ other than 90°. Other wave polarizations are possible that lie between the TE and TM cases, but these can always be expressed as superpositions of TE and TM modes.

---

[3] Other types of modes can exist in other structures (not the parallel-plate guide) in which *both* **E** and **H** have z components. These are known as *hybrid* modes, and they typically occur in guides with cylindrical cross sections, such as the optical fiber.

13.3   PLANE WAVE ANALYSIS OF THE PARALLEL-PLATE WAVEGUIDE

Let us now investigate the conditions under which waveguide modes will occur, using our plane wave model for the mode fields. In Figure 13.13a, a zig-zag path is again shown, but this time phase fronts are drawn that are associated with two of the upward-propagating waves. The first wave has reflected twice (at the top and bottom surfaces) to form the second wave (the downward-propagating phase fronts are not shown). Note that the phase fronts of the second wave do not coincide with those of the first wave, and so the two waves are out of phase. In Figure 13.13b, the wave angle has been adjusted so that the two waves are now in phase. Having satisfied this condition for the two waves, we will find that *all* upward-propagating waves will have coincident phase fronts. The same condition will automatically occur for all downward-propagating waves. This is the requirement to establish a guided mode.

In Figure 13.14 we show the wavevector, $\mathbf{k}_u$, and its components, along with a series of phase fronts. A drawing of this kind for $\mathbf{k}_d$ would be the same, except the

(a)

(b)
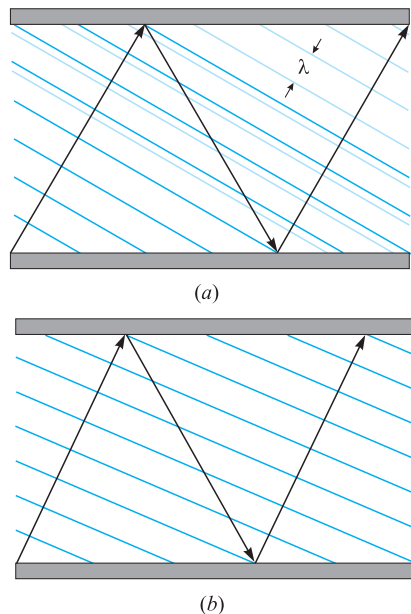
Figure 13.13   (a) Plane wave propagation in a parallel-plate guide in which the wave angle is such that the upward-propagating waves are not in phase. (b) The wave angle has been adjusted so that the upward waves are in phase, resulting in a guided mode.
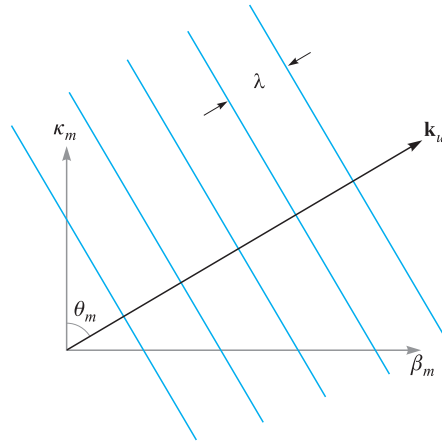
Figure 13.14 The components of the upward wavevector are κ_m and β_m, the transverse and axial phase constants. To form the downward wavevector, $\mathbf{k}_d$, the direction of κ_m is reversed.

x component, κ_m, would be reversed. In Section 12.4, we measured the phase shift per unit distance along the x and z directions by the components, k_x and k_z, which varied continuously as the direction of $\mathbf{k}$ changed. In our discussion of waveguides, we introduce a different notation, where κ_m and β_m are used for k_x and k_z. The subscript m is an integer indicating the *mode number*. This provides a subtle hint that β_m and κ_m will assume only certain discrete values that correspond to certain allowed directions of $\mathbf{k}_u$ and $\mathbf{k}_d$, such that our coincident phase front requirement is satisfied.[4] From the geometry we see that for any value of m,

$$\beta_m = \sqrt{k^2 - \kappa_m^2} \tag{35}$$

Use of the symbol β_m for the z components of $\mathbf{k}_u$ and $\mathbf{k}_d$ is appropriate because β_m will ultimately be the phase constant for the mth waveguide mode, measuring phase shift per distance down the guide; it is also used to determine the phase velocity of the mode, ω/β_m, and the group velocity, $d\omega/d\beta_m$.

Throughout our discussion, we will assume that the medium within the guide is lossless and nonmagnetic, so that

$$k = \omega\sqrt{\mu_0\epsilon'} = \frac{\omega\sqrt{\epsilon_r'}}{c} = \frac{\omega n}{c} \tag{36}$$

---

[4] Subscripts (m) are not shown on $\mathbf{k}_u$ and $\mathbf{k}_d$ but are understood. Changing m does not affect the magnitudes of these vectors, only their directions.

which we express either in terms of the dielectric constant, ϵ_r', or the refractive index, n, of the medium.

It is κ_m, the x component of $\mathbf{k}_u$ and $\mathbf{k}_d$, that will be useful to us in quantifying our requirement on coincident phase fronts through a condition known as *transverse resonance*. This condition states that the net phase shift measured during a round trip over the full transverse dimension of the guide must be an integer multiple of 2π radians. This is another way of stating that all upward- (or downward-) propagating plane waves must have coincident phases. The various segments of this round trip are illustrated in Figure 13.15. We assume for this exercise that the waves are frozen in time and that an observer moves vertically over the round trip, measuring phase shift along the way. In the first segment (Figure 13.15*a*), the observer starts at a position just above the lower conductor and moves vertically to the top conductor through distance d. The measured phase shift over this distance is $\kappa_m d$ rad. On reaching the top surface, the observer will note a possible phase shift on reflection (Figure 13.15*b*). This will be π if the wave is TE polarized and will be zero if the wave is TM polarized (see Figure 13.16 for a demonstration of this). Next, the observer moves along the reflected wave phases down to the lower conductor and again measures a phase shift of $\kappa_m d$ (Figure 13.15*c*). Finally, after including the phase shift on reflection at the bottom conductor, the observer is back at the original starting point and is noting the phase of the next upward-propagating wave.

The total phase shift over the round trip is required to be an integer multiple of 2π:

$$\boxed{\kappa_m d + \phi + \kappa_m d + \phi = 2m\pi} \tag{37}$$

where ϕ is the phase shift on reflection at each boundary. Note that with $\phi = \pi$ (TE waves) or 0 (TM waves), the net reflective phase shift over a round trip is 2π or 0, regardless of the angle of incidence. Thus the reflective phase shift has no bearing on the current problem, and we may simplify (37) to read

$$\boxed{\kappa_m = \frac{m\pi}{d}} \tag{38}$$

which is valid for *both* TE and TM modes. Note from Figure 13.14 that $\kappa_m = k\cos\theta_m$. Thus the wave angles for the allowed modes are readily found from (38) with (36):

$$\theta_m = \cos^{-1}\left(\frac{m\pi}{kd}\right) = \cos^{-1}\left(\frac{m\pi c}{\omega n d}\right) = \cos^{-1}\left(\frac{m\lambda}{2nd}\right) \tag{39}$$

where λ is the wavelength in free space.

We can next solve for the phase constant for each mode, using (35) with (38):

$$\beta_m = \sqrt{k^2 - \kappa_m^2} = k\sqrt{1 - \left(\frac{m\pi}{kd}\right)^2} = k\sqrt{1 - \left(\frac{m\pi c}{\omega n d}\right)^2} \tag{40}$$

(*a*)

Reflection with 0 or π phase shift

(*b*)

Reflection with 0 or π phase shift
(*c*)
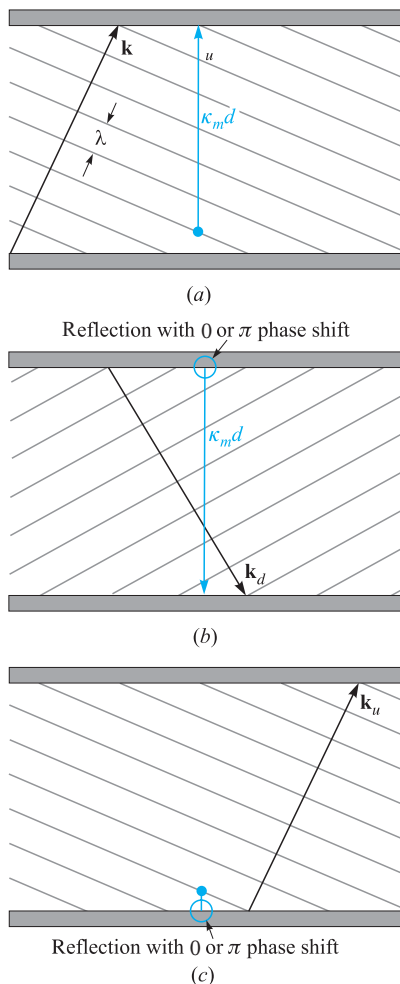
Figure 13.15 The net phase shift over a round trip in the parallel-plate guide is found by first measuring the transverse phase shift between plates of the initial upward wave (*a*); next, the transverse phase shift in the reflected (downward) wave is measured, while accounting for the reflective phase shift at the top plate (*b*); finally, the phase shift on reflection at the bottom plate is added, thus returning to the starting position, but with a new upward wave (*c*). Transverse resonance occurs if the phase at the final point is the same as that at the starting point (the two upward waves are in phase).
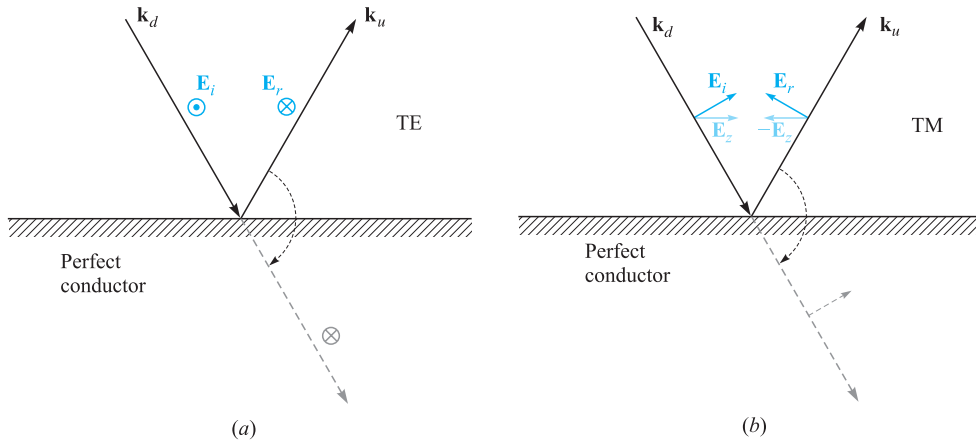
Figure 13.16 The phase shift of a wave on reflection from a perfectly conducting surface depends on whether the incident wave is TE (s-polarized) or TM (p-polarized). In both drawings, electric fields are shown as they would appear immediately adjacent to the conducting boundary. In (a) the field of a TE wave reverses direction upon reflection to establish a zero net field at the boundary. This constitutes a π phase shift, as is evident by considering a fictitious transmitted wave (dashed line) formed by a simple rotation of the reflected wave into alignment with the incident wave. In (b) an incident TM wave experiences a reversal of the *z component* of its electric field. The resultant field of the reflected wave, however, has not been phase-shifted; rotating the reflected wave into alignment with the incident wave (dashed line) shows this.

We define the radian *cutoff frequency* for mode m as

$$\omega_{cm} = \frac{m\pi c}{nd} \tag{41}$$

so that (40) becomes

$$\beta_m = \frac{n\omega}{c}\sqrt{1 - \left(\frac{\omega_{cm}}{\omega}\right)^2} \tag{42}$$

The significance of the cutoff frequency is readily seen from (42): If the operating frequency ω is greater than the cutoff frequency for mode m, then that mode will have phase constant β_m that is real-valued, and so the mode will propagate. For $\omega < \omega_{cm}$, β_m is imaginary, and the mode does not propagate.

Associated with the cutoff frequency is the *cutoff wavelength, λ_{cm}*, defined as the free-space wavelength at which cutoff for mode m occurs. This will be

$$\lambda_{cm} = \frac{2\pi c}{\omega_{cm}} = \frac{2nd}{m} \tag{43}$$

Note, for example, that in an air-filled guide ($n = 1$) the wavelength at which the lowest-order mode first starts to propagate is $\lambda_{c1} = 2d$, or the plate separation is one-half wavelength. Mode m will propagate whenever $\omega > \omega_{cm}$, or equivalently whenever $\lambda < \lambda_{cm}$. Use of the cutoff wavelength enables us to construct a second useful form of Eq. (42):

$$\boxed{\beta_m = \frac{2\pi n}{\lambda}\sqrt{1 - \left(\frac{\lambda}{\lambda_{cm}}\right)^2}} \tag{44}$$

EXAMPLE 13.1

A parallel-plate waveguide has plate separation $d = 1$ cm and is filled with teflon having dielectric constant $\epsilon_r' = 2.1$. Determine the maximum operating frequency such that only the TEM mode will propagate. Also find the range of frequencies over which the TE$_1$ and TM$_1$ ($m = 1$) modes, and no higher-order modes, will propagate.

Solution. Using (41), the cutoff frequency for the first waveguide mode ($m = 1$) will be

$$f_{c1} = \frac{\omega_{c1}}{2\pi} = \frac{2.99 \times 10^{10}}{2\sqrt{2.1}} = 1.03 \times 10^{10} \text{ Hz} = 10.3 \text{ GHz}$$

To propagate only TEM waves, we must have $f < 10.3$ GHz. To allow TE$_1$ and TM$_1$ (along with TEM) only, the frequency range must be $\omega_{c1} < \omega < \omega_{c2}$, where $\omega_{c2} = 2\omega_{c1}$, from (41). Thus, the frequencies at which we will have the $m = 1$ modes and TEM will be 10.3 GHz $< f <$ 20.6 GHz.

EXAMPLE 13.2

In the parallel-plate guide of Example 13.1, the operating wavelength is $\lambda = 2$ mm. How many waveguide modes will propagate?

Solution. For mode m to propagate, the requirement is $\lambda < \lambda_{cm}$. For the given waveguide and wavelength, the inequality becomes, using (43),

$$2 \text{ mm} < \frac{2\sqrt{2.1}\,(10 \text{ mm})}{m}$$

from which

$$m < \frac{2\sqrt{2.1}\,(10 \text{ mm})}{2 \text{ mm}} = 14.5$$

Thus the guide will support modes at the given wavelength up to order $m = 14$. Since there will be a TE and a TM mode for each value of m, this gives, not including the TEM mode, a total of 28 guided modes that are above cutoff.

The field configuration for a given mode can be found through the superposition of the fields of all the reflected waves. We can do this for the TE waves, for example, by writing the electric field phasor in the guide in terms of incident and reflected fields through

$$E_{ys} = E_0 e^{-j\mathbf{k}_u \cdot \mathbf{r}} - E_0 e^{-j\mathbf{k}_d \cdot \mathbf{r}} \tag{45}$$

where the wavevectors, $\mathbf{k}_u$ and $\mathbf{k}_d$, are indicated in Figure 13.12. The minus sign in front of the second term arises from the π phase shift on reflection. From the geometry depicted in Figure 13.14, we write

$$\mathbf{k}_u = \kappa_m \mathbf{a}_x + \beta_m \mathbf{a}_z \tag{46}$$

and

$$\mathbf{k}_d = -\kappa_m \mathbf{a}_x + \beta_m \mathbf{a}_z \tag{47}$$

Then, using

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

Eq. (45) becomes

$$E_{ys} = E_0(e^{-j\kappa_m x} - e^{j\kappa_m x})e^{-j\beta_m z} = 2jE_0 \sin(\kappa_m x)e^{-j\beta_m z} = E_0' \sin(\kappa_m x)e^{-j\beta_m z} \tag{48}$$

where the plane wave amplitude, E_0, and the overall phase are absorbed into E_0'. In real instantaneous form, (48) becomes

$$E_y(z,t) = \mathrm{Re}\left(E_{ys}e^{j\omega t}\right) = E_0' \sin(\kappa_m x)\cos(\omega t - \beta_m z) \qquad \text{(TE mode above cutoff)} \tag{49}$$

We interpret this as a wave that propagates in the positive z direction (down the guide) while having a field profile that varies with x.[5] The TE mode field is the *interference pattern* resulting from the superposition of the upward and downward plane waves. Note that if $\omega < \omega_{cm}$, then (42) yields an imaginary value for β_m, which we may write as $-j|\beta_m| = -j\alpha_m$. Eqs. (48) and (49) then become

$$E_{ys} = E_0' \sin(\kappa_m x)e^{-\alpha_m z} \tag{50}$$

$$E(z,t) = E_0' \sin(\kappa_m x)e^{-\alpha_m z} \cos(\omega t) \qquad \text{(TE mode below cutoff)} \tag{51}$$

This mode does not propagate, but simply oscillates at frequency ω, while exhibiting a field pattern that decreases in strength with increasing z. The attenuation coefficient, α_m, is found from (42) with $\omega < \omega_{cm}$:

$$\alpha_m = \frac{n\omega_{cm}}{c}\sqrt{1 - \left(\frac{\omega}{\omega_{cm}}\right)^2} = \frac{2\pi n}{\lambda_{cm}}\sqrt{1 - \left(\frac{\lambda_{cm}}{\lambda}\right)^2} \tag{52}$$

---

[5] We can also interpret this field as that of a standing wave in x while it is a traveling wave in z.

We note from (39) and (41) that the plane wave angle is related to the cutoff frequency and cutoff wavelength through

$$\cos \theta_m = \frac{\omega_{cm}}{\omega} = \frac{\lambda}{\lambda_{cm}} \tag{53}$$

So we see that at cutoff ($\omega = \omega_{cm}$), $\theta_m = 0$, and the plane waves are just reflecting back and forth over the cross section; they are making no forward progress down the guide. As ω is increased beyond cutoff (or λ is decreased), the wave angle increases, approaching 90° as ω approaches infinity (or as λ approaches zero). From Figure 13.14, we have

$$\beta_m = k \sin \theta_m = \frac{n\omega}{c} \sin \theta_m \tag{54}$$

and so the phase velocity of mode m will be

$$v_{pm} = \frac{\omega}{\beta_m} = \frac{c}{n \sin \theta_m} \tag{55}$$

The velocity minimizes at c/n for all modes, approaching this value at frequencies far above cutoff; v_{pm} approaches infinity as the frequency is reduced to approach the cutoff frequency. Again, phase velocity is the speed of the phases in the z direction, and the fact that this velocity may exceed the speed of light in the medium is not a violation of relativistic principles, as discussed in Section 12.7.

The energy will propagate at the group velocity, $v_g = d\omega/d\beta$. Using (42), we have

$$v_{gm}^{-1} = \frac{d\beta_m}{d\omega} = \frac{d}{d\omega} \left[\frac{n\omega}{c} \sqrt{1 - \left(\frac{\omega_{cm}}{\omega} \right)^2} \right] \tag{56}$$

The derivative is straightforward. Carrying it out and taking the reciprocal of the result yields:

$$v_{gm} = \frac{c}{n} \sqrt{1 - \left(\frac{\omega_{cm}}{\omega} \right)^2} = \frac{c}{n} \sin \theta_m \tag{57}$$

Group velocity is thus identified as the projection of the velocity associated with $\mathbf{k}_u$ or $\mathbf{k}_d$ into the z direction. This will be less than or equal to the velocity of light in the medium, c/n, as expected.

EXAMPLE 13.3

In the guide of Example 13.1, the operating frequency is 25 GHz. Consequently, modes for which $m = 1$ and $m = 2$ will be above cutoff. Determine the *group delay difference* between these two modes over a distance of 1 cm. This is the difference in propagation times between the two modes when energy in each propagates over the 1-cm distance.

Solution. The group delay difference is expressed as

$$\Delta t = \left(\frac{1}{v_{g2}} - \frac{1}{v_{g1}} \right) \text{ (s/cm)}$$

From (57), along with the results of Example 13.1, we have

$$v_{g1} = \frac{c}{\sqrt{2.1}} \sqrt{1 - \left(\frac{10.3}{25} \right)^2} = 0.63c$$

$$v_{g2} = \frac{c}{\sqrt{2.1}} \sqrt{1 - \left(\frac{20.6}{25} \right)^2} = 0.39c$$

Then

$$\Delta t = \frac{1}{c} \left[\frac{1}{.39} - \frac{1}{.63} \right] = 3.3 \times 10^{-11} \text{ s/cm} = 33 \text{ ps/cm}$$

This computation gives a rough measure of the *modal dispersion* in the guide, applying to the case of having only two modes propagating. A pulse, for example, whose center frequency is 25 GHz would have its energy divided between the two modes. The pulse would broaden by approximately 33 ps/cm of propagation distance as the energy in the modes separates. If, however, we include the TEM mode (as we really must), then the broadening will be even greater. The group velocity for TEM will be $c/\sqrt{2.1}$. The group delay difference of interest will then be between the TEM mode and the $m = 2$ mode (TE or TM). We would therefore have

$$\Delta t_{\text{net}} = \frac{1}{c} \left[\frac{1}{.39} - 1 \right] = 52 \text{ ps/cm}$$

D13.5. Determine the wave angles θ_m for the first four modes ($m = 1, 2, 3, 4$) in a parallel-plate guide with $d = 2$ cm, $\epsilon_r' = 1$, and $f = 30$ GHz.

Ans. 76°; 60°; 41°; 0°

D13.6. A parallel-plate guide has plate spacing $d = 5$ mm and is filled with glass ($n = 1.45$). What is the maximum frequency at which the guide will operate in the TEM mode only?

Ans. 20.7 GHz

D13.7. A parallel-plate guide having $d = 1$ cm is filled with air. Find the cutoff wavelength for the $m = 2$ mode (TE or TM).

Ans. 1 cm

13.4 PARALLEL-PLATE GUIDE ANALYSIS USING THE WAVE EQUATION

The most direct approach in the analysis of any waveguide is through the wave equation, which we solve subject to the boundary conditions at the conducting walls. The form of the equation that we will use is that of Eq. (28) in Section 11.1, which was written for the case of free-space propagation. We account for the dielectric properties in the waveguide by replacing k_0 in that equation with k to obtain:

$$\nabla^2 \mathbf{E}_s = -k^2 \mathbf{E}_s \qquad (58)$$

where $k = n\omega/c$ as before.

We can use the results of the last section to help us visualize the process of solving the wave equation. For example, we may consider TE modes first, in which there will be only a y component of $\mathbf{E}$. The wave equation becomes:

$$\frac{\partial^2 E_{ys}}{\partial x^2} + \frac{\partial^2 E_{ys}}{\partial y^2} + \frac{\partial^2 E_{ys}}{\partial z^2} + k^2 E_{ys} = 0 \qquad (59)$$

We assume that the width of the guide (in the y direction) is very large compared to the plate separation d. Therefore we can assume no y variation in the fields (fringing fields are ignored), and so $\partial^2 E_{ys}/\partial y^2 = 0$. We also know that the z variation will be of the form $e^{-j\beta_m z}$. The form of the field solution will thus be

$$E_{ys} = E_0 f_m(x) e^{-j\beta_m z} \qquad (60)$$

where E_0 is a constant, and where $f_m(x)$ is a normalized function to be determined (whose maximum value is unity). We have included subscript m on β, κ, and $f(x)$, since we anticipate several solutions that correspond to discrete modes, to which we associate mode number m. We now substitute (60) into (59) to obtain

$$\frac{d^2 f_m(x)}{dx^2} + \left(k^2 - \beta_m^2\right) f_m(x) = 0 \qquad (61)$$

where E_0 and $e^{-j\beta_m z}$ have divided out, and where we have used the fact that

$$\frac{d^2}{dz^2} e^{-j\beta_m z} = -\beta_m^2 e^{-j\beta_m z}$$

Note also that we have written (61) using the total derivative d^2/dx^2, as f_m is a function only of x. We next make use of the geometry of Figure 13.14, and we note

that $k^2 - \beta_m^2 = \kappa_m^2$. Using this in (61) we obtain

$$\frac{d^2 f_m(x)}{dx^2} + \kappa_m^2 f_m(x) = 0 \tag{62}$$

The general solution of (62) will be

$$f_m(x) = \cos(\kappa_m x) + \sin(\kappa_m x) \tag{63}$$

We next apply the appropriate boundary conditions in our problem to evaluate κ_m. From Figure 13.6, conducting boundaries appear at $x = 0$ and $x = d$, at which the tangential electric field (E_y) must be zero. In Eq. (63), only the $\sin(\kappa_m x)$ term will allow the boundary conditions to be satisfied, so we retain it and drop the cosine term. The $x = 0$ condition is automatically satisfied by the sine function. The $x = d$ condition is met when we choose the value of κ_m such that

$$\kappa_m = \frac{m\pi}{d} \tag{64}$$

We recognize Eq. (64) as the same result that we obtained using the transverse resonance condition of Section 13.3. The final form of E_{ys} is obtained by substituting $f_m(x)$ as expressed through (63) and (64) into (60), yielding a result that is consistent with the one expressed in Eq. (48):

$$E_{ys} = E_0 \sin\left(\frac{m\pi x}{d}\right) e^{-j\beta_m z} \tag{65}$$

An additional significance of the mode number m is seen when considering the form of the electric field of (65). Specifically, m is the number of spatial half-cycles of electric field that occur over the distance d in the transverse plane. This can be understood physically by considering the behavior of the guide at cutoff. As we learned in the last section, the plane wave angle of incidence in the guide at cutoff is zero, meaning that the wave simply bounces up and down between the conducting walls. The wave must be resonant in the structure, which means that the net round trip phase shift is $2m\pi$. With the plane waves oriented vertically, $\beta_m = 0$, and so $\kappa_m = k = 2n\pi/\lambda_{cm}$. So at cutoff,

$$\frac{m\pi}{d} = \frac{2n\pi}{\lambda_{cm}} \tag{66}$$

which leads to

$$d = \frac{m\lambda_{cm}}{2n} \qquad \text{at cutoff} \tag{67}$$

Eq. (65) at cutoff then becomes

$$E_{ys} = E_0 \sin\left(\frac{m\pi x}{d}\right) = E_0 \sin\left(\frac{2n\pi x}{\lambda_{cm}}\right) \tag{68}$$

The waveguide is simply a one-dimensional *resonant cavity,* in which a wave can oscillate in the x direction if its wavelength as measured in the medium is an integer multiple of $2d$ where the integer is m.
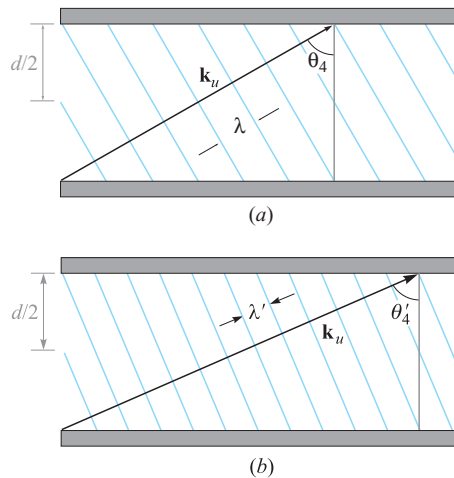
Figure 13.17 (a) A plane wave associated with an $m = 4$ mode, showing a net phase shift of 4π (two wavelengths measured in x) occurring over distance d in the transverse plane. (b) As frequency increases, an increase in wave angle is required to maintain the 4π transverse phase shift.

Now, as the frequency increases, wavelength will decrease, and so the requirement of wavelength equaling an integer multiple of $2d$ is no longer met. The response of the mode is to establish z components of $\mathbf{k}_u$ and $\mathbf{k}_d$, which results in the decreased wavelength being compensated by an increase in wavelength *as measured in the x direction*. Figure 13.17 shows this effect for the $m = 4$ mode, in which the wave angle, θ_4, steadily increases with increasing frequency. Thus, the mode retains precisely the functional form of its field in the x direction, but it establishes an increasing value of β_m as the frequency is raised. This invariance in the transverse spatial pattern means that the mode will retain its identity at all frequencies. Group velocity, expressed in (57), is changing as well, meaning that the changing wave angle with frequency is a mechanism for group velocity dispersion, known simply as *waveguide dispersion*. Pulses, for example, that propagate in a single waveguide mode will thus experience broadening in the manner considered in Section 12.8.

Having found the electric field, we can find the magnetic field using Maxwell's equations. We note from our plane wave model that we expect to obtain x and z components of $\mathbf{H}_s$ for a TE mode. We use the Maxwell equation

$$\nabla \times \mathbf{E}_s = -j\omega\mu\mathbf{H}_s \tag{69}$$

where, in the present case of having only a y component of $\mathbf{E}_s$, we have

$$\nabla \times \mathbf{E}_s = \frac{\partial E_{ys}}{\partial x}\mathbf{a}_z - \frac{\partial E_{ys}}{\partial z}\mathbf{a}_x = \kappa_m E_0 \cos(\kappa_m x)e^{-j\beta_m z}\mathbf{a}_z + j\beta_m E_0 \sin(\kappa_m x)e^{-j\beta_m z}\mathbf{a}_x \tag{70}$$

We solve for $\mathbf{H}_s$ by dividing both sides of (69) by $-j\omega\mu$. Performing this operation on (70), we obtain the two magnetic field components:

$$H_{xs} = -\frac{\beta_m}{\omega\mu} E_0 \sin(\kappa_m x) e^{-j\beta_m z} \tag{71}$$

$$H_{zs} = j\frac{\kappa_m}{\omega\mu} E_0 \cos(\kappa_m x) e^{-j\beta_m z} \tag{72}$$

Together, these two components form closed-loop patterns for $\mathbf{H}_s$ in the x, z plane, as can be verified using the streamline plotting methods developed in Section 2.6.

It is interesting to consider the magnitude of $\mathbf{H}_s$, which is found through

$$|\mathbf{H}_s| = \sqrt{\mathbf{H}_s \cdot \mathbf{H}_s^*} = \sqrt{H_{xs} H_{xs}^* + H_{zs} H_{zs}^*} \tag{73}$$

Carrying this out using (71) and (72) results in

$$|\mathbf{H}_s| = \frac{E_0}{\omega\mu} \left(\kappa_m^2 + \beta_m^2\right)^{1/2} \left(\sin^2(\kappa_m x) + \cos^2(\kappa_m x)\right)^{1/2} \tag{74}$$

Using the fact that $\kappa_m^2 + \beta_m^2 = k^2$ and using the identity $\sin^2(\kappa_m x) + \cos^2(\kappa_m x) = 1$, (74) becomes

$$|\mathbf{H}_s| = \frac{k}{\omega\mu} E_0 = \frac{\omega\sqrt{\mu\epsilon}}{\omega\mu} = \frac{E_0}{\eta} \tag{75}$$

where $\eta = \sqrt{\mu/\epsilon}$. This result is consistent with our understanding of waveguide modes based on the superposition of plane waves, in which the relation between $\mathbf{E}_s$ and $\mathbf{H}_s$ is through the medium intrinsic impedance, η.

D13.8. Determine the group velocity of the $m = 1$ (TE or TM) mode in an air-filled parallel-plate guide with $d = 0.5$ cm at $f = (a)$ 30 GHz, (b) 60 GHz, and (c) 100 GHz.

Ans. 0; 2.6×10^8 m/s; 2.9×10^8 m/s

D13.9. A TE mode in a parallel-plate guide is observed to have three maxima in its electric field pattern between $x = 0$ and $x = d$. What is the value of m?

Ans. 3

13.5 RECTANGULAR WAVEGUIDES

In this section we consider the rectangular waveguide, a structure that is usually used in the microwave region of the electromagnetic spectrum. The guide is shown in Figure 13.7. As always, the propagation direction is along the z axis. The guide is of width a along x and height b along y. We can relate the geometry to that of the parallel-plate guide of previous sections by thinking of the rectangular guide as two

parallel-plate guides of orthogonal orientation that are assembled to form one unit. We have a pair of horizontal conducting walls (along the x direction) and a pair of vertical walls (along y), all of which form one continuous boundary. The wave equation in its full three-dimensional form [Eq. (59)] must now be solved, for in general we may have field variations in all three coordinate directions.

In the parallel-plate guide, we found that the TEM mode can exist, along with TE and TM modes. The rectangular guide will support the TE and TM modes, but it *will not support a TEM mode*. This is because, in contrast to the parallel-plate guide, we now have a conducting boundary that completely surrounds the transverse plane. The nonexistence of TEM can be understood by remembering that any electric field must have a zero tangential component at the boundary. This means that it is impossible to set up an electric field that will not exhibit the sideways variation that is necessary to satisfy this boundary condition. Because **E** varies in the transverse plane, the computation of **H** through $\nabla \times \mathbf{E} = -j\omega\mu\mathbf{H}$ *must* lead to a z component of **H**, and so we cannot have a TEM mode. We cannot find any other orientation of a completely transverse **E** in the guide that will allow a completely transverse **H**.

13.5.1 Using Maxwell's Equations to Relate Field Components

With the modes dividing into TE and TM types, the standard approach is to first solve the wave equation for the z components. By definition, $E_z = 0$ in a TE mode, and $H_z = 0$ in a TM mode. Therefore, we will find the TE mode solution by solving the wave equation for H_z, and we will obtain the TM mode solution by solving for E_z. Using these results, all transverse field components are obtained directly through Maxwell's equations. This procedure may sound a little tedious (which it is), but we can be certain to find all the modes this way. First, we will handle the problem of finding transverse components in terms of the z components.

To begin the process, we assume that the phasor electric and magnetic fields will be forward-z propagating functions that exhibit spatial variation in the xy plane; the only z variation is that of a forward-propagating wave:

$$\mathbf{E}_s(x, y, z) = \mathbf{E}_s(x, y, 0)e^{-j\beta z} \tag{76a}$$

$$\mathbf{H}_s(x, y, z) = \mathbf{H}_s(x, y, 0)e^{-j\beta z} \tag{76b}$$

We can then obtain expressions for the transverse components of the phasor fields by evaluating the x and y components of the Maxwell curl equations in sourceless media. In evaluating the curl, it is evident from (76) that $\partial/\partial z = -j\beta$. The result is

$$\nabla \times \mathbf{E}_s = -j\omega\mu\mathbf{H}_s \;\rightarrow\; \begin{cases} \partial E_{zs}/\partial y + j\beta E_{ys} = -j\omega\mu H_{xs} & (x \text{ component}) \quad (77a) \\ j\beta E_{xs} + \partial E_{zs}/\partial x = j\omega\mu H_{ys} & (y \text{ component}) \quad\;\; (77b) \end{cases}$$

$$\nabla \times \mathbf{H}_s = j\omega\epsilon\mathbf{E}_s \;\rightarrow\; \begin{cases} \partial H_{zs}/\partial y + j\beta H_{ys} = j\omega\epsilon E_{xs} & (x \text{ component}) \quad\;\;\; (78a) \\ j\beta H_{xs} + \partial H_{zs}/\partial x = -j\omega\epsilon E_{ys} & (y \text{ component}) \quad\;\;\; (78b) \end{cases}$$

Now, pairs of the above equations can be solved together in order to express the individual transverse field components in terms of derivatives of the z components of

E and **H**. For example, (77a) and (78b) can be combined, eliminating E_{ys}, to give

$$H_{xs} = \frac{-j}{\kappa^2} \left[\beta \frac{\partial H_{zs}}{\partial x} - \omega\epsilon \frac{\partial E_{zs}}{\partial y} \right] \tag{79a}$$

Then, using (76b) and (77a), eliminate E_{xs} between them to obtain

$$H_{ys} = \frac{-j}{\kappa^2} \left[\beta \frac{\partial H_{zs}}{\partial y} + \omega\epsilon \frac{\partial E_{zs}}{\partial x} \right] \tag{79b}$$

Using the same equation pairs, the transverse electric field components are then found:

$$E_{xs} = \frac{-j}{\kappa^2} \left[\beta \frac{\partial E_{zs}}{\partial x} + \omega\mu \frac{\partial H_{zs}}{\partial y} \right] \tag{79c}$$

$$E_{ys} = \frac{-j}{\kappa^2} \left[\beta \frac{\partial E_{zs}}{\partial y} - \omega\mu \frac{\partial H_{zs}}{\partial x} \right] \tag{79d}$$

κ is defined in the same manner as in the parallel-plate guide [Eq. (35)]:

$$\kappa = \sqrt{k^2 - \beta^2} \tag{80}$$

where $k = \omega\sqrt{\mu\epsilon}$. In the parallel-plate geometry, we found that discrete values of κ and β resulted from the analysis, which we then subscripted with the integer mode number, m (κ_m and β_m). The interpretation of m was the number of field maxima that occurred between plates (in the x direction). In the rectangular guide, field variations will occur in both x and y, and so we will find it necessary to assign *two* integer subscripts to κ and β, thus leading to

$$\kappa_{mp} = \sqrt{k^2 - \beta_{mp}^2} \tag{81}$$

where m and p indicate the number of field variations in the x and y directions. The form of Eq. (81) suggests that plane wave (ray) theory could be used to obtain the mode fields in the rectangular guide, as was accomplished in Section 13.3 for the parallel-plate guide. This is, in fact, the case, and is readily accomplished for cases in which plane wave reflections occur between only two opposing boundaries (either top to bottom or side to side), and this would be true only for certain TE modes. The method becomes complicated when reflections occur at all four surfaces; but in any case, the interpretation of κ_{mp} is the transverse (xy plane) component of the plane wave-vector k, while β_{mp} is the z component, as before.

The next step is to solve the wave equation for the z components of **E** and **H**, from which we will find the fields of the TM and TE modes.

13.5.2 TM Modes

Finding the TM modes begins with the wave equation [Eq. (59)], in which derivatives with respect to z are equivalent to multiplying by $j\beta$. We write the equation for the z

component of $\mathbf{E}_s$:

$$\frac{\partial^2 E_{zs}}{\partial x^2} + \frac{\partial^2 E_{zs}}{\partial y^2} + (k^2 - \beta_{mp}^2)E_{zs} = 0 \tag{82}$$

The solution of (82) can be written as a sum of terms, each of which involves the product of three functions that exhibit individual variation with x, y, and z:

$$\boxed{E_{zs}(x, y, z) = \sum_{m,p} F_m(x)\, G_p(y) \exp(-j\beta_{mp}\, z)} \tag{83}$$

where the functions $F_m(x)$ and $G_p(y)$ (not normalized) are to be determined. Each term in (83) corresponds to one mode of the guide, and will by itself be a solution to (82). To determine the functions, a single term in (83) is substituted into (82). Noting that all derivatives are applied to functions of a single variable (and thus partial derivatives become total derivatives), and using (81), the result is

$$G_p(y)\frac{d^2 F_m}{dx^2} + F_m(x)\frac{d^2 G_p}{dy^2} + \kappa_{mp}^2\, F_m(x)\, G_p(y) = 0 \tag{84}$$

in which the $\exp(-j\beta_{mp}\, z)$ term has divided out. Rearranging (84), we get

$$\underbrace{\frac{1}{F_m}\frac{d^2 F_m}{dx^2}}_{-\kappa_m^2} + \underbrace{\frac{1}{G_p}\frac{d^2 G_p}{dy^2}}_{-\kappa_p^2} + \kappa_{mp}^2 = 0 \tag{85}$$

Terms in (85) are grouped such that all of the x variation is in the first term, which varies *only* with x, and all y variation is in the second term, which varies *only* with y. Now, consider what would happen if x is allowed to vary while holding y fixed. The second and third terms would be fixed, and Eq. (85) must always hold. Therefore, the x-varying first term must be a constant. This constant is denoted $-\kappa_m^2$, as indicated in (85). The same is true for the second term, which must also turn out to be a constant if y is allowed to vary while x is fixed. We assign the second term the constant value $-\kappa_p^2$ as indicated. Eq. (85) then states that

$$\boxed{\kappa_{mp}^2 = \kappa_m^2 + \kappa_p^2} \tag{86}$$

which suggests an immediate geometrical interpretation: As κ_{mp} is the transverse plane component of the wavevector k, κ_m and κ_p are clearly the x and y components of κ_{mp} (and of k)—again if one thinks in terms of plane waves and how they would bounce around in the waveguide to form the overall mode. Also indicated in (86) is the fact that κ_m and κ_p will be functions, respectively, of the integers m and p, as we will find.

Under the above conditions, Eq. (85) will now separate into two equations—one in each variable:

$$\frac{d^2 F_m}{dx^2} + \kappa_m^2\, F_m = 0 \tag{87a}$$

$$\frac{d^2 G_p}{dy^2} + \kappa_p^2\, G_p = 0 \tag{87b}$$

Equation (87) is now easily solved. We obtain

$$F_m(x) = A_m \cos(\kappa_m x) + B_m \sin(\kappa_m x) \tag{88a}$$

$$G_p(y) = C_p \cos(\kappa_p y) + D_p \sin(\kappa_p y) \tag{88b}$$

Using these, along with (83), the general solution for z component of $\mathbf{E}_s$ for a single TM mode can be constructed:

$$E_{zs} = [A_m \cos(\kappa_m x) + B_m \sin(\kappa_m x)][C_p \cos(\kappa_p y) + D_p \sin(\kappa_p y)] \exp(-j\beta_{mp} z) \tag{89}$$

The constants in (89) can be evaluated by applying the boundary conditions of the field on all four surfaces. Specifically, as E_{zs} is tangent to all the conducting surfaces, it must vanish on all of them. Referring to Figure 13.7, the boundary conditions are

$$E_{zs} = 0 \text{ at } x = 0, \ y = 0, \ x = a, \ \text{and } y = b$$

Obtaining zero field at $x = 0$ and $y = 0$ is accomplished by dropping the cosine terms in (89) (setting $A_m = C_p = 0$). The values of κ_m and κ_p that appear in the remaining sine terms are then set to the following, in order to assure zero field at $x = a$ and $y = b$:

$$\kappa_m = \frac{m\pi}{a} \tag{90a}$$

$$\kappa_p = \frac{p\pi}{b} \tag{90b}$$

Using these results, and defining $B = B_m D_p$, Eq. (89) becomes:

$$E_{zs} = B \sin(\kappa_m x) \sin(\kappa_p y) \exp(-j\beta_{mp} z) \tag{91a}$$

Now, to find the remaining (transverse) field components, we substitute Eq. (91a) into Eqs. (79) to obtain:

$$E_{xs} = -j\beta_{mp} \frac{\kappa_m}{\kappa_{mp}^2} B \cos(\kappa_m x) \sin(\kappa_p y) \exp(-j\beta_{mp} z) \tag{91b}$$

$$E_{ys} = -j\beta_{mp} \frac{\kappa_p}{\kappa_{mp}^2} B \sin(\kappa_m x) \cos(\kappa_p y) \exp(-j\beta_{mp} z) \tag{91c}$$

$$H_{xs} = j\omega\epsilon \frac{\kappa_p}{\kappa_{mp}^2} B \sin(\kappa_m x) \cos(\kappa_p y) \exp(-j\beta_{mp} z) \tag{91d}$$

$$H_{ys} = -j\omega\epsilon \frac{\kappa_m}{\kappa_{mp}^2} B \cos(\kappa_m x) \sin(\kappa_p y) \exp(-j\beta_{mp} z) \tag{91e}$$

The above field components pertain to modes designated TM$_{mp}$. Note that for these modes, both m and p must be greater than or equal to 1. A zero value for either integer will zero all fields.

13.5.3 TE Modes

To obtain the TE mode fields, we solve the wave equation for the z component of $\mathbf{H}$ and then use Eq. (79) as before to find the transverse components. The wave equation is now the same as (82), except that E_{zs} is replaced by H_{zs}:

$$\frac{\partial^2 H_{zs}}{\partial x^2} + \frac{\partial^2 H_{zs}}{\partial y^2} + (k^2 - \beta_{mp}^2)H_{zs} = 0 \tag{92}$$

and the solution is of the form:

$$H_{zs}(x, y, z) = \sum_{m,p} F'_m(x)\, G'_p(y) \exp(-j\beta_{mp}\, z) \tag{93}$$

The procedure from here is identical to that involving TM modes, and the general solution will be

$$H_{zs} = [A'_m \cos(\kappa_m x) + B'_m \sin(\kappa_m x)][C'_p \cos(\kappa_p y) + D'_p \sin(\kappa_p y)] \exp(-j\beta_{mp}\, z) \tag{94}$$

Again, the expression is simplified by using the appropriate boundary conditions. We know that tangential electric field must vanish on all conducting boundaries. When we relate the electric field to magnetic field derivatives using (79c) and (79d), the following conditions develop:

$$E_{xs}\Big|_{y=0,b} = 0 \;\Rightarrow\; \frac{\partial H_{zs}}{\partial y}\Big|_{y=0,b} = 0 \tag{95a}$$

$$E_{ys}\Big|_{x=0,a} = 0 \;\Rightarrow\; \frac{\partial H_{zs}}{\partial x}\Big|_{x=0,a} = 0 \tag{95b}$$

The boundary conditions are now applied to Eq. (94), giving, for Eq. (95a)

$$\frac{\partial H_{zs}}{\partial y} = [A'_m \cos(\kappa_m x) + B'_m \sin(\kappa_m x)]$$

$$\times \underline{[-\kappa_p C'_p \sin(\kappa_p y) + \kappa_p D'_p \cos(\kappa_p y)]} \exp(-j\beta_{mp}\, z)$$

The underlined terms are those that were modified by the partial differentiation. Requiring this result to be zero at $y = 0$ and $y = b$ leads to dropping the $\cos(\kappa_p y)$ term (set $D'_p = 0$) and requiring that $\kappa_p = p\pi/b$ as before. Applying Eq. (95b) to (94) results in

$$\frac{\partial H_{zs}}{\partial x} = \underline{[-\kappa_m A'_m \sin(\kappa_m x) + \kappa_m B'_m \cos(\kappa_m x)]}$$

$$\times [C'_p \cos(\kappa_p y) + D'_p \sin(\kappa_p y)] \exp(-j\beta_{mp}\, z)$$

where again, the underlined term has been modified by partial differentiation with respect to x. Setting this result to zero at $x = 0$ and $x = a$ leads to dropping the $\cos(\kappa_m x)$ term (setting $B'_m = 0$), and requiring that $\kappa_m = m\pi/a$ as before. With all the above boundary conditions applied, the final expression for H_{zs} is now

$$H_{zs} = A \cos(\kappa_m x) \cos(\kappa_p y) \exp(-j\beta_{mp}\, z) \tag{96a}$$

where we define $A = A'_m C'_p$. Applying Eqs. (79a) through (79d) to (96a) gives the transverse field components:

$$H_{xs} = j\beta_{mp} \frac{\kappa_m}{\kappa_{mp}^2} A \sin(\kappa_m x) \cos(\kappa_p y) \exp(-j\beta_{mp} z) \tag{96b}$$

$$H_{ys} = j\beta_{mp} \frac{\kappa_p}{\kappa_{mp}^2} A \cos(\kappa_m x) \sin(\kappa_p y) \exp(-j\beta_{mp} z) \tag{96c}$$

$$E_{xs} = j\omega\mu \frac{\kappa_p}{\kappa_{mp}^2} A \cos(\kappa_m x) \sin(\kappa_p y) \exp(-j\beta_{mp} z) \tag{96d}$$

$$E_{ys} = -j\omega\mu \frac{\kappa_m}{\kappa_{mp}^2} A \sin(\kappa_m x) \cos(\kappa_p y) \exp(-j\beta_{mp} z) \tag{96e}$$

These field components pertain to modes designated TE$_{mp}$. For these modes, either m or p may be zero, thus allowing the possibility of the important TE$_{m0}$ or TE$_{0p}$ cases, as will be discussed later. Some very good illustrations of TE and TM modes are presented in Ref. 3.

13.5.4 Cutoff Conditions

The phase constant for a given mode can be expressed using Eq. (81):

$$\beta_{mp} = \sqrt{k^2 - \kappa_{mp}^2} \tag{97}$$

Then, using (86), along with (90a) and (90b), we have

$$\beta_{mp} = \sqrt{k^2 - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{p\pi}{b}\right)^2} \tag{98}$$

This result can be written in a manner consistent with Eq. (42) by using $k = \omega\sqrt{\mu\epsilon}$, and defining a radian cutoff frequency, $\omega_{c_{mp}}$, appropriate for the rectangular guide. We obtain:

$$\beta_{mp} = \omega\sqrt{\mu\epsilon}\sqrt{1 - \left(\frac{\omega_{Cmp}}{\omega}\right)^2} \tag{99}$$

where

$$\omega_{Cmp} = \frac{1}{\sqrt{\mu\epsilon}}\left[\left(\frac{m\pi}{a}\right)^2 + \left(\frac{p\pi}{b}\right)^2\right]^{1/2} \tag{100}$$

As discussed for the parallel-plate guide, it is again clear from (99) that the operating frequency, ω, must exceed the cutoff frequency, ω_{Cmp}, to achieve a real value for β_{mp} (and thus enabling mode mp to propagate). Equation (100) applies to both TE and TM modes, and thus some combination of both mode types may be present (or above cutoff) at a given frequency. It is evident that the choice of guide dimensions,

a and b, along with the material properties, ϵ_r and μ_r, will determine the number of modes that will propagate. For the typical case in which $\mu_r = 1$, using $n = \sqrt{\epsilon_r}$, and identifying the speed of light, $c = 1/\sqrt{\mu_0 \epsilon_0}$, we may re-write (100) in a manner consistent with Eq. (41):

$$\omega_{Cmp} = \frac{c}{n} \left[\left(\frac{m\pi}{a} \right)^2 + \left(\frac{p\pi}{b} \right)^2 \right]^{1/2} \tag{101}$$

This would lead to an expression for the cutoff wavelength, λ_{Cmp}, in a manner consistent with Eq. (43):

$$\lambda_{Cmp} = \frac{2\pi c}{\omega_{Cmp}} = 2n \left[\left(\frac{m}{a} \right)^2 + \left(\frac{p}{b} \right)^2 \right]^{-1/2} \tag{102}$$

λ_{Cmp} is the *free space* wavelength at cutoff. If measured in the medium that fills the waveguide, the cutoff wavelength would be given by Eq. (102) divided by n.

Now, in a manner consistent with Eq. (44), Eq. (99) becomes

$$\beta_{mp} = \frac{2\pi n}{\lambda} \sqrt{1 - \frac{\lambda}{\lambda_{Cmp}}} \tag{103}$$

where λ is the free space wavelength. As we saw before, a TE_{mp} or TM_{mp} mode can propagate if its operating wavelength, λ, is less than λ_{Cmp}.

13.5.5 Special Cases: TE_{m0} and TE_{0p} Modes

The most important mode in the rectangular guide is the one that can propagate by itself. As we know, this will be the mode that has the lowest cutoff frequency (or the highest cutoff wavelength), so that over a certain range of frequencies, this mode will be above cutoff, while all others are below cutoff. By inspecting Eq. (101), and noting that $a > b$, the lowest cutoff frequency will occur for the mode in which $m = 1$ and $p = 0$, which will be the TE_{10} mode (remember that a TM_{10} mode does not exist, as can be shown in (91)). It turns out that this mode, and those of the same general type, are of the same form as those of the parallel-plate structure.

The specific fields for the TE_{m0} family of modes are obtained from (96a) through (96e) by setting $p = 0$, which means, using (86) and (90), that

$$\kappa_m = \kappa_{mp} \Big|_{p=0} = \frac{m\pi}{a} \tag{104}$$

and $\kappa_p = 0$. Under these conditions, the only surviving field components in (91) will be E_{ys}, H_{xs}, and H_{zs}. It is convenient to define the field equations in terms of an electric field amplitude, E_0, which is composed of all the amplitude terms in Eq. (96e). Specifically, define

$$E_0 = -j\omega\mu \frac{\kappa_m}{\kappa_{m0}^2} A = -j \frac{\omega\mu}{\kappa_m} A \tag{105}$$

Substituting (104) and (105) into Eqs. (96e), (96c), and (96a) leads to the following expressions for the TE$_{m0}$ mode fields:

$$E_{ys} = E_0 \sin{(\kappa_m x)} e^{-j\beta_{m0}z} \tag{106}$$

$$H_{xs} = -\frac{\beta_{m0}}{\omega\mu} E_0 \sin{(\kappa_m x)} e^{-j\beta_{m0}z} \tag{107}$$

$$H_{zs} = j\frac{\kappa_m}{\omega\mu} E_0 \cos(\kappa_m x) e^{-j\beta_{m0}z} \tag{108}$$

It can be seen that these expressions are identical to the parallel-plate fields, Eqs. (65), (71), and (72). For TE$_{m0}$, we again note that the subscripts indicate that there are m half cycles of the electric field over the x dimension and there is zero variation in y. The cutoff frequency for the TE$_{m0}$ mode is given by (101), appropriately modified:

$$\omega_{Cm0} = \frac{m\pi c}{na} \tag{109}$$

Using (109) in (99), the phase constant is

$$\beta_{m0} = \frac{n\omega}{c}\sqrt{1 - \left(\frac{m\pi c}{\omega na}\right)^2} \tag{110}$$

All of the implications on mode behavior above and below cutoff are exactly the same as we found for the parallel-plate guide. The plane wave analysis is also carried out in the same manner. TE$_{m0}$ modes can be modeled as plane waves that propagate down the guide by reflecting between the vertical side walls.

The electric field of the fundamental (TE$_{10}$) mode is, from (106):

$$E_{ys} = E_0 \sin\left(\frac{\pi x}{a}\right) e^{-j\beta_{10}z} \tag{111}$$

This field is plotted in Figure 13.18a. The field is vertically polarized, terminates on the top and bottom plates, and becomes zero at the two vertical walls, as is required from our boundary condition on a tangential electric field at a conducting surface. Its cutoff wavelength is found from (102) to be

$$\lambda_{C10} = 2na \tag{112}$$

which means that cutoff for this mode is achieved when the guide horizontal dimension, a, is equal to a half-wavelength (as measured in the medium).

Another possibility is the TE$_{0p}$ field configuration, which consists of a horizontally polarized electric field. Figure 13.18b shows the field for TE$_{01}$. The specific fields for the TE$_{0p}$ family are obtained from Eqs. (96a) through (96e) by setting $m = 0$,
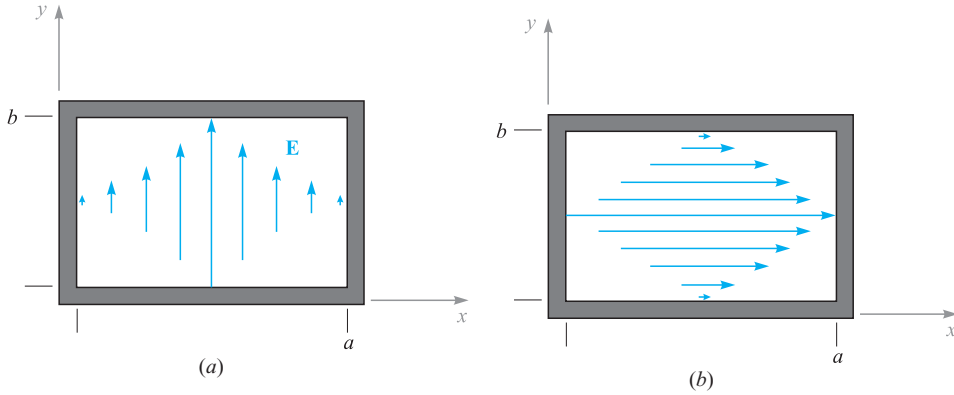
Figure 13.18 (a) TE$_{10}$ and (b) TE$_{01}$ mode electric field configurations in a rectangular waveguide.

which means, using (86) and (90), that

$$\kappa_p = \kappa_{mp}\Big|_{m=0} = \frac{p\pi}{b} \tag{113}$$

and $\kappa_m = 0$. Now, the surviving field components in Eqs. (91a) through (91e) will be E_{xs}, H_{ys}, and H_{zs}. Now, define the electric field amplitude, E_0', which is composed of all the amplitude terms in Eq. (96d):

$$E_0' = j\omega\mu\frac{\kappa_p}{\kappa_{0p}^2} A = j\frac{\omega\mu}{\kappa_p} A \tag{114}$$

Using (113) and (114) in Eqs. (96d), (96b), and (96a) leads to the following expressions for the TE$_{0p}$ mode fields:

$$E_{xs} = E_0 \sin\left(\kappa_p y\right) e^{-j\beta_{0p}z} \tag{115}$$

$$H_{ys} = \frac{\beta_{0p}}{\omega\mu} E_0 \sin\left(\kappa_p y\right) e^{-j\beta_{0p}z} \tag{116}$$

$$H_{zs} = -j\frac{\kappa_p}{\omega\mu} E_0 \cos(\kappa_p y) e^{-j\beta_{0p}z} \tag{117}$$

where the cutoff frequency will be

$$\omega_{C0p} = \frac{p\pi c}{nb} \tag{118}$$

EXAMPLE 13.4

An air-filled rectangular waveguide has dimensions $a = 2$ cm and $b = 1$ cm. Determine the range of frequencies over which the guide will operate single mode (TE_{10}).

Solution. Since the guide is air-filled, $n = 1$, and (109) gives, for $m = 1$:

$$f_{C10} = \frac{\omega_{C10}}{2\pi} = \frac{c}{2a} = \frac{3 \times 10^{10}}{2(2)} = 7.5 \text{ GHz}$$

The next higher-order mode will be either TE_{20} or TE_{01}, which, from (100) will have the same cutoff frequency because $a = 2b$. This frequency will be twice that found for TE_{10}, or 15 GHz. Thus the operating frequency range over which the guide will be single mode is 7.5 GHz $< f <$ 15 GHz.

Having seen how rectangular waveguides work, questions arise: why are they used and when are they useful? Let us consider for a moment the operation of a transmission line at frequencies high enough such that waveguide modes can occur. The onset of guided modes in a transmission line, known as *moding*, is in fact a problem that needs to be avoided, because signal distortion may result. A signal that is input to such a line will find its power divided in some proportions among the various modes. The signal power in each mode propagates at a group velocity unique to that mode. With the power thus distributed, distortion will occur over sufficient distances, as the signal components among the modes lose synchronization with each other, owing to the different delay times (group delays) associated with the different modes. We encountered this concept in Example 13.3.

The above problem of *modal dispersion* in transmission lines is avoided by ensuring that only the TEM mode propagates, and that all waveguide modes are below cutoff. This is accomplished either by using line dimensions that are smaller than one-half the signal wavelength, or by assuring an upper limit to the operating frequency in a given line. But it is more complicated than this.

In Section 13.1, we saw that increasing the frequency increases the line loss as a result of the skin effect. This is manifested through the increase in the series resistance per unit length, R. One can compensate by increasing one or more dimensions in the line cross section, as shown for example in Eqs. (7) and (12), but only to the point at which moding may occur. Typically, the increasing loss with increasing frequency will render the transmission line useless before the onset of moding, but one still cannot increase the line dimensions to reduce losses without considering the possibility of moding. This limitation on dimensions also limits the power handling capability of the line, as the voltage at which dielectric breakdown occurs decreases with decreasing conductor separation. Consequently, the use of transmission lines, as frequencies are increased beyond a certain point, becomes undesirable, as losses will become excessive, and as the limitation on dimensions will limit the power-handling capability. Instead, we look to other guiding structures, among which is the rectangular guide.

Because the rectangular guide will not support a TEM mode, it will not operate until the frequency exceeds the cutoff frequency of the lowest-order guided mode of

the structure. Thus, the guide must be constructed large enough to accomplish this for a given frequency; the required transverse dimensions will consequently be larger than those of a transmission line that is designed to support only the TEM mode. The increased size, coupled with the fact that there is more conductor surface area than in a transmission line of equal volume, means that losses will be substantially lower in the rectangular waveguide structure. Additionally, the guides will support more power at a given electric field strength than a transmission line, as the rectangular guide will have a higher cross-sectional area.

Still, hollow pipe guides must operate in a single mode in order to avoid the signal distortion problems arising from multimode transmission. This means that the guides must be of dimension such that they operate above the cutoff frequency of the lowest-order mode, but below the cutoff frequency of the next higher-order mode, as demonstrated in Example 13.4. Increasing the operating frequency again means that the guide transverse dimensions must be decreased to maintain single mode operation. This can be accomplished to a point at which skin effect losses again become problematic (remember that the skin depth is decreasing with increasing frequency, in addition to the decrease in metal surface area with diminishing guide size). In addition, the guides become too difficult to fabricate, with machining tolerances becoming more stringent. So again, as frequencies are further increased, we look for another type of structure.

D13.10. Specify the minimum width, a, and the maximum height, b, of an air-filled rectangular guide so that it will operate in a single mode over the frequency range $15\,\text{GHz} < f < 20\,\text{GHz}$.

Ans. 1 cm; 0.75 cm

13.6   PLANAR DIELECTRIC WAVEGUIDES

When skin effect losses become excessive, a good way to remove them is to remove the metal in the structure entirely and use interfaces between dielectrics for the confining surfaces. We thus obtain a *dielectric waveguide;* a basic form, the *symmetric slab waveguide,* is shown in Figure 13.19. The structure is so named because of its vertical symmetry about the z axis. The guide is assumed to have width in y much greater than the slab thickness d, so the problem becomes two-dimensional, with fields presumed to vary with x and z while being independent of y. The slab waveguide works in very much the same way as the parallel-plate waveguide, except wave reflections occur at the interfaces between dielectrics, having different refractive indices, n_1 for the slab and n_2 for the surrounding regions above and below. In the dielectric guide, total reflection is needed, so the incident angle must exceed the critical angle. Consequently, as discussed in Section 12.6, the slab index, n_1, must be greater than that of the surrounding materials, n_2. Dielectric guides differ from conducting guides in that power is not completely confined to the slab but resides partially above and below.
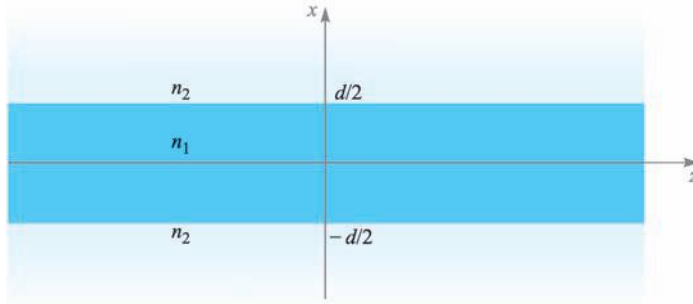
Figure 13.19 Symmetric dielectric slab waveguide structure, in which waves propagate along z. The guide is assumed to be infinite in the y direction, thus making the problem two-dimensional.

Dielectric guides are used primarily at optical frequencies (on the order of 10^{14} Hz). Again, guide transverse dimensions must be kept on the order of a wavelength to achieve operation in a single mode. A number of fabrication methods can be used to accomplish this. For example, a glass plate can be doped with materials that will raise the refractive index. The doping process allows materials to be introduced only within a thin layer adjacent to the surface that is a few micrometers thick.

To understand how the guide operates, consider Figure 13.20, which shows a wave propagating through the slab by multiple reflections, but where *partial transmission* into the upper and lower regions occurs at each bounce. Wavevectors are shown in the middle and upper regions, along with their components in the x and z directions. As we found in Chapter 12, the z components (β) of all wavevectors are equal, as must be true if the field boundary conditions at the interfaces are to be satisfied for all positions and times. Partial transmission at the boundaries is, of course, an undesirable situation, since power in the slab will eventually leak away. We thus have a *leaky wave* propagating in the structure, whereas we need to have a guided mode. Note that in either case, we still have the two possibilities of wave polarization, and the resulting mode designation—either TE or TM.

Total power reflection at the boundaries for TE or TM waves implies, respectively, that $|\Gamma_s|^2$ or $|\Gamma_p|^2$ is unity, where the reflection coefficients are given in Eqs. (71) and (69) in Chapter 12.

$$\Gamma_s = \frac{\eta_{2s} - \eta_{1s}}{\eta_{2s} + \eta_{1s}} \tag{119}$$

and

$$\Gamma_p = \frac{\eta_{2p} - \eta_{1p}}{\eta_{2p} + \eta_{1p}} \tag{120}$$

As discussed in Section 12.6, we require that the effective impedances, η_{2s} or η_{2p}, be purely imaginary, zero, or infinite if (119) or (120) is to have unity magnitudes.
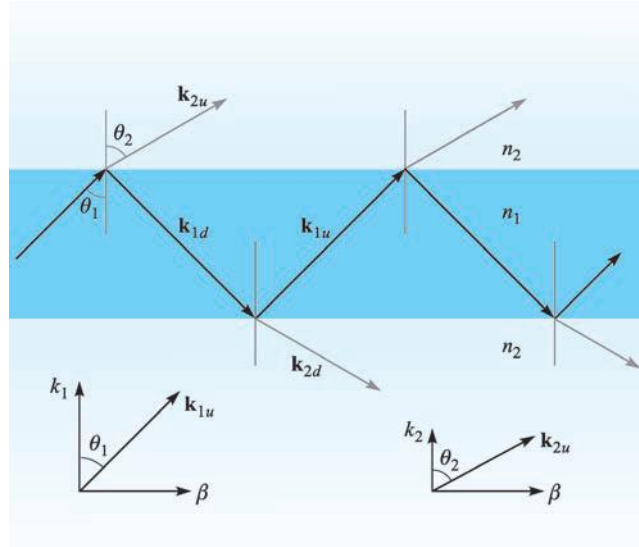
Figure 13.20 Plane wave geometry of a leaky wave in a symmetric slab waveguide. For a guided mode, total reflection occurs in the interior, and the x components of $\mathbf{k}_{2u}$ and $\mathbf{k}_{2d}$ are imaginary.

Knowing that

$$\eta_{2s} = \frac{\eta_2}{\cos \theta_2} \tag{121}$$

and

$$\eta_{2p} = \eta_2 \cos \theta_2 \tag{122}$$

the requirement is that $\cos \theta_2$ be zero or imaginary, where, from Eq. (75), Section 12.6,

$$\cos \theta_2 = \left[1 - \sin^2 \theta_2 \right]^{1/2} = \left[1 - \left(\frac{n_1}{n_2} \right)^2 \sin^2 \theta_1 \right]^{1/2} \tag{123}$$

As a result, we require that

$$\theta_1 \geq \theta_c \tag{124}$$

where the critical angle is defined through

$$\sin \theta_c = \frac{n_2}{n_1} \tag{125}$$

Now, from the geometry of Figure 13.20, we can construct the field distribution of a TE wave in the guide using plane wave superposition. In the slab region ($-d/2 < x < d/2$), we have

$$E_{y1s} = E_0 e^{-j\mathbf{k}_{1u} \cdot \mathbf{r}} \pm E_0 e^{-j\mathbf{k}_{1d} \cdot \mathbf{r}} \qquad \left(-\frac{d}{2} < x < \frac{d}{2} \right) \tag{126}$$

where

$$\mathbf{k}_{1u} = \kappa_1 \mathbf{a}_x + \beta \mathbf{a}_z \tag{127}$$

and

$$\mathbf{k}_{1d} = -\kappa_1 \mathbf{a}_x + \beta \mathbf{a}_z \tag{128}$$

The second term in (126) may either add to or subtract from the first term, since either case would result in a symmetric intensity distribution in the x direction. We expect this because the guide is symmetric. Now, using $\mathbf{r} = x\mathbf{a}_x + z\mathbf{a}_z$, (126) becomes

$$E_{y1s} = E_0[e^{j\kappa_1 x} + e^{-j\kappa_1 x}]e^{-j\beta z} = 2E_0 \cos(\kappa_1 x)e^{-j\beta z} \tag{129}$$

for the choice of the plus sign in (126), and

$$E_{y1s} = E_0[e^{j\kappa_1 x} - e^{-j\kappa_1 x}]e^{-j\beta z} = 2jE_0 \sin(\kappa_1 x)e^{-j\beta z} \tag{130}$$

if the minus sign is chosen. Because $\kappa_1 = n_1 k_0 \cos\theta_1$, we see that larger values of κ_1 imply smaller values of θ_1 at a given frequency. In addition, larger κ_1 values result in a greater number of spatial oscillations of the electric field over the transverse dimension, as (129) and (130) show. We found similar behavior in the parallel-plate guide. In the slab waveguide, as with the parallel-plate guide, we associate higher-order modes with increasing values of κ_1.[6]

In the regions above and below the slab, waves propagate according to wave-vectors $\mathbf{k}_{2u}$ and $\mathbf{k}_{2d}$ as shown in Figure 13.20. Above the slab, for example $(x > d/2)$, the TE electric field will be of the form

$$E_{y2s} = E_{02}e^{-j\mathbf{k}_2 \cdot \mathbf{r}} = E_{02}e^{-j\kappa_2 x}e^{-j\beta z} \tag{131}$$

However, $\kappa_2 = n_2 k_0 \cos\theta_2$, where $\cos\theta_2$, given in (123), is imaginary. We may therefore write

$$\kappa_2 = -j\gamma_2 \tag{132}$$

where γ_2 is real and is given by (using 123)

$$\gamma_2 = j\kappa_2 = jn_2 k_0 \cos\theta_2 = jn_2 k_0(-j)\left[\left(\frac{n_1}{n_2}\right)^2 \sin^2\theta_1 - 1\right]^{1/2} \tag{133}$$

Equation (131) now becomes

$$E_{y2s} = E_{02}e^{-\gamma_2(x-d/2)}e^{-j\beta z} \qquad \left(x > \frac{d}{2}\right) \tag{134}$$

---

[6] It would be appropriate to add the mode number subscript, m, to κ_1, κ_2, β, and θ_1, because, as was true with the metal guides, we will obtain discrete values of these quantities. To keep notation simple, the m subscript is suppressed, and we will assume it to be understood. Again, subscripts 1 and 2 in this section indicate, respectively, the slab and surrounding *regions,* and have nothing to do with mode number.
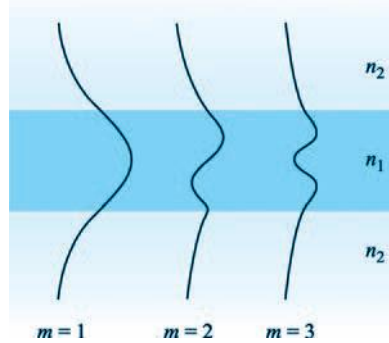
Figure 13.21 Electric field amplitude distributions over the transverse plane for the first three TE modes in a symmetric slab waveguide.

where the x variable in (131) has been replaced by $x - (d/2)$ to position the field magnitude, E_{02}, at the boundary. Using similar reasoning, the field in the region below the lower interface, where x is negative, and where $\mathbf{k}_{2d}$ is involved, will be

$$E_{y2s} = E_{02}e^{\gamma_2(x+d/2)}e^{-j\beta z} \qquad \left(x < -\frac{d}{2}\right) \tag{135}$$

The fields expressed in (134) and (135) are those of *surface* waves. Note that they propagate in the z direction only, according to $e^{-j\beta z}$, but simply reduce in amplitude with increasing $|x|$, according to the $e^{-\gamma_2(x-d/2)}$ term in (134) and the $e^{\gamma_2(x+d/2)}$ term in (135). These waves represent a certain fraction of the total power in the mode, and so we see an important fundamental difference between dielectric waveguides and metal waveguides: in the dielectric guide, the fields (and guided power) exist over a cross section that extends beyond the confining boundaries, and in principle they exist over an infinite cross section. In practical situations, the exponential decay of the fields above and below the boundaries is typically sufficient to render the fields negligible within a few slab thicknesses from each boundary.

The total electric field distribution is composed of the field in all three regions and is sketched in Figure 13.21 for the first few modes. Within the slab, the field is oscillatory and is of a similar form to that of the parallel-plate waveguide. The difference is that the fields in the slab waveguide do not reach zero at the boundaries but connect to the evanescent fields above and below the slab. The restriction is that the TE fields on either side of a boundary (being tangent to the interface) must match at the boundary. Specifically,

$$E_{y1s}|_{x=\pm d/2} = E_{y2s}|_{x=\pm d/2} \tag{136}$$

Applying this condition to (129), (130), (134), and (135) results in the final expressions for the TE electric field in the symmetric slab waveguide, for the cases of even and

odd symmetry:

$$E_{se}(\text{even TE}) = \begin{cases} E_{0e}\cos(\kappa_1 x)e^{-j\beta z} & \left(-\frac{d}{2} < x < \frac{d}{2}\right) \\ E_{0e}\cos\left(\kappa_1\frac{d}{2}\right)e^{-\gamma_2(x-d/2)}e^{-j\beta z} & \left(x > \frac{d}{2}\right) \\ E_{0e}\cos\left(\kappa_1\frac{d}{2}\right)e^{\gamma_2(x+d/2)}e^{-j\beta z} & \left(x < -\frac{d}{2}\right) \end{cases} \tag{137}$$

$$E_{so}(\text{odd TE}) = \begin{cases} E_{0o}\sin(\kappa_1 x)e^{-j\beta z} & \left(-\frac{d}{2} < x < \frac{d}{2}\right) \\ E_{0o}\sin\left(\kappa_1\frac{d}{2}\right)e^{-\gamma_2(x-d/2)}e^{-j\beta z} & \left(x > \frac{d}{2}\right) \\ -E_{0o}\sin\left(\kappa_1\frac{d}{2}\right)e^{\gamma_2(x+d/2)}e^{-j\beta z} & \left(x < -\frac{d}{2}\right) \end{cases} \tag{138}$$

Solution of the wave equation yields (as it must) results identical to these. The reader is referred to References 2 and 3 for the details. The magnetic field for the TE modes will consist of x and z components, as was true for the parallel-plate guide. Finally, the TM mode fields will be nearly the same in form as those of TE modes, but with a simple rotation in polarization of the plane wave components by $90°$. Thus, in TM modes, H_y will result, and it will have the same form as E_y for TE, as presented in (137) and (138).

Apart from the differences in the field structures, the dielectric slab waveguide operates in a manner that is qualitatively similar to the parallel-plate guide. Again, a finite number of discrete modes will be allowed at a given frequency, and this number increases as frequency increases. Higher-order modes are characterized by successively smaller values of θ_1.

An important difference in the slab waveguide occurs at cutoff for any mode. We know that $\theta = 0$ at cutoff in the metal guides. In the dielectric guide at cutoff, the wave angle, θ_1, is equal to the *critical angle, θ_c*. Then, as the frequency of a given mode is raised, its θ_1 value increases beyond θ_c in order to maintain transverse resonance, while maintaining the same number of field oscillations in the transverse plane.

As wave angle increases, however, the character of the evanescent fields changes significantly. This can be understood by considering the wave angle dependence on evanescent decay coefficient, γ_2, as given by (133). Note, in that equation, that as θ_1 increases (as frequency goes up), γ_2 also increases, leading to a more rapid falloff of the fields with increasing distance above and below the slab. The mode therefore becomes more tightly confined to the slab as frequency is raised. Also, at a given frequency, lower-order modes, having smaller wave angles, will have lower values of γ_2 as (133) indicates. Consequently, when considering several modes propagating together at a single frequency, the higher-order modes will carry a greater percentage of their power in the upper and lower regions surrounding the slab than will modes of lower order.

One can determine the conditions under which modes will propagate by using the transverse resonance condition, as we did with the parallel-plate guide. We perform the transverse round trip analysis in the slab region in the same manner that was done in Section 13.3, and obtain an equation similar to (37):

$$\boxed{\kappa_1 d + \phi_{TE} + \kappa_1 d + \phi_{TE} = 2m\pi} \tag{139}$$

for TE waves and

$$\kappa_1 d + \phi_{TM} + \kappa_1 d + \phi_{TM} = 2m\pi \tag{140}$$

for the TM case. Eqs. (139) and (140) are called the *eigenvalue equations* for the symmetric dielectric slab waveguide. The phase shifts on reflection, ϕ_{TE} and ϕ_{TM}, are the phases of the reflection coefficients, Γ_s and Γ_p, given in (119) and (120). These are readily found, but they turn out to be functions of θ_1. As we know, κ_1 also depends on θ_1, but in a different way than ϕ_{TE} and ϕ_{TM}. Consequently, (139) and (140) are *transcendental* in θ_1, and they cannot be solved in closed form. Instead, numerical or graphical methods must be used (see References 4 or 5). Emerging from this solution process, however, is a fairly simple cutoff condition for any TE or TM mode:

$$k_0 d \sqrt{n_1^2 - n_2^2} \geq (m - 1)\pi \qquad (m = 1, 2, 3, \ldots) \tag{141}$$

For mode m to propagate, (141) must hold. The physical interpretation of the mode number m is again the number of half-cycles of the electric field (for TE modes) or magnetic field (for TM modes) that occur over the transverse dimension. The lowest-order mode ($m = 1$) is seen to have no cutoff—it will propagate from zero frequency on up. We will thus achieve single-mode operation (actually a single pair of TE and TM modes) if we can assure that the $m = 2$ modes are below cutoff. Using (141), our single-mode condition will thus be:

$$k_0 d \sqrt{n_1^2 - n_2^2} < \pi \tag{142}$$

Using $k_0 = 2\pi/\lambda$, the wavelength range over which single-mode operation occurs is

$$\lambda > 2d \sqrt{n_1^2 - n_2^2} \tag{143}$$

EXAMPLE 13.5

A symmetric dielectric slab waveguide is to guide light at wavelength $\lambda = 1.30 \ \mu$m. The slab thickness is to be $d = 5.00 \ \mu$m, and the refractive index of the surrounding material is $n_2 = 1.450$. Determine the maximum allowable refractive index of the slab material that will allow single TE and TM mode operation.

Solution. Equation (143) can be rewritten in the form,

$$n_1 < \sqrt{\left(\frac{\lambda}{2d}\right)^2 + n_2^2}$$

Thus

$$n_1 < \sqrt{\left(\frac{1.30}{2(5.00)}\right)^2 + (1.450)^2} = 1.456$$

Clearly, fabrication tolerances are very exacting when constructing dielectric guides for single-mode operation!

D13.11. A 0.5-mm-thick slab of glass ($n_1 = 1.45$) is surrounded by air ($n_2 = 1$). The slab waveguides infrared light at wavelength $\lambda = 1.0\ \mu$m. How many TE and TM modes will propagate?

Ans. 2102

13.7 OPTICAL FIBER

Optical fiber works on the same principle as the dielectric slab waveguide, except of course for the round cross section. A *step index* fiber is shown in Figure 13.10, in which a high index *core* of radius a is surrounded by a lower-index *cladding* of radius b. Light is confined to the core through the mechanism of total reflection, but again some fraction of the power resides in the cladding as well. As we found in the slab waveguide, the cladding power again moves in toward the core as frequency is raised. Additionally, as is true in the slab waveguide, the fiber supports a mode that has no cutoff.

Analysis of the optical fiber is complicated. This is mainly because of the round cross section, along with the fact that it is generally a three-dimensional problem; the slab waveguide had only two dimensions to be concerned about. It is possible to analyze the fiber using rays within the core that reflect from the cladding boundary as light progresses down the fiber. We did this with the slab waveguide and obtained results fairly quickly. The method is difficult in fiber, however, because ray paths are complicated. There are two types of rays in the core: (1) those that pass through the fiber axis (z axis), known as *meridional rays,* and (2) those that avoid the axis but progress in a spiral-like path as they propagate down the guide. These are known as *skew rays;* their analysis, although possible, is tedious. Fiber modes are developed that can be associated with the individual ray types, or with combinations thereof, but it is easier to obtain these by solving the wave equation directly. Our purpose in this section is to provide a first exposure to the optical fiber problem (and to avoid an excessively long treatment). To accomplish this, we will solve the simplest case in the quickest way.

The simplest fiber configuration is that of a step index, but with the core and cladding indices of values that are very close, that is $n_1 \doteq n_2$. This is the *weak-quidance* condition, whose simplifying effect on the analysis is significant. We already saw how core and cladding indices in the slab waveguide need to be very close in value in order to achieve single-mode or few-mode operation. Fiber manufacturers have taken this result to heart, such that the weak-guidance condition is in fact satisfied by most commercial fibers today. Typical dimensions of a single-mode fiber are between 5 and 10 μm for the core diameter, with the cladding diameter usually 125 μm. Refractive index differences between core and cladding are typically a small fraction of a percent.

The main result of the weak-guidance condition is that a set of modes appears in which each mode is *linearly polarized*. This means that light having x-polarization, for example, will enter the fiber and establish itself in a mode or in a set of modes that preserve the x-polarization. Magnetic field is essentially orthogonal to **E**, and so it would in that case lie in the y direction. The z components of both fields, although present, are too weak to be of significance; the nearly equal core and cladding indices lead to ray paths that are essentially parallel to the guide axis—deviating only slightly. In fact, we may write for a given mode, $E_x \doteq \eta H_y$, when η is approximated as the intrinsic impedance of the cladding. Therefore, in the weak-guidance approximation, the fiber mode fields are treated as plane waves (nonuniform, of course). The designation for these modes is $\mathrm{LP}_{\ell m}$, meaning linearly polarized, with integer order parameters ℓ and m. The latter express the numbers of variations over the two dimensions in the circular transverse plane. Specifically, ℓ, the *azimuthal mode number,* is one-half the number of power density maxima (or minima) that occur at a given radius as ϕ varies from 0 to 2π. The *radial mode number, m*, expresses the number of maxima that occur along a radial line (at constant ϕ) that extends from zero to infinity.

Although we may assume a linearly polarized field in a rectangular coordinate system, we are obliged to work in cylindrical coordinates for obvious reasons. In a manner that reminds us of the rectangular waveguide, it is possible to write the x-polarized phasor electric field within a weakly guiding cylindrical fiber as a product of three functions, each of which varies with one of the coordinate variables, ρ, ϕ, and z:

$$E_{xs}(\rho, \phi, z) = \sum_i R_i(\rho)\Phi_i(\phi)\exp(-j\beta_i z) \qquad (144)$$

Each term in the summation is an individual mode of the fiber. Note that the z function is just the propagation term, $e^{-j\beta z}$, since we are assuming an infinitely long lossless fiber.

The wave equation is Eq. (58), which we may write for the assumed x component of **E**$_s$, but in which the Laplacian operator is written in cylindrical coordinates:

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial^2 E_{xs}}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2 E_{xs}}{\partial\phi^2} + (k^2 - \beta^2)E_{xs} = 0 \qquad (145)$$

where we recognize that the $\partial^2/\partial z^2$ operation, when applied to (144), leads to a factor of $-\beta^2$. We now substitute a single term of (144) into (145) [since each term in (144) should alone satisfy the wave equation]. Dropping the subscript i, expanding the radial derivative, and rearranging terms, we obtain:

$$\underbrace{\frac{\rho^2}{R}\frac{d^2 R}{d\rho^2} + \frac{\rho}{R}\frac{dR}{d\rho} + \rho^2(k^2 - \beta^2)}_{\ell^2} = \underbrace{-\frac{1}{\Phi}\frac{d^2\Phi}{d\phi^2}}_{\ell^2} \qquad (146)$$

We note that the left-hand side of (146) varies only with ρ, whereas the right-hand side varies only with ϕ. Since the two variables are independent, it must follow that each side of the equation must be equal to a constant. Calling this constant ℓ^2, as shown,

we may write separate equations for each side; the variables are now separated:

$$\frac{d^2\Phi}{d\phi^2} + \ell^2\Phi = 0 \tag{147a}$$

$$\frac{d^2R}{d\rho^2} + \frac{1}{\rho}\frac{dR}{d\rho} + \left[k^2 - \beta^2 - \frac{\ell^2}{\rho^2}\right]R = 0 \tag{147b}$$

The solution of (147a) is of the form of the sine or cosine of ϕ:

$$\Phi(\phi) = \begin{cases} \cos(\ell\phi + \alpha) \\ \sin(\ell\phi + \alpha) \end{cases} \tag{148}$$

where α is a constant. The form of (148) dictates that ℓ must be an integer, since the same mode field must occur in the transverse plane as ϕ is changed by 2π radians. Since the fiber is round, the orientation of the x and y axes in the transverse plane is immaterial, so we may choose the cosine function and set $\alpha = 0$. We will thus use $\Phi(\phi) = \cos(\ell\phi)$.

The solution of (147b) to obtain the radial function is more complicated. Eq. (147b) is a form of Bessel's equation, whose solutions are Bessel functions of various forms. The key parameter is the function $\beta_t = (k^2 - \beta^2)^{1/2}$, the square of which appears in (147b). Note that β_t will differ in the two regions: Within the core ($\rho < a$), $\beta_t = \beta_{t1} = (n_1^2 k_0^2 - \beta^2)^{1/2}$; within the cladding ($\rho > a$), we have $\beta_t = \beta_{t2} = (n_2^2 k_0^2 - \beta^2)^{1/2}$. Depending on the relative magnitudes of k and β, β_t may be real or imaginary. These possibilities lead to two solution forms of (147b):

$$R(\rho) = \begin{cases} AJ_\ell(\beta_t\rho) & \beta_t \text{ real} \\ BK_\ell(|\beta_t|\rho) & \beta_t \text{ imaginary} \end{cases} \tag{149}$$

where A and B are constants. $J_\ell(\beta_t\rho)$ is the ordinary Bessel function of the first kind, of order ℓ and of argument $\beta_t\rho$. $K_\ell(|\beta_t|\rho)$ is the modified Bessel function of the second kind, of order ℓ, and having argument $|\beta_t|\rho$. The first two orders of each of these functions are illustrated in Figures 13.22a and b. In our study, it is necessary to know the precise zero crossings of the J_0 and J_1 functions. Those shown in Figure 13.22a are as follows: For J_0, the zeros are 2.405, 5.520, 8.654, 11.792, and 14.931. For J_1, the zeros are 0, 3.832, 7.016, 10.173, and 13.324. Other Bessel function types would contribute to the solutions in Eq. (149), but these exhibit nonphysical behavior with radius and are not included.

We next need to determine which of the two solutions is appropriate for each region. Within the core ($\rho < a$) we expect to get an oscillatory solution for the field—much in the same manner as we found in the slab waveguide. Therefore, we assign the ordinary Bessel function solutions to that region by requiring that $\beta_{t1} = (n_1^2 k_0^2 - \beta^2)^{1/2}$ is real. In the cladding ($\rho > a$), we expect surface waves that decrease in amplitude with increasing radius away from the core/cladding boundary.
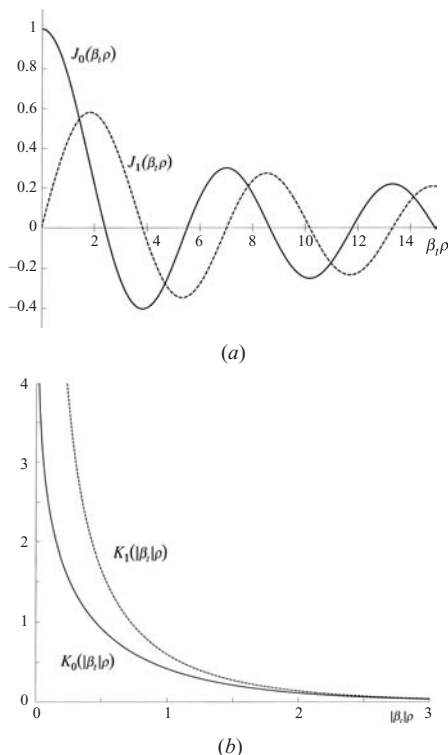
Figure 13.22 (a) Ordinary Bessel functions of the first kind, of orders 0 and 1, and of argument $\beta_t \rho$, where β_t is real. (b) Modified Bessel functions of the second kind, of orders 0 and 1, and of argument $|\beta_t|\rho$, where β_t is imaginary.

The Bessel K functions provide this behavior and will apply if β_{t2} is imaginary. Requiring this, we may therefore write $|\beta_{t2}| = (\beta^2 - n_2^2 k_0^2)^{1/2}$. The diminishing field amplitude with increasing radius within the cladding allows us to neglect the effect of the outer cladding boundary (at $\rho = b$), as fields there are presumed too weak for this boundary to have any effect on the mode field.

Because β_{t1} and β_{t2} are in units of m^{-1}, it is convenient to normalize these quantities (while making them dimensionless) by multiplying both by the core radius, a. Our new normalized parameters become

$$u \equiv a\beta_{t1} = a\sqrt{n_1^2 k_0^2 - \beta^2} \qquad (150a)$$

$$w \equiv a|\beta_{t2}| = a\sqrt{\beta^2 - n_2^2 k_0^2} \qquad (150b)$$

u and w are in direct analogy with the quantities $\kappa_1 d$ and $\kappa_2 d$ in the slab waveguide. As in those parameters, β is the z component of *both* $n_1 k_0$ and $n_2 k_0$ and is the phase constant of the guided mode. β must be the same in both regions so that the field boundary conditions will be satisfied at $\rho = a$ for all z and t.

We may now construct the total solution for E_{xs} for a single guided mode, using (144) along with (148), (149), (150a), and (150b):

$$E_{xs} = \begin{cases} E_0 J_\ell(u\rho/a)\cos(\ell\phi)e^{-j\beta z} & \rho \le a \\ E_0[J_\ell(u)/K_\ell(w)]K_\ell(w\rho/a)\cos(\ell\phi)e^{-j\beta z} & \rho \ge a \end{cases} \tag{151}$$

Note that we have let the coefficient A in (149) equal E_0, and $B = E_0[J_\ell(u)/K_\ell(w)]$. These choices assure that the expressions for E_{xs} in the two regions become equal at $\rho = a$, a condition approximately true as long as $n_1 \doteq n_2$ (the weak-guidance approximation).

Again, the weak-guidance condition also allows the approximation $H \doteq E/\eta$, with η taken as the intrinsic impedance of the cladding. Having $\mathbf{E}_s$ and $\mathbf{H}_s$ enables us to find the LP$_{\ell m}$ mode average power density (or light intensity) through

$$|\langle \mathbf{S} \rangle| = \left| \frac{1}{2}\text{Re}\{\mathbf{E}_s \times \mathbf{H}_s^*\} \right| = \frac{1}{2}\text{Re}\{E_{xs}H_{ys}^*\} = \frac{1}{2\eta}|E_{xs}|^2 \tag{152}$$

Using (151) in (152), the mode intensity in W/m^2 becomes

$$I_{\ell m} = I_0 J_\ell^2\left(\frac{u\rho}{a}\right)\cos^2(\ell\phi) \qquad \rho \le a \tag{153a}$$

$$I_{\ell m} = I_0 \left(\frac{J_\ell(u)}{K_\ell(w)}\right)^2 K_\ell^2\left(\frac{w\rho}{a}\right)\cos^2(\ell\phi) \qquad \rho \ge a \tag{153b}$$

where I_0 is the peak intensity value. The role of the azimuthal mode number ℓ, as evident in (153a) and (153b), is to determine the number of intensity variations around the circle, $0 < \phi < 2\pi$; it also determines the order of the Bessel functions that are used. The influence of the radial mode number, m, is not immediately apparent in (153a) and (153b). Briefly stated, m determines the range of allowed values of u that occur in the Bessel function, $J(u\rho/a)$. The greater the value of m, the greater the allowed values of u; with larger u, the Bessel function goes through more oscillations over the range $0 < \rho < a$, and so more radial intensity variations occur with larger m. In the slab waveguide, the mode number (also m) determines the allowed ranges of κ_1. As we saw in Section 13.6, increasing κ_1 at a given frequency means that the slab ray propagates closer to the normal (smaller θ_1), and so more spatial oscillations of the field occur in the transverse direction (larger m).

The final step in the analysis is to obtain an equation from which values of mode parameters (u, w, and β, for example) can be determined for a given operating frequency and fiber construction. In the slab waveguide, two equations, (139) and (140), were found using transverse resonance arguments, and these were associated

with TE and TM waves in the slab. In our fiber, we do not apply transverse resonance directly, but rather *implicitly,* by requiring that all fields satisfy the boundary conditions at the core/cladding interface, $\rho = a$.[7] We have already applied conditions on the transverse fields to obtain Eq. (151). The remaining condition is continuity of the z components of **E** and **H**. In the weak-guidance approximation, we have neglected all z components, but we will consider them now for this last exercise. Using Faraday's law in point form, continuity of H_{zs} at $\rho = a$ is the same as the continuity of the z component of $\nabla \times \mathbf{E}_s$, provided that $\mu = \mu_0$ (or is the same value) in both regions. Specifically

$$(\nabla \times \mathbf{E}_{s1})_z\big|_{\rho=a} = (\nabla \times \mathbf{E}_{s2})_z\big|_{\rho=a} \tag{154}$$

The procedure begins by expressing the electric field in (151) in terms of ρ and ϕ components and then applying (154). This is a lengthy procedure and is left as an exercise (or may be found in Reference 5). The result is the eigenvalue equation for LP modes in the weakly guiding step index fiber:

$$\frac{J_{\ell-1}(u)}{J_\ell(u)} = -\frac{w}{u}\frac{K_{\ell-1}(w)}{K_\ell(w)} \tag{155}$$

This equation, like (139) and (140), is transcendental, and it must be solved for u and w numerically or graphically. This exercise in all of its aspects is beyond the scope of our treatment. Instead, we will obtain from (155) the conditions for cutoff for a given mode and some properties of the most important mode—that which has no cutoff, and which is therefore the mode that is present in single-mode fiber.

The solution of (155) is facilitated by noting that u and w can be combined to give a new parameter that is independent of β and depends only on the fiber construction and on the operating frequency. This new parameter, called the *normalized frequency,* or V number, is found using (150a) and (150b):

$$V \equiv \sqrt{u^2 + w^2} = ak_0\sqrt{n_1^2 - n_2^2} \tag{156}$$

We note that an increase in V is accomplished through an increase in core radius, frequency, or index difference.

The cutoff condition for a given mode can now be found from (155) in conjunction with (156). To do this, we note that cutoff in a dielectric guide means that total reflection at the core/cladding boundary just ceases, and power just begins to propagate radially, away from the core. The effect on the electric field of Eq. (151) is to produce a cladding field that no longer diminishes with increasing radius. This occurs in the modified Bessel function, $K(w\rho/a)$, when $w = 0$. This is our general cutoff condition,

---

[7] Recall that the equations for reflection coefficient (119) and (120), from which the phase shift on reflection used in transverse resonance is determined, originally came from the application of the field boundary conditions.

which we now apply to (155), whose right-hand side becomes zero when $w = 0$. This leads to cutoff values of u and V (u_c and V_c), and, by (156), $u_c = V_c$. Eq. (155) at cutoff now becomes:

$$J_{\ell-1}(V_c) = 0 \tag{157}$$

Finding the cutoff condition for a given mode is now a matter of finding the appropriate zero of the relevant ordinary Bessel function, as determined by (157). This gives the value of V at cutoff for that mode.

For example, the lowest-order mode is the simplest in structure; therefore it has no variations in ϕ and one variation (one maximum) in ρ. The designation for this mode is therefore LP$_{01}$, and with $\ell = 0$, (157) gives the cutoff condition as $J_{-1}(V_c) = 0$. Because $J_{-1} = J_1$ (true only for the J_1 Bessel function), we take the first zero of J_1, which is $V_c(01) = 0$. The LP$_{01}$ mode therefore has no cutoff and will propagate at the exclusion of all other modes provided V for the fiber is greater than zero but less than V_c for the next-higher-order mode. By inspecting Figure 13.22a, we see that the next Bessel function zero is 2.405 (for the J_0 function). Therefore, $\ell - 1 = 0$ in (156), and so $\ell = 1$ for the next-higher-order mode. Also, we use the lowest value of $m_\ell (m = 1)$, and the mode is therefore identified as LP$_{11}$. Its cutoff V is $V_c(11) = 2.405$. If $m = 2$ were to be chosen instead, we would obtain the cutoff V number for the LP$_{12}$ mode. We use the next zero of the J_0 function, which is 5.520, or $V_c(12) = 5.520$. In this way, the radial mode number, m, numbers the zeros of the Bessel function of order $\ell - 1$, taken in order of increasing value.

When we follow the reasoning just described, the condition for single-mode operation in a step index fiber is found to be

$$V < V_c(11) = 2.405 \tag{158}$$

Then, using (156) along with $k_0 = 2\pi/\lambda$, we find

$$\lambda > \lambda_c = \frac{2\pi a}{2.405}\sqrt{n_1^2 - n_2^2} \tag{159}$$

as the requirement on free-space wavelength to achieve single-mode operation in a step index fiber. The similarity to the single-mode condition in the slab waveguide [Eq. (143)] is apparent. The *cutoff wavelength, λ_c,* is that for the LP$_{11}$ mode. Its value is quoted as a specification of most commercial single-mode fiber.

EXAMPLE 13.6

The cutoff wavelength of a step index fiber is quoted as $\lambda_c = 1.20$ μm. If the fiber is operated at wavelength $\lambda = 1.55$ μm, what is V?

Solution. Using (156) and (159), we find

$$V = 2.405\frac{\lambda_c}{\lambda} = 2.405\left(\frac{1.20}{1.55}\right) = 1.86$$

The intensity profiles of the first two modes can be found using (153a) and (153b), having determined u and w values for each mode from (155). For LP$_{01}$, we find

$$
I_{01} = \begin{cases} I_0 J_0^2(u_{01}\rho/a) & \rho \le a \\ I_0 \left(\frac{J_0(u_{01})}{K_0(w_{01})} \right)^2 K_0^2(w_{01}\rho/a) & \rho \ge a \end{cases} \tag{160}
$$

and for LP$_{11}$ we find

$$
I_{11} = \begin{cases} I_0 J_1^2(u_{11}\rho/a)\cos^2\phi & \rho \le a \\ I_0 \left(\frac{J_1(u_{11})}{K_1(w_{11})} \right)^2 K_1^2(w_{11}\rho/a)\cos^2\phi & \rho \ge a \end{cases} \tag{161}
$$

The two intensities for a single V value are plotted as functions of radius at $\phi = 0$ in Figure 13.23. We again note the lower confinement of the higher-order mode to the core, as was true in the slab waveguide.

As V increases (accomplished by increasing the frequency, for example), existing modes become more tightly confined to the core, while new modes of higher order may begin to propagate. The behavior of the lowest-order mode with changing V is depicted in Figure 13.24, where we again note that the mode becomes more tightly confined as V increases. In determining the intensities, Eq. (155) must in general be solved numerically to obtain u and w. Various analytic approximations to the exact numerical solution exist, the best of which is the Rudolf-Neumann formula for the LP$_{01}$ mode, valid over the range $1.3 < V < 3.5$:

$$
w_{01} \doteq 1.1428V - 0.9960 \tag{162}
$$

Having w_{01}, u_{01} can be found from (156), knowing V.

Another important simplification for the LP$_{01}$ mode is the approximation of its intensity profile by a Gaussian function. An inspection of any of the intensity plots of Figure 13.24 shows a resemblence to a Gaussian, which would be expressed as

$$
I_{01} \approx I_0 e^{-2\rho^2/\rho_0^2} \tag{163}
$$

where ρ_0, termed the *mode field radius,* is defined as the radius from the fiber axis at which the mode intensity falls to $1/e^2$ times its on-axis value. This radius depends on frequency, and most generally on V. A similar approximation can be made for the fundamental symmetric slab waveguide mode intensity. In step index fiber, the best fit between the Gaussian approximation and the actual mode intensity as given in (160) is given by the Marcuse formula:

$$
\frac{\rho_0}{a} \approx 0.65 + \frac{1.619}{V^{3/2}} + \frac{2.879}{V^6} \tag{164}
$$
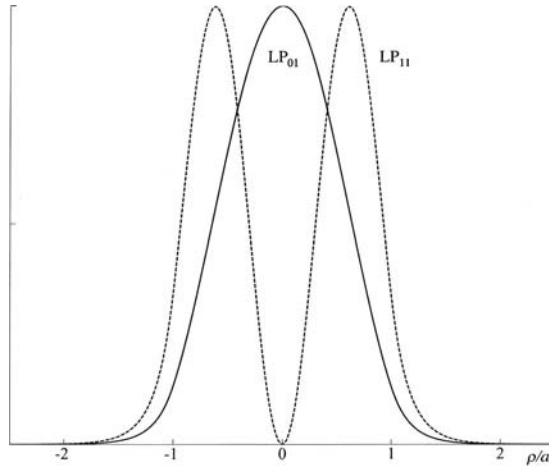
Figure 13.23 Intensity plots from Eqs. (160) and (161) of the first two LP modes in a weakly guiding step index fiber, as functions of normalized radius, ρ/a. Both functions were evaluated at the same operating frequency; the relatively weak confinement of the LP_{11} mode compared to that of LP_{01} is evident.

The mode field radius (at a quoted wavelength) is another important specification (along with the cutoff wavelength) of commercial single-mode fiber. It is important to know for several reasons: First, in splicing or connecting two single-mode fibers together, the lowest connection loss will be attained if both fibers have the same mode field radius, and if the fiber axes are precisely aligned. Different radii or displaced axes result in increased loss, but this can be calculated and compared with measurement. Alignment tolerance (allowable deviation from precise axis alignment) is relaxed somewhat if the fibers have larger mode field radii. Second, a smaller mode field radius means that the fiber is less likely to suffer loss as a result of bending. A loosely confined mode tends to radiate away more as the fiber is bent. Finally, mode field radius is directly related to the mode phase constant, β, since if u and w are known (found from ρ_0), β can be found from (150a) or (150b). Therefore, knowledge of how β changes with frequency (leading to the quantification of dispersion) can be found by measuring the change in mode field radius with frequency. Again, References 4 and 5 (and references therein) provide more detail.

D13.12. For the fiber of Example 13.6, the core radius is given as $a = 5.0\,\mu m$. Find the mode field radius at wavelengths (a) 1.55 μm; (b) 1.30 μm.

Ans. 6.78 μm; 5.82 μm
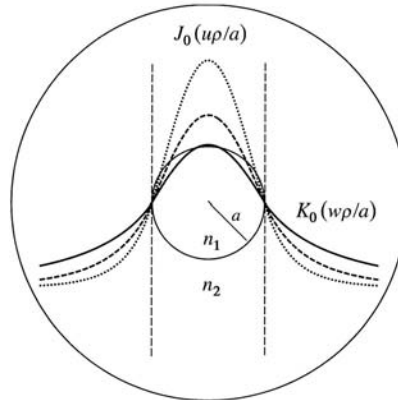
Figure 13.24   Intensity plots for the LP_{01} mode in a weakly guiding step index fiber. Traces are shown for $V = 1.0$ (solid), $V = 1.2$ (dashed), and $V = 1.5$ (dotted), corresponding to increases in frequency in those proportions. Dashed vertical lines indicate the core/cladding boundary, at which for all three cases, the J_0 radial dependence in the core connects to the K_0 radial dependence in the cladding, as demonstrated in Eq. (160). The migration of mode power toward the fiber axis as frequency increases is evident.

REFERENCES

1.   Weeks, W. L. *Transmission and Distribution of Electrical Energy*. New York: Harper and Row, 1981. Line parameters for various configurations of power transmission and distribution systems are discussed in Chapter 2, along with typical parameter values.

2.   Edwards, T. C. *Foundations for Microstrip Circuit Design*. Chichester, N.Y.: Wiley-Interscience, 1981. Chapters 3 and 4 provide an excellent treatment of microstrip lines, with many design formulas.

3.   Ramo, S., J. R. Whinnery, and T. Van Duzer. *Fields and Waves in Communication Electronics*. 3d ed. New York: John Wiley & Sons, 1990. In-depth treatment of parallel-plate and rectangular waveguides is presented in Chapter 8.

4.   Marcuse, D. *Theory of Dielectric Optical Waveguides*. 2d ed. New York: Academic Press, 1990. This book provides a very general and complete discussion of dielectric slab waveguides, plus other types.

5.   Buck, J. A. *Fundamentals of Optical Fibers*. 2d ed. New York: Wiley-Interscience, 2004. Symmetric slab dielectric guides and weakly guiding fibers are emphasized in this book by one of the coauthors.

CHAPTER 13   PROBLEMS

13.1 The conductors of a coaxial transmission line are copper ($\sigma_c = 5.8 \times 10^7$ S/m), and the dielectric is polyethylene ($\epsilon'_r = 2.26$, $\sigma/\omega\epsilon' = 0.0002$). If the inner radius of the outer conductor is 4 mm, find the radius of the inner conductor so that (a) $Z_0 = 50$ Ω; (b) $C = 100$ pF/m; (c) $L = 0.2\,\mu$H/m. A lossless line can be assumed.

13.2 Find R, L, C, and G for a coaxial cable with $a = 0.25$ mm, $b = 2.50$ mm, $c = 3.30$ mm, $\epsilon_r = 2.0$, $\mu_r = 1$, $\sigma_c = 1.0 \times 10^7$ S/m, $\sigma = 1.0 \times 10^{-5}$ S/m, and $f = 300$ MHz.

13.3 Two aluminum-clad steel conductors are used to construct a two-wire transmission line. Let $\sigma_{Al} = 3.8 \times 10^7$ S/m, $\sigma_{St} = 5 \times 10^6$ S/m, and $\mu_{St} = 100\,\mu$H/m. The radius of the steel wire is 0.5 in., and the aluminum coating is 0.05 in. thick. The dielectric is air, and the center-to-center wire separation is 4 in. Find C, L, G, and R for the line at 10 MHz.

13.4 Find R, L, C, and G for a two-wire transmission line in polyethylene at $f = 800$ MHz. Assume copper conductors of radius 0.50 mm and separation 0.80 cm. Use $\epsilon_r = 2.26$ and $\sigma/(\omega\epsilon') = 4.0 \times 10^{-4}$.

13.5 Each conductor of a two-wire transmission line has a radius of 0.5 mm; their center-to-center separation is 0.8 cm. Let $f = 150$ MHz, and assume σ and σ_c are zero. Find the dielectric constant of the insulating medium if (a) $Z_0 = 300$ Ω; (b) $C = 20$ pF/m; (c) $v_p = 2.6 \times 10^8$ m/s.

13.6 The transmission line in Fig. 6.8 is filled with polyethylene. If it were filled with air, the capacitance would be 57.6 pF/m. Assuming that the line is lossless, find C, L, and Z_0.

13.7 Pertinent dimensions for the transmission line shown in Figure 13.2 are $b = 3$ mm and $d = 0.2$ mm. The conductors and the dielectric are nonmagnetic. (a) If the characteristic impedance of the line is 15 Ω, find ϵ'_r. Assume a low-loss dielectric. (b) Assume copper conductors and operation at 2×10^8 rad/s. If $RC = GL$, determine the loss tangent of the dielectric.

13.8 A transmission line constructed from perfect conductors and an air dielectric is to have a maximum dimension of 8 mm for its cross section. The line is to be used at high frequencies. Specify the dimensions if it is (a) a two-wire line with $Z_0 = 300$ Ω; (b) a planar line with $Z_0 = 15$ Ω; (c) a 72 Ω coax having a zero-thickness outer conductor.

13.9 A microstrip line is to be constructed using a lossless dielectric for which $\epsilon'_r = 7.0$. If the line is to have a 50 Ω characteristic impedance, determine (a) $\epsilon_{r,\text{eff}}$; (b) w/d.

13.10 Two microstrip lines are fabricated end-to-end on a 2-mm-thick wafer of lithium niobate ($\epsilon'_r = 4.8$). Line 1 is of 4 mm width; line 2 (unfortunately)
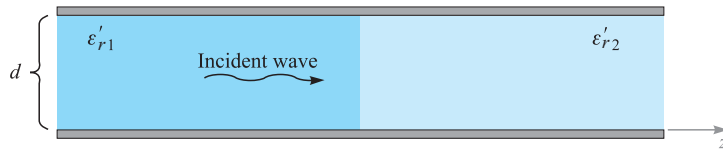
Figure 13.25   See Problems 13.17 and 13.18.

has been fabricated with a 5 mm width. Determine the power loss in dB for waves transmitted through the junction.

13.11 A parallel-plate waveguide is known to have a cutoff wavelength for the $m = 1$ TE and TM modes of $\lambda_{c1} = 4.1$ mm. The guide is operated at wavelength $\lambda = 1.0$ mm. How many modes propagate?

13.12 A parallel-plate guide is to be constructed for operation in the TEM mode only over the frequency range $0 < f < 3$ GHz. The dielectric between plates is to be teflon ($\epsilon_r' = 2.1$). Determine the maximum allowable plate separation, d.

13.13 A lossless parallel-plate waveguide is known to propagate the $m = 2$ TE and TM modes at frequencies as low as 10 GHz. If the plate separation is 1 cm, determine the dielectric constant of the medium between plates.

13.14 A $d = 1$ cm parallel-plate guide is made with glass ($n = 1.45$) between plates. If the operating frequency is 32 GHz, which modes will propagate?

13.15 For the guide of Problem 13.14, and at the 32 GHz frequency, determine the difference between the group delays of the highest-order mode (TE or TM) and the TEM mode. Assume a propagation distance of 10 cm.

13.16 The cutoff frequency of the $m = 1$ TE and TM modes in an air-filled parallel-plate guide is known to be $f_{c1} = 7.5$ GHz. The guide is used at wavelength $\lambda = 1.5$ cm. Find the group velocity of the $m = 2$ TE and TM modes.

13.17 A parallel-plate guide is partially filled with two lossless dielectrics (Figure 13.25) where $\epsilon_{r1}' = 4.0$, $\epsilon_{r2}' = 2.1$, and $d = 1$ cm. At a certain frequency, it is found that the TM$_1$ mode propagates through the guide without suffering any reflective loss at the dielectric interface. (a) Find this frequency. (b) Is the guide operating at a single TM mode at the frequency found in part (a)? *Hint*: Remember Brewster's angle?

13.18 In the guide of Figure 13.25, it is found that $m = 1$ modes propagating from left to right totally reflect at the interface, so that no power is transmitted into the region of dielectric constant ϵ_{r2}'. (a) Determine the range of frequencies over which this will occur. (b) Does your part (a) answer in any way relate to the cutoff frequency for $m = 1$ modes in either region? *Hint*: Remember the critical angle?

13.19 A rectangular waveguide has dimensions $a = 6$ cm and $b = 4$ cm. (*a*) Over what range of frequencies will the guide operate single mode? (*b*) Over what frequency range will the guide support *both* TE_{10} and TE_{01} modes and no others?

13.20 Two rectangular waveguides are joined end-to-end. The guides have identical dimensions, where $a = 2b$. One guide is air-filled; the other is filled with a lossless dielectric characterized by ϵ_r'. (*a*) Determine the maximum allowable value of ϵ_r' such that single-mode operation can be simultaneously assured in *both* guides at some frequency. (*b*) Write an expression for the frequency range over which single-mode operation will occur in both guides; your answer should be in terms of ϵ_r', guide dimensions as needed, and other known constants.

13.21 An air-filled rectangular waveguide is to be constructed for single-mode operation at 15 GHz. Specify the guide dimensions, a and b, such that the design frequency is 10 percent higher than the cutoff frequency for the TE_{10} mode, while being 10 percent lower than the cutoff frequency for the next-higher-order mode.

13.22 Using the relation $\langle S \rangle = \frac{1}{2}\text{Re}\{\mathbf{E}_s \times \mathbf{H}_s^*\}$ and Eqs. (106) through (108), show that the average power density in the TE_{10} mode in a rectangular waveguide is given by

$$\langle S \rangle = \frac{\beta_{10}}{2\omega\mu} E_0^2 \sin^2(\kappa_{10}x)\mathbf{a}_z \text{ W/m}^2$$

13.23 Integrate the result of Problem 13.22 over the guide cross section, $0 < x < a$, $0 < y < b$, to show that the average power in watts transmitted down the guide is given as

$$P_{\text{av}} = \frac{\beta_{10}ab}{4\omega\mu} E_0^2 = \frac{ab}{4\eta} E_0^2 \sin\theta_{10} \text{ W}$$

where $\eta = \sqrt{\mu/\epsilon}$ and θ_{10} is the wave angle associated with the TE_{10} mode. Interpret.

13.24 Show that the group dispersion parameter, $d^2\beta/d\omega^2$, for a given mode in a parallel-plate or rectangular waveguide is given by

$$\frac{d^2\beta}{d\omega^2} = -\frac{n}{\omega c} \left(\frac{\omega_c}{\omega}\right)^2 \left[1 - \left(\frac{\omega_c}{\omega}\right)^2\right]^{-3/2}$$

where ω_c is the radian cutoff frequency for the mode in question [note that the first derivative form was already found, resulting in Eq. (57)].

13.25 Consider a transform-limited pulse of center frequency $f = 10$ GHz, and of full-width $2T = 1.0$ ns. The pulse propagates in a lossless single-mode rectangular guide which is air-filled and in which the 10 GHz operating frequency is 1.1 times the cutoff frequency of the TE_{10} mode. Using the result of Problem 13.24, determine the length of guide over which the
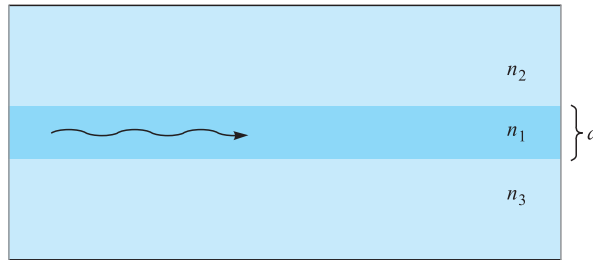
Figure 13.26 See Problem 13.29.

pulse broadens to twice its initial width. What simple step can be taken to reduce the amount of pulse broadening in this guide, while maintaining the same initial pulse width? Additional background for this problem is found in Section 12.6.

13.26 A symmetric dielectric slab waveguide has a slab thickness $d = 10\,\mu$m, with $n_1 = 1.48$ and $n_2 = 1.45$. If the operating wavelength is $\lambda = 1.3\,\mu$m, what modes will propagate?

13.27 A symmetric slab waveguide is known to support only a single pair of TE and TM modes at wavelength $\lambda = 1.55\,\mu$m. If the slab thickness is 5 μm, what is the maximum value of n_1 if $n_2 = 3.30$?

13.28 In a symmetric slab waveguide, $n_1 = 1.50$, $n_2 = 1.45$, and $d = 10\,\mu$m. (*a*) What is the phase velocity of the $m = 1$ TE or TM mode at cutoff? (*b*) How will your part (*a*) result change for higher-order modes (if at all)?

13.29 An *asymmetric* slab waveguide is shown in Figure 13.26. In this case, the regions above and below the slab have unequal refractive indices, where $n_1 > n_3 > n_2$. (*a*) Write, in terms of the appropriate indices, an expression for the minimum possible wave angle, θ_1, that a guided mode may have. (*b*) Write an expression for the maximum phase velocity a guided mode may have in this structure, using given or known parameters.

13.30 A step index optical fiber is known to be single mode at wavelengths $\lambda > 1.2\,\mu$m. Another fiber is to be fabricated from the same materials, but it is to be single mode at wavelengths $\lambda > 0.63\,\mu$m. By what percentage must the core radius of the new fiber differ from the old one, and should it be larger or smaller?

13.31 Is the mode field radius greater than or less than the fiber core radius in single-mode step index fiber?

13.32 The mode field radius of a step index fiber is measured as 4.5 μm at free-space wavelength $\lambda = 1.30\,\mu$m. If the cutoff wavelength is specified as $\lambda_c = 1.20\,\mu$m, find the expected mode field radius at $\lambda = 1.55\,\mu$m.

ELECTROMAGNETIC RADIATION AND ANTENNAS

We are used to the idea that loss mechanisms in electrical devices, including transmission lines and waveguides, are associated with resistive effects in which electrical power is transformed into heat. We have also assumed that time-varying electric and magnetic fields are totally confined to a waveguide or circuit. In fact, confinement is rarely complete, and electromagnetic power will *radiate* away from the device to some degree. Radiation may generally be an unwanted effect, as it represents an additional power loss mechanism, or a device may receive unwanted signals from the surrounding region. On the other hand, a well-designed antenna provides an efficient interface between guided waves and free-space waves for purposes of intentionally radiating or receiving electromagnetic power. In either case, it is important to understand the radiation phenomenon so that it can either be used most effectively or be reduced to a minimum. In this chapter, our goal is to establish such an understanding and to explore several practical examples of antenna design.■

14.1 BASIC RADIATION PRINCIPLES: THE HERTZIAN DIPOLE

The essential point of this chapter is that *any* time-varying current distribution will radiate electromagnetic power. So our first task is to find the fields that radiate from a specific time-varying source. This problem is different from any that we have explored. In our treatment of waves and fields in bulk media and in waveguides, only the wave motion in the medium was investigated, and the sources of the fields were not considered. Earlier in Chapter 11, we found the current distribution in a conductor by relating it to assumed electric and magnetic field intensities at the conductor
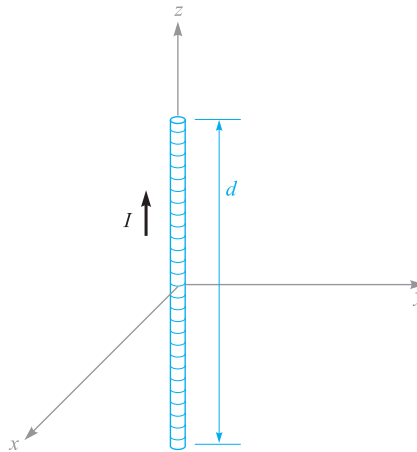
Figure 14.1   A differential current filament of length d carries a current $I = I_0 \cos \omega t$.

surface. Although this would relate the current source to the field, it is not practical for our purposes because the conductors were considered infinite in size in at least one dimension.

We begin by studying a current filament of infinitesimally small cross-section, positioned within an infinite lossless medium that is specified by permeability μ and permittivity ϵ (both real). The filament is specified as having a differential length, but we will later extend the results easily to larger dimensions that are on the order of a wavelength. The filament is positioned with its center at the origin and is oriented along the z axis as shown in Figure 14.1. The positive sense of the current is taken in the $\mathbf{a}_z$ direction. A uniform current $I(t) = I_0 \cos \omega t$ is assumed to flow in this short length d. The existence of such a current would imply the existence of time-varying charges of equal and opposite instantaneous amplitude on each end of the wire. For this reason, the wire is termed an *elemental* or *Hertzian dipole*. This is distinct in meaning from the more general definition of a dipole antenna that we will use later in this chapter.

The first step is the application of the retarded vector magnetic potential expression, as presented in Section 9.5,

$$\mathbf{A} = \int \frac{\mu \, I[t - R/v] \, d\mathbf{L}}{4\pi R} \tag{1}$$

where I is a function of the retarded time $t - R/v$.

When a single frequency is used to drive the antenna, v is the phase velocity of a wave at that frequency in the medium around the current element, and is given by

$v = 1/\sqrt{\mu\epsilon}$. Since no integration is required for the short filament, we have

$$\mathbf{A} = \frac{\mu \, I[t - R/v] \, d}{4\pi R} \mathbf{a}_z \tag{2}$$

Only the z component of $\mathbf{A}$ is present, for the current is only in the $\mathbf{a}_z$ direction. At any point P at distance R from the origin, the vector potential is retarded by R/v and so we use

$$I[t - R/v] = I_0 \cos\left[\omega\left(t - \frac{R}{v}\right)\right] = I_0 \cos\left[\omega t - kR\right] \tag{3}$$

where the wavenumber in the lossless medium is $k = \omega/v = \omega\sqrt{\mu\epsilon}$. In phasor form, Eq. (3) becomes

$$I_s = I_0 \, e^{-jkR} \tag{4}$$

where the current amplitude, I_0, is assumed to be real (as it will be throughout this chapter). Incorporating (4) into (2), we find the phasor retarded potential:

$$\mathbf{A}_s = A_{zs}\, \mathbf{a}_z = \frac{\mu I_0 d}{4\pi R} e^{-jkR} \, \mathbf{a}_z \tag{5}$$

Using a mixed coordinate system for the moment, we now replace R by the small r of the spherical coordinate system and then determine which spherical components are represented by A_{zs}. Using the projections as illustrated in Figure 14.2, we find

$$A_{rs} = A_{zs} \cos\theta \tag{6a}$$

$$A_{\theta s} = -A_{zs} \sin\theta \tag{6b}$$

and therefore

$$A_{rs} = \frac{\mu I_0 d}{4\pi r} \cos\theta \, e^{-jkr} \tag{7a}$$
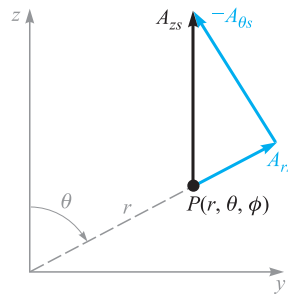
Figure 14.2 The resolution of A_{zs} at $P(r, \theta, \phi)$ into the two spherical components A_{rs} and $A_{\theta s}$. The sketch is arbitrarily drawn in the $\phi = 90°$ plane.

$$A_{\theta s} = -\frac{\mu I_0 d}{4\pi r} \sin\theta \, e^{-jkr} \tag{7b}$$

From these two components of the vector magnetic potential at P we can now find $\mathbf{B}_s$ or $\mathbf{H}_s$ from the definition of $\mathbf{A}_s$,

$$\mathbf{B}_s = \mu\mathbf{H}_s = \nabla \times \mathbf{A}_s \tag{8}$$

Taking the indicated partial derivatives as specified by the curl operator in spherical coordinates, we are able to separate Eq. (8) into its three spherical components, of which only the ϕ component is non-zero:

$$H_{\phi s} = \frac{1}{\mu r}\frac{\partial}{\partial r}(r A_{\theta s}) - \frac{1}{\mu r}\frac{\partial A_{rs}}{\partial\theta} \tag{9}$$

Now, substituting (7a) and (7b) into (9), we find the magnetic field:

$$H_{\phi s} = \frac{I_0 d}{4\pi} \sin\theta \, e^{-jkr}\left(j\frac{k}{r} + \frac{1}{r^2}\right) \tag{10}$$

The electric field that is associated with Eq. (10) is found from one of Maxwell's equations—specifically the point form of Ampere's circuital law as applied to the surrounding region (where conduction and convection current are absent). In phasor form, this is Eq. (23) in Chapter 11, except that in the present case we allow for a lossless medium having permittivity ϵ:

$$\nabla \times \mathbf{H}_s = j\omega\epsilon\mathbf{E}_s \tag{11}$$

Using (11), we expand the curl in spherical coordinates, assuming the existence of only a ϕ component for $\mathbf{H}_s$. The resulting electric field components are:

$$E_{rs} = \frac{1}{j\omega\epsilon}\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(H_{\phi s}\sin\theta) \tag{12a}$$

$$E_{\theta s} = \frac{1}{j\omega\epsilon}\left(-\frac{1}{r}\right)\frac{\partial}{\partial r}(r H_{\phi s}) \tag{12b}$$

Then on substituting (10) into (12a) and (12b) we find:

$$E_{rs} = \frac{I_0 d}{2\pi}\eta\cos\theta \, e^{-jkr}\left(\frac{1}{r^2} + \frac{1}{jkr^3}\right) \tag{13a}$$

$$E_{\theta s} = \frac{I_0 d}{4\pi}\eta\sin\theta \, e^{-jkr}\left(\frac{jk}{r} + \frac{1}{r^2} + \frac{1}{jkr^3}\right) \tag{13b}$$

where the intrinsic impedance is, as always, $\eta = \sqrt{\mu/\epsilon}$.

Equations (10), (13a), and (13b) are the fields that we set out to find. The next step is to interpret them. We first notice the e^{-jkr} factor appearing with each component. By itself, this term describes a spherical wave, propagating outward from the origin in the positive r direction with a phase constant $k = 2\pi/\lambda$. λ is the wavelength as

measured in the medium. Matters are complicated by the complex r-dependent terms in parentheses that appear in all three equations. These terms can be expressed in polar form (magnitude and phase), leading to the following modified versions of the three field equations for the Hertzian dipole:

$$H_{\phi s} = \frac{I_0 k d}{4\pi r} \left[1 + \frac{1}{(kr)^2} \right]^{1/2} \sin\theta \, \exp[-j(kr - \delta_\phi)] \tag{14}$$

$$E_{rs} = \frac{I_0 d}{2\pi r^2} \eta \left[1 + \frac{1}{(kr)^2} \right]^{1/2} \cos\theta \, \exp[-j(kr - \delta_r)] \tag{15}$$

$$E_{\theta s} = \frac{I_0 k d}{4\pi r} \eta \left[1 - \frac{1}{(kr)^2} + \frac{1}{(kr)^4} \right]^{1/2} \sin\theta \, \exp[-j(kr - \delta_\theta)] \tag{16}$$

where the additional phase terms are

$$\delta_\phi = \tan^{-1} [\, kr \,] \tag{17a}$$

$$\delta_r = \tan^{-1} [\, kr \,] - \frac{\pi}{2} \tag{17b}$$

and

$$\delta_\theta = \tan^{-1} \left[kr \left(1 - \frac{1}{(kr)^2} \right) \right] \tag{18}$$

In (17) and (18), the principal value is always taken when evaluating the inverse tangent. This means that the phases as expressed in (17) and (18) will occur within the range $\pm\pi/2$ as kr varies between zero and infinity. Suppose a single frequency (k value) is chosen, and the fields are observed at a fixed instant in time. Consider observing the field along a path in the direction of increasing r, in which spatial oscillations will be seen as r varies. As a result of the phase terms in (17) and (18), the oscillation period will change with increasing r. We may demonstrate this by considering the H_ϕ component as a function of r under the following conditions:

$$I_0 d = 4\pi \qquad \theta = 90° \qquad t = 0$$

Using $k = 2\pi/\lambda$, Eq. (14) becomes

$$H_{\phi s} = \frac{2\pi}{\lambda r} \left[1 + \left(\frac{\lambda}{2\pi r} \right)^2 \right]^{1/2} \exp\left\{ -j \left[\frac{2\pi r}{\lambda} - \tan^{-1} \left(\frac{2\pi r}{\lambda} \right) \right] \right\} \tag{19}$$

The real part of (19) gives the real instantaneous field at $t = 0$:

$$\mathcal{H}_\phi(r, 0) = \frac{2\pi}{\lambda r} \left[1 + \left(\frac{\lambda}{2\pi r} \right)^2 \right]^{1/2} \cos\left[\tan^{-1} \left(\frac{2\pi r}{\lambda} \right) - \frac{2\pi r}{\lambda} \right] \tag{20}$$

We next use the identity, $\cos(a - b) = \cos a \cos b + \sin a \sin b$, in addition to $\cos(\tan^{-1} x) = 1/\sqrt{1 + x^2}$ and $\sin(\tan^{-1} x) = x/\sqrt{1 + x^2}$. With these, Eq. (20)

simplifies to

$$\mathcal{H}_\phi = \frac{1}{r^2}\left[\cos\left(\frac{2\pi r}{\lambda}\right) + \frac{2\pi r}{\lambda}\sin\left(\frac{2\pi r}{\lambda}\right)\right] \tag{21}$$

A few important points emerge when studying Eq. (21). First, at distances r that are on the order of a wavelength, the expression consists of two sinusoidal functions having the same period but in which the amplitude of the second one increases with increasing r. This leads to significant nonsinusoidal behavior, in that the field as a function of r/λ will oscillate, but with nonuniform periodicity, and with positive and negative amplitudes that differ in each cycle. Second, at distances r that are much greater than a wavelength, the second term in (21) dominates, and the field variation with r approaches that of a pure sinusoid. We may therefore say that, for all practical purposes, the wave at large distances, where $r \gg \lambda$, is a uniform plane wave having a sinusoidal variation with distance (and time, of course) and a well-defined wavelength. This wave evidently carries power away from the differential antenna.

We should now take a more careful look at the expressions containing terms varying as $1/r^3$, $1/r^2$, and $1/r$ in Eqs. (10), (13a), and (13b). At points very close to the current element, the $1/r^3$ term must be dominant. In the numerical example we have used, the relative values of the terms in $1/r^3$, $1/r^2$, and $1/r$ in the $E_{\theta s}$ expression are about 250, 16, and 1, respectively, when r is 1 cm. The variation of an electric field as $1/r^3$ should remind us of the *electrostatic* field of the dipole (Chapter 4). The development of this concept is the subject of Problem 14.4. The near-field terms represent energy stored in a reactive (capacitive) field, and they do not contribute to the radiated power. The inverse-square term in the $H_{\phi s}$ expression is similarly important only in the region very near to the current element. It corresponds to the *induction* field of the dc element, as found through the Biot-Savart law (Problem 14.5).

At distances corresponding to, say, 10 or more wavelengths from the current element, the product $kr = 2\pi r/\lambda > 20\pi$, and the fields dramatically simplify. In Eqs. (14)–(16), the terms within brackets involving $1/(kr)^2$ and $1/(kr)^4$ can be considered much less than unity, and can be neglected. In addition, the phases (Eqs. (17) and (18)) all approach $\pi/2$. The effect is also seen in Eqs. (10), (13a), and (13b), in which all terms except the inverse-distance $(1/r)$ term may be neglected. At such distances, at which $kr \gg 1$ (equivalently $r \gg \lambda$), we are said to be in the *far-field* or *far-zone*. The remaining field terms that have the $1/r$ dependence are the *radiation* fields. This produces an approximately zero E_{rs} field, leaving only $E_{\theta s}$ and $H_{\phi s}$. Thus, in the far zone:

$$E_{rs} \doteq 0$$

$$E_{\theta s} = j\frac{I_0 kd}{4\pi r}\eta \sin\theta\, e^{-jkr} \tag{22}$$

$$H_{\phi s} = j\frac{I_0 kd}{4\pi r}\sin\theta\, e^{-jkr} \tag{23}$$
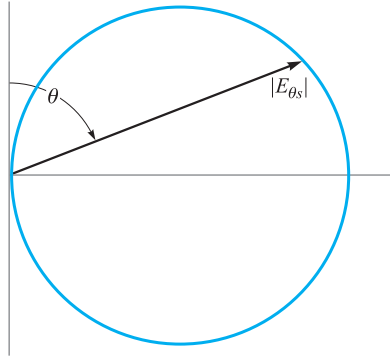
Figure 14.3 The polar plot of the
E-plane pattern of a vertical current
element. The crest amplitude of $E_{\theta s}$ is
plotted as a function of the polar angle
θ at a constant distance r. The locus is
a circle.

The relation between these fields is evidently the same as that of a uniform plane wave, which an expanding spherical wave approximates at large radii, and over regions in which $1/r$ is approximately constant. Specifically,

$$E_{\theta s} = \eta H_{\phi s} \qquad (kr \gg 1 \text{ or } r \gg \lambda) \qquad (24)$$

The variation of both radiation fields with the polar angle θ is the same; the fields maximize in the equatorial plane (xy plane) of the current element and vanish off the ends of the element. The variation with angle may be shown by plotting a vertical, or *E-plane* pattern (assuming a vertical orientation of the current element). The E plane is simply the coordinate plane that contains the electric field, which in our present case, is any surface of constant ϕ in the spherical coordinate system. Figure 14.3 shows an E-plane plot of Eq. (22) in polar coordinates, in which the relative magnitude of $E_{\theta s}$ is plotted against θ for a constant r. The length of the vector shown in the figure represents the magnitude of E_θ, normalized to unity at $\theta = 90°$; the vector length is just $|\sin \theta|$, and so as θ varies, the tip of the vector traces out a circle as shown.

A horizontal, or *H-plane* pattern may also be plotted for this or more complicated antenna systems. In the present case, this would show the variation of field intensity with ϕ. The H-plane of the current element (the plane that contains the magnetic field) is any plane that is normal to the z axis. As E_θ is not a function of ϕ, the H-plane plot would be simply a circle centered at the origin.

D14.1. A short antenna with a uniform current distribution in air has $I_0 d = 3 \times 10^{-4}$ A $\cdot$ m and $\lambda = 10$ cm. Find $|E_{\theta s}|$ at $\theta = 90°$, $\phi = 0°$, and $r =: (a)$ 1 cm; (b) 2 cm; (c) 20 cm; (d) 200 cm; (e) 2 m.

Ans. 125 V/m; 25 V/m; 2.8 V/m; 0.28 V/m; 0.028 V/m

14.2 ANTENNA SPECIFICATIONS

It is important to fully describe and quantify the radiation from a general antenna. To do this, we need to be aware of a few new concepts and definitions.

In order to evaluate the radiated power, the time-average Poynting vector must be found (Eq. (77), Chapter 11). In the present case, this will become

$$< \mathbf{S} > = \frac{1}{2} \mathcal{R}e \{E_{\theta s} H_{\phi s}^*\} \mathbf{a}_r \quad \text{W/m}^2 \tag{25}$$

Substituting (22) and (23) into (25), we obtain the time-average Poynting vector magnitude:

$$| < \mathbf{S} > | = S_r = \frac{1}{2} \left(\frac{I_0 k d}{4 \pi r} \right)^2 \eta \sin^2 \theta \tag{26}$$

From this we find the time-average power that crosses the surface of a sphere of radius r, centered at the antenna:

$$P_r = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} S_r \, r^2 \sin\theta \, d\theta \, d\phi = 2\pi \left(\frac{1}{2} \right) \left(\frac{I_0 k d}{4\pi} \right)^2 \eta \int_0^{\pi} \sin^3 \theta \, d\theta \tag{27}$$

The integral is evaluated, and we substitute $k = 2\pi/\lambda$. We will also assume that the medium is free space, where $\eta = \eta_0 \doteq 120\pi$. We finally obtain:

$$P_r = 40\pi^2 \left(\frac{I_0 d}{\lambda} \right)^2 \quad \text{W} \tag{28}$$

This is the same average power that would be dissipated in a resistance R_{rad} by sinusoidal current of amplitude I_0 in the absence of any radiation, where

$$P_r = \frac{1}{2} I_0^2 R_{\text{rad}} \tag{29}$$

We call this effective resistance the *radiation resistance* of the antenna. For the differential antenna, this becomes

$$R_{\text{rad}} = \frac{2 P_r}{I_0^2} = 80\pi^2 \left(\frac{d}{\lambda} \right)^2 \tag{30}$$

If, for example, the differential length is 0.01λ, then R_{rad} is about $0.08 \, \Omega$. This small resistance is probably comparable to the *ohmic* resistance of a practical antenna (providing a measure of the power dissipated through heat), and thus the efficiency of the antenna is likely to be too low. Effective matching to the source also becomes very difficult to achieve, for the input reactance of an electrically short antenna is much greater in magnitude than the input resistance R_{rad}.

Evaluating the net power from the antenna, as carried out in (27), involved the integration of the Poynting vector over a spherical shell of presumed large radius, such that the antenna appeared as a point source at the sphere center. In view of this,

a new concept of power density can be introduced; this involves the power that is carried within a very thin cone with its vertex at the antenna location. The axis of the cone extends along a line of radius, and thus the cone intersects the spherical surface over which the integral in (27) is taken. That portion of the sphere area that the cone intersects will have area A. We define the *solid angle* of the cone in the following manner: If $A = r^2$, where r is the sphere radius, then the cone is defined as having a solid angle, Ω, equal to one *steradian* (sr).[1] As the total sphere area is $4\pi r^2$, we see that the total solid angle contained within a sphere is 4π steradians.

As a consequence of this definition, differential area on the sphere surface can be expressed in terms of a differential solid angle through:

$$dA = r^2\, d\Omega \tag{31}$$

The total sphere area can then be expressed as an integral over solid angle, or equivalently by an integral using spherical coordinates:

$$A_{net} = 4\pi r^2 = \int_0^{4\pi} r^2 d\Omega = \int_0^{2\pi}\int_0^\pi r^2 \sin\theta\, d\theta\, d\phi \tag{32}$$

from which we identify the differential solid angle as expressed in spherical coordinates:

$$\boxed{d\Omega = \sin\theta\, d\theta\, d\phi} \tag{33}$$

D14.2. A cone is centered on the positive z axis with its vertex at the origin. The cone angle in spherical coordinates is θ_1. (*a*) If the cone subtends 1 sr of solid angle, determine θ_1; (*b*) If $\theta_1 = 45°$, find the solid angle subtended.

Ans. $32.8°$; $\pi\sqrt{2}$

We can now express the Poynting vector magnitude as found in Eq. (26) in units of power per unit solid angle. To do this, we multiply the Watts/m^2 power density in (26) by the sphere area encompassed in one steradian—which is r^2. The result, known as the *radiation intensity*, is

$$\boxed{K(\theta,\phi) = r^2 S_r \quad \text{W/Sr}} \tag{34}$$

For the Hertzian dipole, the intensity is independent of ϕ, and we would have (using (26)):

$$K(\theta) = \frac{1}{2}\left(\frac{I_0 k d}{4\pi}\right)^2 \eta \sin^2\theta \quad \text{W/Sr} \tag{35}$$

---

[1] This definition is related to that of the radian, in which the arc length on a circle traced out by a change of angle of one radian is the circle radius, r.
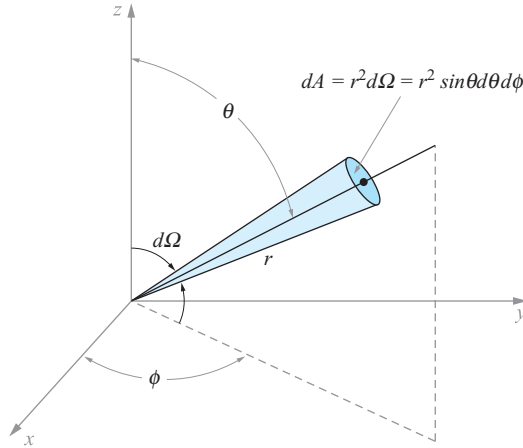
Figure 14.4   A cone having differential solid angle, $d\Omega$, subtends a (shaded) differential area on the surface of a sphere of radius r. This area, given by $dA = r^2 d\Omega$, can also be expressed in our more familiar spherical coordinate system as $dA = r^2 \sin\theta d\theta d\phi$.

In the general case, the total radiated power is then

$$P_r = \int_0^{4\pi} K d\Omega = \int_0^{2\pi}\int_0^{\pi} K(\theta,\phi)\sin\theta d\theta d\phi \quad \text{W} \tag{36}$$

which, for the Hertzian dipole, gives the same result as found in (28).

The advantage of using the radiation intensity for power density is that this quantity is independent of the radius. This is true, however, only if the original power density exhibits a $1/r^2$ dependence. In fact, *all* antennas have this functional dependence on radius in the far zone, in that when far enough away, the antenna appears as a point source of power. Assuming the surrounding medium does not absorb any power, the integral of the Poynting vector over a closed sphere of *any* radius must give the same result. This fact demands an inverse-square dependence on radius for the power density. With the radial dependence removed, one can concentrate on the angular dependence of the power density as expressed by K, and this will differ significantly among different antennas.

A special case of a power source is an *isotropic radiator*, defined as having a *constant* radiation intensity (i.e., $K = K_{iso}$ is independent of θ and ϕ). This gives a simple relation between K and the total radiated power:

$$P_r = \int_0^{4\pi} K_{iso}\, d\Omega = 4\pi K_{iso} \quad \Rightarrow \quad K_{iso} = P_r/4\pi \quad \text{(isotropic radiator)} \tag{37}$$

Generally, K *will* vary with angle, giving more intensity in some directions than in others. It is useful to compare the radiation intensity in a given direction to that

which would occur if the antenna were to radiate *the same net power* isotropically. The *directivity* function, $D(\theta, \phi)$, does this.[2] Using (36) and (37), we can write the directivity:

$$D(\theta, \phi) = \frac{K(\theta, \phi)}{K_{iso}} = \frac{K(\theta, \phi)}{P_r/4\pi} = \frac{4\pi K(\theta, \phi)}{\oint K \, d\Omega} \tag{38}$$

Of particular interest in most cases is the maximum value of the directivity, D_{max}, which is sometimes called simply D (without the θ and ϕ dependence indicated):

$$D = D_{max} = \frac{4\pi K_{max}}{\oint K \, d\Omega} \tag{39}$$

in which the maximum radiation intensity, K_{max}, will usually occur at more than one set of values of θ and ϕ. Typically, the directivity is quoted in decibels, according to the definition:

$$D_{dB} = 10 \log_{10} (D_{max}) \quad \text{dB} \tag{40}$$

EXAMPLE 14.1

Evaluate the directivity of the Hertzian dipole.

Solution. Use Eqs. (35) and (28), with $k = 2\pi/\lambda$ and $\eta = \eta_0 = 120\pi$ in the expression:

$$D(\theta, \phi) = \frac{4\pi K(\theta, \phi)}{P_r} = \frac{2\pi \left(\frac{I_0 d}{2\lambda}\right)^2 120\pi \sin^2 \theta}{40\pi^2 \left(\frac{I_0 d}{\lambda}\right)^2} = \frac{3}{2} \sin^2 \theta$$

The maximum of this result, occurring at $\theta = \pi/2$, is:

$$D_{max} = \frac{3}{2} \quad \text{Or, in decibels:} \quad D_{dB} = 10 \log_{10} \left(\frac{3}{2}\right) = 1.76 \text{ dB}$$

D14.3. What is the directivity in dB of a power source at the origin that radiates: a) uniformly into the upper half-space, but nothing into the lower half-space, b) into all space with a $\cos^2 \theta$ power density dependence, c) into all space with a $|\cos^n \theta|$ dependence?

Ans. 3; 4.77; $10 \log_{10}(n + 1)$

Usually, one would like to have a much higher directivity than what we just found for the Hertzian dipole. One implication of a low directivity (and a problem

---

[2] In earlier times (and in older texts), the directivity function was called the *directive gain*. The latter term has since been discarded by the Antenna Standards Committee of the IEEE Antennas and Propagation Society, in favor of the term "directivity." Details are found in IEEE Std 145-1993.

with the short antenna) is that power is radiated over a broad angular range in the
E plane. In most cases, it is desired to confine the power to a narrow range, or
small *beamwidth*, thus increasing the directivity. The *3-dB beamwidth* is defined
as the separation between the two angles at which the directivity falls to one-half
its maximum value. For the Hertzian dipole, and using the $D(\theta, \phi)$ result from the
previous example, the beamwidth will be the span between the two θ values on either
side of $90°$ at which $\sin^2 \theta = 1/2$, or $|\sin \theta| = 1/\sqrt{2} = 0.707$. These two values
are $45°$ and $135°$, representing a 3-dB beamwidth of $135° - 45° = 90°$. We will
see that using a longer antenna leads to both a narrower beamwidth and an increased
radiation resistance. In the H plane, radiation is uniform at all values of ϕ, no matter
what length is used. It is necessary to use multiple antennas in an *array* in order to
narrow the beam in the H-plane.

We have based several definitions on the total average power that is radiated by
the antenna, P_r. It is desirable, however, to distinguish the radiated power from the
input power that is *supplied* to the antenna, P_{in}. It is likely that P_{in} will be somewhat
greater than P_r because of resistive losses in the conducting materials that make up
the antenna. To overcome this resistance, a greater input voltage amplitude would be
necessary to generate a given current amplitude, I_0, on which all of our power compu-
tations have been based. The antenna *gain* is defined in such a way to accommodate
this difference.[3]

Specifically, suppose that the antenna in question were to isotropically radiate
all of the electrical power that is supplied to it, which is P_{in}. The radiation intensity
would simply be $K_s = P_{in}/4\pi$. Gain is defined as the ratio of the actual radiation
intensity in a specified direction, to K_s:

$$G(\theta, \phi) = \frac{K(\theta, \phi)}{K_s} = \frac{4\pi K(\theta, \phi)}{P_{in}} \tag{41}$$

Note that the term $4\pi K(\theta, \phi)$ would be the radiated power of an isotropic antenna
whose (in that case, constant) radiation intensity is $K(\theta, \phi)$. The gain thus expresses
the ratio of the radiated power of an antenna to the input power *as if the antenna were
to radiate isotropically* with constant K as evaluated at a selected θ and ϕ. Using
(38), we can relate directivity to gain:

$$D(\theta, \phi) = \frac{4\pi K(\theta, \phi)}{P_r} = \frac{P_{in}}{P_r} G(\theta, \phi) = \frac{1}{\eta_r} G(\theta, \phi) \tag{42}$$

where η_r is the *radiation efficiency* of the antenna, defined as the ratio of the radiated
power to the input power. Other ways of writing this are:

$$\eta_r = \frac{P_r}{P_{in}} = \frac{G(\theta, \phi)}{D(\theta, \phi)} = \frac{G_{max}}{D_{max}} \tag{43}$$

which expresses η_r as the maximum gain divided by the maximum directivity.

---

[3] The antenna gain defined in this way is sometimes called the power gain.

14.3 MAGNETIC DIPOLE

An interesting device that is closely related to the Hertzian dipole is the *magnetic dipole* antenna. Shown in Figure 14.5, the antenna consists of a circular current loop of radius a, centered at the origin, and in the xy plane. The loop current is sinusoidal and is given by $I(t) = I_0 \cos \omega t$, as was the case in the Hertzian dipole. Although it is possible to work out the fields for this antenna, beginning with the retarded potentials as in the previous section, there is a much faster way.

We first note that the circulating current implies the existence of a circulating electric field that overlaps the wire and that has the same time dependence. So one could simply replace the wire by a circular electric field that we could designate as $\mathbf{E}(a, t) = E_0(a) \cos(\omega t) \mathbf{a}_\phi$. Such a change would replace conduction current with displacement current, which will have no effect on the surrounding field solutions for $\mathbf{E}$ and $\mathbf{H}$. Next, suppose that we could replace the electric field by a magnetic field, again of the form $\mathbf{H}(a, t) = H_0 \cos(\omega t) \mathbf{a}_\phi$. This is the magnetic field that would be generated by the Hertzian dipole at radius a in the xy plane, and it enables us to obtain the solution for the current loop field through the following method:

We begin with Maxwell's equations in a sourceless medium ($\rho_v = \mathbf{J} = 0$):

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} \tag{44a}$$

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \tag{44b}$$

$$\nabla \cdot \mathbf{E} = 0 \tag{44c}$$

$$\nabla \cdot \mathbf{H} = 0 \tag{44d}$$
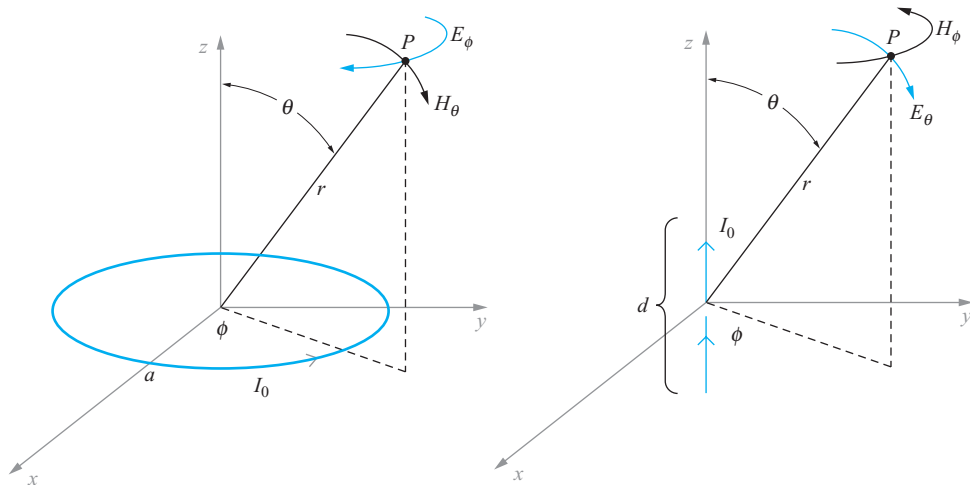
Figure 14.5 Magnetic (left) and electric dipole antennas are dual structures, producing identical field patterns but with the roles of **E** and **H** interchanged.

By inspection, we see that the equations would be unchanged if we replace $\mathbf{E}$ with $\mathbf{H}$, $\mathbf{H}$ with $-\mathbf{E}$, ϵ with μ, and μ with ϵ. This illustrates the concept of *duality* in electromagnetics. The fact that the current loop *electric* field will have the same functional form as the electric dipole *magnetic* field means that with the above substitutions, we can construct the current loop fields directly from the electric dipole results. It is because of this duality between field solutions of the two devices that the name, *magnetic dipole antenna*, is applied to the current loop device.

Before making the substitutions, we must relate the currents and geometries of the two devices. To do this, consider first the static electric dipole result of Chapter 4 (Eq. (35)). We can specialize this result by finding the electric field on the z axis ($\theta = 0$). We find

$$\mathbf{E}|_{\theta=0} = \frac{Qd}{2\pi\epsilon z^3}\,\mathbf{a}_z \tag{45}$$

We can next study the current loop magnetic field as found on the z axis, in which a steady current I_0 is present. This result can be obtained using the Biot–Savart Law:

$$\mathbf{H}|_{\theta=0} = \frac{\pi a^2 I_0}{2\pi z^3}\,\mathbf{a}_z \tag{46}$$

Now the current associated with a harmonically time-varying charge on the electric dipole, $Q(t)$, is

$$\boxed{I_0 = \frac{dQ}{dt} = j\omega Q \;\Rightarrow\; Q = \frac{I_0}{j\omega}} \tag{47}$$

If we substitute Eq. (47) into Eq. (45), and replace d with $j\omega\epsilon(\pi a^2)$, we find that Eq. (45) is transformed to Eq. (46). We now perform these substitutions, along with the replacements, $\mathbf{H}$ for $\mathbf{E}$, $-\mathbf{E}$ for $\mathbf{H}$, ϵ for μ, and μ for ϵ, on Eqs. (14), (15), and (16). The results are

$$E_{\phi s} = -j\frac{\omega\mu(\pi a^2)I_0 k}{4\pi r}\left[1 + \frac{1}{(kr)^2}\right]^{1/2}\sin\theta\,\exp[-j(kr - \delta_\phi)] \tag{48}$$

$$H_{rs} = j\frac{\omega\mu(\pi a^2)I_0}{2\pi r^2}\frac{1}{\eta}\left[1 + \frac{1}{(kr)^2}\right]^{1/2}\cos\theta\,\exp[-j(kr - \delta_r)] \tag{49}$$

$$H_{\theta s} = j\frac{\omega\mu(\pi a^2)I_0 k}{4\pi r}\frac{1}{\eta}\left[1 - \frac{1}{(kr)^2} + \frac{1}{(kr)^4}\right]^{1/2}\sin\theta\,\exp[-j(kr - \delta_\theta)] \tag{50}$$

where δ_r, δ_θ, and δ_ϕ are as defined in Eqs. (17) and (18). In the far-field ($kr \gg 1$), $E_{\phi s}$ and $H_{\theta s}$ survive, and these simplify to compare closely with (22) and (23). This process of exploiting duality in electromagnetics is a very powerful method that can be applied in many situations.

14.4 THIN WIRE ANTENNAS

In addition to giving insights on radiation fundamentals, the Hertzian dipole results provide us with a basis from which the fields associated with more complicated antennas can be derived. In this section this methodology is applied to the more practical problem of straight thin wire antennas of any length. We will find that for a given wavelength, changes in antenna length lead to dramatic variations in (and control of) the radiation pattern. We will also note improvement in directivity and efficiency when using certain antenna lengths.

The basic arrangement is shown in Figure 14.6. In a simplistic way, it is possible to think of the antenna as having been formed by bending the two wires of an open-ended transmission line down and up by 90°. The midpoint, at which the bends occur, is known as the *feed* point. The current, originally present, persists and is instantaneously flowing in the same direction in the lower and upper sections of the antenna. If the current is sinusoidal, a standing wave is set up in the antenna wires, with zeros occurring at the wire ends at $z = \pm\ell$. A symmetric antenna of this type is called a *dipole*.

The actual current distribution on a very thin wire antenna is very nearly sinusoidal. With zero current at the ends, maxima occur one-quarter wavelength from each end, and the current continues to vary in this manner toward the feed point. The current at the feed will be small for an antenna whose overall length, 2ℓ, is an integral number of wavelengths; but it will be equal to the maximum found at any point on the antenna if the antenna length is an odd multiple of a half wavelength.
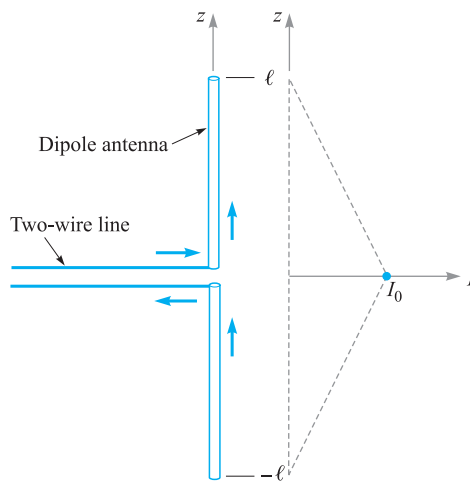
Figure 14.6 A thin dipole antenna driven sinusoidally by a two-wire line. The current amplitude distribution, shown in the adjacent plot, is approximately linear if the overall length is sufficiently less than a half-wavelength. Current amplitude maximizes at the center (feed) point.

On a short antenna, where 2ℓ is significantly less than a half-wavelength, we see only the first portion of the sine wave; the amplitude of the current increases in an approximately linear manner, from zero at the ends to a maximum value at the feed, as indicated in Figure 14.6. The gap at the feed point is small and has negligible effects. The short antenna approximation (in which a linear current variation along the length can be assumed) is reasonable for antennas having an overall length that is less than about one-tenth of a wavelength.

A simple extension of the Hertzian dipole results can be performed in the short antenna regime ($\ell < \lambda/20$). If this is the case, then retardation effects may be neglected. That is, signals arriving at any field point P from the two ends of the antenna are approximately in phase. The average current along the antenna is $I_0/2$, where I_0 is the input current at the feed. The electric and magnetic field intensities will thus be one-half the values given in (22) and (23), and there are no changes in the vertical and horizontal patterns. The power will be one-quarter of its previous value, and thus the radiation resistance will also be one-quarter of the value given by (30). Matters improve as the antenna length is increased, but retardation effects must then be included.

For longer lengths, the current distribution is treated the same as that of an open-ended transmission line that propagates a TEM wave. This will be a standing wave in which the current phasor is given by

$$I_s(z) \doteq I_0 \sin(kz) \tag{51}$$

where the open end is located at $z = 0$. Also, for a TEM wave on a transmission line, the phase constant will be $\beta = k = \omega\sqrt{\mu\epsilon}$. When the line is unfolded to form the antenna, the z axis is rotated to the vertical orientation with $z = 0$ occurring at the feed point. The current in (51) is then modified to be

$$I_s(z) \doteq \left.\begin{cases} I_0 \sin k(\ell - z) & (z > 0) \\ I_0 \sin k(\ell + z) & (z < 0) \end{cases}\right\} = I_0 \sin k(\ell - |z|) \tag{52}$$

The strategy from here is to consider the antenna as made up of a stack of Hertzian dipoles, each having length dz (Figure 14.7). The current amplitude in each Hertzian is determined according to its position z along the length, and is given by (52). The far-zone field from each Hertzian can then be written using Eq. (22), suitably modified. We write this as a *differential* field contribution at a far point at distance r' and spherical coordinate angle, θ':

$$dE_{\theta s} = j\frac{I_s(z)\,k\,dz}{4\pi r'}\,\eta\sin\theta'\,e^{-jkr'} \tag{53}$$

The coordinates r' and θ' are, of course, referenced from the center of the Hertzian, which itself is at a position z along the antenna length. We need to reference these local coordinates to the origin, which occurs at the antenna feed point. To do this, we borrow from the methods used to analyze the static electric dipole as presented in Sec. 4.7. Referring to Fig. 14.7, we can write the relation between
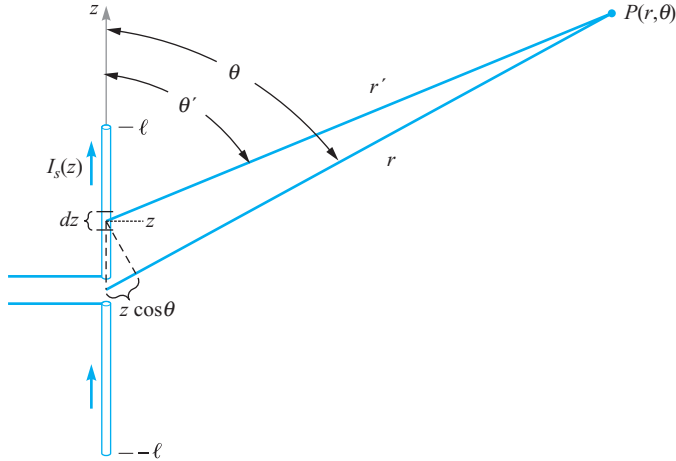
Figure 14.7   A dipole antenna can be represented as a stack of Hertzian dipoles whose individual phasor currents are given by $I_s(z)$. One Hertzian dipole is shown at location z, and has length dz. When the observation point, P, lies in the far zone, distance lines r and r' are approximately parallel, so they differ in length by $z \cos \theta$.

the distance r' from the Hertzian at location z, and the distance r from the origin to the same point as

$$r' \doteq r - z \cos \theta \tag{54}$$

where, in the far field, $\theta' \doteq \theta$, and distance lines r' and r are approximately parallel. Eq. (53) is then modified to read

$$dE_{\theta s} = j \frac{I_s(z)\, k\, dz}{4\pi r} \, \eta \sin \theta \, e^{-jk(r - z \cos \theta)} \tag{55}$$

Notice that in obtaining (55) from (53) we have approximated $r' \doteq r$ in the denominator, as the use of Eq. (54) will make little difference when considering *amplitude* variations with z and θ. The exponential term in (55) *does* include (54) because slight variations in z or θ will greatly impact the phase.

Now, the total electric field at the far-zone position (r, θ) will be the sum of all the Hertzian dipole contributions along the antenna length, which becomes the integral:

$$
\begin{aligned}
E_{\theta s}(r, \theta) = \int dE_{\theta s} &= \int_{-\ell}^{\ell} j \frac{I_s(z)\, k\, dz}{4\pi r} \, \eta \sin \theta \, e^{-jk(r - z \cos \theta)} \\
&= \left[j \frac{I_0 k}{4\pi r} \, \eta \, \sin \theta \, e^{-jkr} \right] \int_{-\ell}^{\ell} \sin k(\ell - |z|) \, e^{jkz \cos \theta} \, dz
\end{aligned}
\tag{56}
$$

To evaluate the last integral, we first express the complex exponential in terms of sine and cosine terms using the Euler identity. Denoting the bracketed terms outside the

integral as A, we write:

$$E_{\theta s}(r, \theta) = A \int_{-\ell}^{\ell} \underbrace{\sin k(\ell - |z|)}_{even} \underbrace{\cos(kz \cos \theta)}_{even} + j \underbrace{\sin k(\ell - |z|)}_{even} \underbrace{\sin(kz \cos \theta)}_{odd} \, dz$$

in which the even or odd parity of each term is indicated. The imaginary part of the integrand, consisting of the product of even and odd functions, yields a term with net odd parity; it thus integrates to zero over the symmetric limits of $-\ell$ to ℓ. This leaves the real part, whose integral can be expressed over the positive z range and then further simplified using trigonometric identities:

$$E_{\theta s}(r, \theta) = 2A \int_0^{\ell} \sin k(\ell - z) \cos(kz \cos \theta) \, dz$$

$$= A \int_0^{\ell} \sin[k(\ell - z) + kz \cos \theta] + \sin[k(\ell - z) - kz \cos \theta] \, dz$$

$$= A \int_0^{\ell} \sin[kz(\cos \theta - 1) + k\ell] - \sin[kz(\cos \theta + 1) - k\ell] \, dz$$

The last integral is straightforward and evaluates as

$$E_{\theta s}(r, \theta) = 2A \left[\frac{\cos(k\ell \cos \theta) - \cos(k\ell)}{k \sin^2 \theta} \right]$$

Now, reincorporating the expression for A gives the final result:

$$E_{\theta s}(r, \theta) = j \frac{I_0 \eta}{2\pi r} e^{-jkr} \left[\frac{\cos(k\ell \cos \theta) - \cos(k\ell)}{\sin \theta} \right] = E_0 \, F(\theta) \left[\frac{e^{-jkr}}{r} \right]$$

(57)

where we identify the field amplitude

$$E_0 = j \frac{I_0 \eta}{2\pi}$$

(58)

and where the terms involving θ and ℓ are isolated to form the E-plane *pattern function* for the dipole antenna:

$$F(\theta) = \left[\frac{\cos(k\ell \cos \theta) - \cos(k\ell)}{\sin \theta} \right]$$

(59)

This important function, when normalized, is the E-plane pattern of the dipole antenna. It explicitly shows how choices in dipole length affect the θ dependence in the pattern, and it ultimately determines the dependence on ℓ of the directive gain, directivity, and radiated power for a given current.

Plots of the magnitude of $F(\theta)$ in the E-plane are shown in Figure 14.8a and b for selected dipole lengths. In these, the xz plane is chosen, although the results will be the same in any plane that contains the z axis. The plots show a trend toward narrower radiation beams as length increases, but to the point at which secondary maxima, or *sidelobes*, develop for overall antenna lengths (2ℓ) that exceed one wavelength.
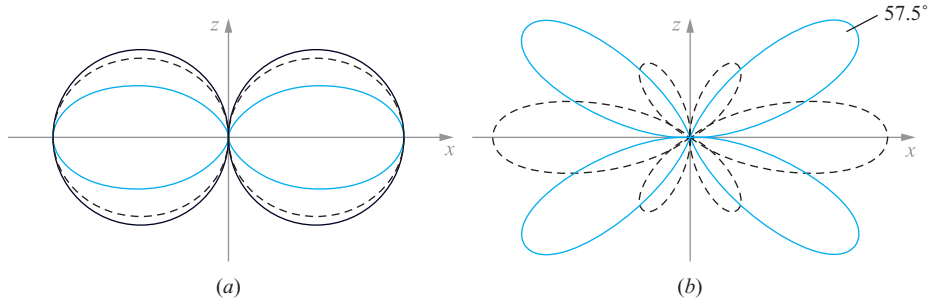
Figure 14.8 *E*-plane plots, normalized to maxima of 1.0, found from $F(\theta)$ for dipole antennas having overall lengths, 2ℓ, of (a) $\lambda/16$ (solid black), $\lambda/2$ (dashed), and λ (red), and (b) 1.3λ (dashed), and 2λ (red). In (a), the beam-narrowing trend is evident as length increases (or as wavelength decreases). Note that the $\lambda/16$ curves are nearly circular and thus approximate the Hertzian dipole pattern. At lengths that exceed one wavelength, sidelobes begin to develop, as exhibited in the smaller beams in the 1.3λ pattern in (b). As length increases, the sidelobes grow to form the four symmetrically arranged main lobes of the 2λ antenna, where the lobe in the first quadrant maximizes at $\theta = 57.5°$. The main lobes along x that were present in the 1.3λ antenna diminish with increasing length, and have vanished completely when the length reaches 2λ.

The presence of sidelobes is usually not wanted, mainly because they represent radiated power in directions other than that of the main beam ($\theta = \pi/2$). Sidelobe power will therefore likely miss the intended receiver. In addition, the sidelobe directions change with wavelength, and will therefore impart an angular spread to a radiated signal, to an extent which will of course increase with increasing signal bandwidth. These problems are avoided by using antenna lengths that are less than one wavelength.

The radiation intensity can now be found for the dipole antenna by using Eq. (34), along with (25):

$$K(\theta) = r^2 S_r = \frac{1}{2}\mathcal{R}e\left\{E_{\theta s} H_{\phi s}^*\right\}r^2$$

where $H_{\phi s} = E_{\theta s}/\eta$. Substituting (57), we obtain

$$\boxed{K(\theta) = \frac{\eta I_0^2}{8\pi^2}[F(\theta)]^2 = \frac{15\,I_0^2}{\pi}[F(\theta)]^2 \qquad \text{W/Sr}} \tag{60}$$

where in the last equality free space is assumed, in which $\eta = \eta_0 = 120\pi$. The total radiated power is now the integral of the radiation intensity over all solid angles, or

$$P_r = \int_0^{4\pi} K\,d\Omega = \int_0^{2\pi}\int_0^{\pi} K(\theta)\sin\theta\,d\theta\,d\phi \tag{61}$$

Again assuming free space we find

$$\boxed{P_r = 30\,I_0^2\int_0^{\pi}[F(\theta)]^2\sin\theta\,d\theta \quad \text{W}} \tag{62}$$

Using this result, expressions for the directivity and radiation resistance can now be found. From Eq. (42), and using (60) and (62), the directivity in free space is

$$D(\theta) = \frac{4\pi K(\theta)}{P_r} = \frac{2[F(\theta)]^2}{\int_0^\pi [F(\theta)]^2 \sin\theta \, d\theta} \qquad (63)$$

whose maximum value is

$$D_{\max} = \frac{2\,[F(\theta)]_{\max}^2}{\int_0^\pi [F(\theta)]^2 \sin\theta \, d\theta} \qquad (64)$$

Finally, the radiation resistance will be

$$R_{\mathrm{rad}} = \frac{2P_r}{I_0^2} = 60 \int_0^\pi [F(\theta)]^2 \sin\theta \, d\theta \qquad (65)$$

EXAMPLE 14.2

Write the specific pattern function, and evaluate the beamwidth, directivity, and radiation resistance of a half-wave dipole.

Solution. The term "half-wave" refers to the overall length, in which $2\ell = \lambda/2$, or $\ell = \lambda/4$. Therefore, $k\ell = (2\pi/\lambda)(\lambda/4) = \pi/2$, which is now substituted into Eq. (59) to obtain:

$$F(\theta) = \frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} \qquad (66)$$

The magnitude of this function is plotted as the dashed curve in Figure 14.8a. Its maxima (equal to 1) occur at $\theta = \pi/2, 3\pi/2$, whereas zeros occur at $\theta = 0$ and π. Beamwidth is found by evaluating the solutions of

$$\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta} = \frac{1}{\sqrt{2}}$$

Numerically, it is found that the two angles on either side of the maximum at $\theta = 90°$ that satisfy the above equation are $\theta_{1/2} = 51°$ and $129°$. The half-power beamwidth is thus $129° - 51° = 78°$.

Directivity and radiation resistance are then found using (64) and (65), in which the integral of $[F(\theta)]^2$ can be performed numerically. The results are $D_{\max} = 1.64$ (or 2.15 dB), and $R_{\mathrm{rad}} = 73$ ohms.

D14.4. Evaluate the percentage of the maximum power density that is found in the direction $\theta = 45°$ for dipole antennas of overall length (a) $\lambda/4$, (b) $\lambda/2$, (c) λ.

Ans. 45.7%; 38.6%; 3.7%

In the half-wave dipole, the standing wave current amplitude maximizes at the feed point, and the antenna is said to be operated *on resonance*. As a result, the driving point impedance, one-quarter wavelength in front of the open ends, would *in principle* be purely real [4] and equal to the 73-Ω radiation resistance, assuming that the antenna is otherwise lossless. This is the primary motivation for using half-wave dipoles, in that they provide a fairly close impedance match to conventional transmission lines (whose characteristic impedances are on the same order).

Actually, because the antenna is essentially an unfolded transmission line, the half-wave dipole does not behave as an ideal quarter-wave transmission line section, as we might suspect considering the discussions in Section 14.1. An appreciable reactive part of the input impedance will likely be present, but the half-wavelength dimension is very close to the length at which the reactance goes to zero. Methods of evaluating the reactance are beyond the scope of our treatment, but are considered in detail in Ref. 1. For a thin lossless dipole of length exactly $\lambda/2$, the input impedance would be $Z_{in} = 73 + jX$, where X is in the vicinity of 40 Ω. The input reactance is extremely sensitive to the antenna length and can be reduced to zero by a very slight reduction in the overall length below $\lambda/2$, leaving the real part essentially unaffected. Similar behavior is seen in dipoles having lengths that are integer multiples of $\lambda/2$, but in these, radiation resistances are considerably higher, thus yielding a poorer impedance match. At dipole lengths between half-wavelength multiples, input reactances can be much higher (in the vicinity of $j600$ Ω) and can become sensitive to the wire thickness, in addition to the length. In practice, when connecting a transmission line feed, the input reactance can be zeroed by length reduction or by using matching techniques such as those discussed in Chapter 10.

Plots of directivity and radiation resistance as functions of antenna length are shown in Figure 14.9. Directivity increases modestly with length, whereas radiation resistance reaches a local maximum at a length between $3\lambda/4$ and λ. At greater lengths, additional peaks in R_{rad} occur at higher levels, but performance is compromised by the presence of sidelobes. Again, half-wave dipoles are typically used because single-lobe behavior is assured over a broad spectral bandwidth, whereas radiation resistance (73 Ω) is close to the impedance of standard transmission lines that are used to feed the antenna.

As a final exercise in wire antennas, we consider the operation of a *monopole* antenna. This is one-half a dipole plus a perfectly conducting plane, as shown in Figure 14.10*a*. The image principle discussed in Section 5.5 provides the image shown in Figure 14.10*b*, so that the monopole and its image form a dipole. Therefore, all field equations that pertain to the dipole apply directly to the upper half-space. The Poynting vector is therefore also the same above the plane, but the integration to find the total power radiated is performed only through the hemisphere that surrounds the upper half-space. So the radiated power and the radiation resistance for the monopole are half the corresponding values for the dipole. As an example, a quarter-wave monopole

---

[4]Think of a half-rotation ($\lambda/4$) around the Smith chart from the open circuit point, toward the generator, where, with loss present, the end position would be somewhere on the negative real axis.
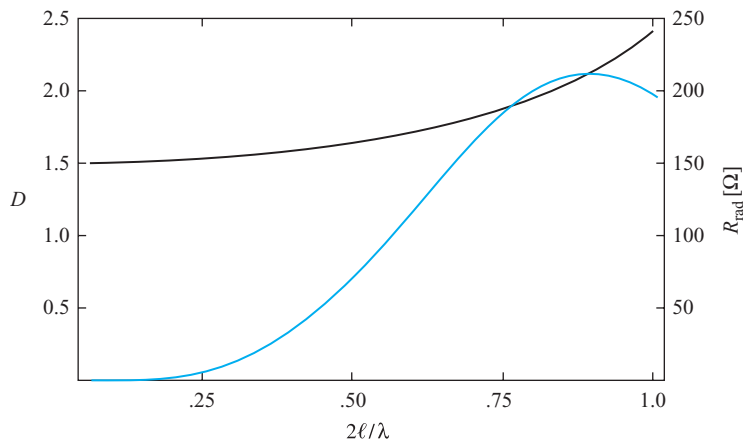
Figure 14.9 Plots of directivity (black) and radiation resistance (red) as functions of overall antenna length, expressed in wavelengths.

(presenting a half-wave dipole when including the image) yields a radiation resistance of $R_{\text{rad}} = 36.5\ \Omega$.

Monopole antennas may be driven by a coaxial cable below the plane, having its center conductor connected to the antenna through a small hole, and having its outer conductor connected to the plane. If the region below the plane is inaccessible or inconvenient, the coax may be laid on top of the plane and its outer conductor connected to it. Examples of this type of antenna include AM broadcasting towers and citizens' band antennas.
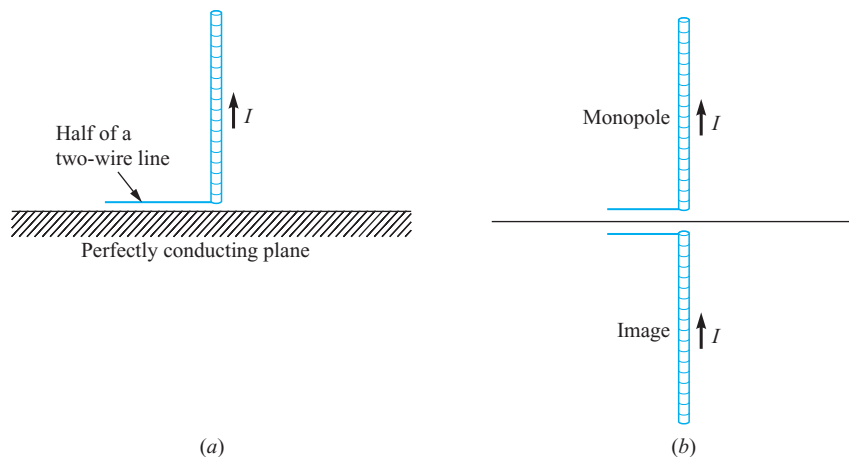
Figure 14.10 (a) An ideal monopole is always associated with a perfectly conducting plane. (b) The monopole plus its image form a dipole.

D14.5. The monopole antenna of Figure 14.10a has a length $d/2 = 0.080$ m and may be assumed to carry a triangular current distribution for which the feed current I_0 is 16.0 A at a frequency of 375 MHz in free space. At point P ($r = 400$ m, $\theta = 60°$, $\phi = 45°$) find (a) $H_{\phi s}$, (b) $E_{\theta s}$, and (c) the amplitude of $\mathcal{P}_r$.

Ans. $j1.7$ mA/m; $j0.65$ V/m; 1.1 mW/m^2

14.5 ARRAYS OF TWO ELEMENTS

We next address the problem of establishing better control of the directional properties of antenna radiation. Although some control of directivity is achieved through adjustment of the length of a wire antenna, these results only appear as changes in the E-plane pattern. The H-plane pattern always remains a circle (no ϕ variation), as long as a single vertical wire antenna is used. By using multiple elements in an *array*, significant improvement in directivity as determined in *both E and H planes* can be achieved. Our objective in this section is to lay the groundwork for the analysis of arrays by considering the simple case of using two elements. The resulting methods are readily extendable to multiple element configurations.

The basic configuration is shown in Figure 14.11. Here, we have our original wire antenna with its feed at the origin, and oriented along the z axis. A second identical antenna, parallel to the first, is positioned at location d on the x axis. The two carry the same current amplitude, I_0 (leading to far-field amplitude E_0), but we allow the second antenna current to exhibit a constant phase difference, ξ, from that of the first. The far-field observation point, P, lies at spherical coordinates, (r, θ, ϕ). From this point, the antennas appear close enough together so that (1) the radial lines, r and
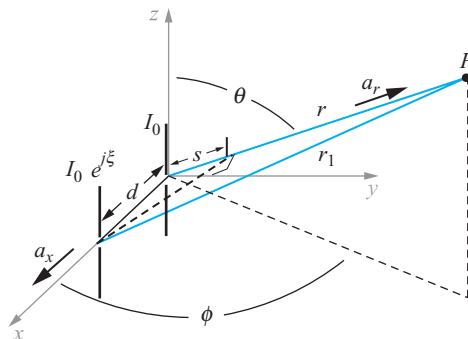
Figure 14.11 The original z-directed wire antenna with its center at the origin is now joined by a second parallel antenna, crossing the x axis at distance d. The second antenna carries the same current amplitude as the first, but with a constant phase shift, ξ. Fields are observed at point P.
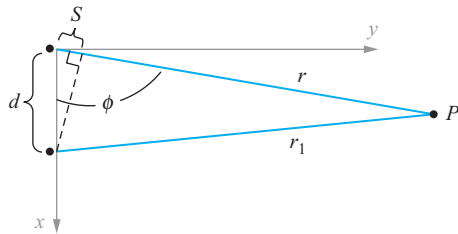
Figure 14.12 Top view of the arrangement
of Figure 14.11 (looking down onto the x-y
plane). In the far-field approximation, the red
lines are essentially parallel, and $r_1 \doteq r - s$.

r_1, are essentially parallel, and (2) the electric field directions at P are essentially the
same (along $\mathbf{a}_\theta$). Using Eq. (57), we may therefore write the total field at P, with the
understanding that the presence of the second antenna on the x axis will introduce a
ϕ dependence in the field that was previously not present:

$$E_{\theta P}(r, \theta, \phi) = E_0\, F(\theta) \left[\frac{e^{-jkr}}{r} + \frac{e^{j\xi}\, e^{-jkr_1}}{r_1} \right] \tag{67}$$

Next, we may express the distance to P from the second antenna, r_1, in terms of the
distance to the first antenna, r (also the spherical coordinate radius), by noting that in
the far-field approximation we have

$$r_1 \doteq r - s$$

where s is one leg of the right triangle formed by drawing a perpendicular line segment
between the second antenna and the line of radius, r, as shown in Figures 14.11 and
14.12. The length, s, is the projection of the antenna separation, d, onto the radial
line, r, and is found through

$$s = d\, \mathbf{a}_x \cdot \mathbf{a}_r = d \sin\theta \cos\phi \tag{68}$$

Therefore,

$$r_1 \doteq r - d \sin\theta \cos\phi \tag{69}$$

In the far-field, the distance, $d \sin\theta \cos\phi$, is very small compared to r, which
allows us to neglect the difference between r and r_1 in the magnitude terms in (67)
(so that $1/r_1 \doteq 1/r$). As we know from the dipole studies, the difference cannot be
neglected in the phase terms in (67) because phase is very sensitive to slight changes
in r. With these considerations in mind, Eq. (67) becomes

$$E_{\theta P}(r, \theta, \phi) = \frac{E_0\, F(\theta)}{r} \left[e^{-jkr} + e^{j\xi}\, e^{-jk(r - d \sin\theta \cos\phi)} \right] \tag{70}$$

which simplifies to

$$E_{\theta P}(r, \theta, \phi) = \frac{E_0\, F(\theta)}{r} e^{-jkr} \left[1 + e^{j\psi}\right] \tag{71}$$

where

$$\psi = \xi + kd\, \sin\theta \cos\phi \tag{72}$$

ψ is the net phase difference between the two antenna fields that is observed at $P(r, \theta, \phi)$. Equation (71) can be further simplified by factoring out the term $e^{j\psi/2}$ to obtain

$$E_{\theta P}(r, \theta, \phi) = \frac{2E_0\, F(\theta)}{r} e^{-jkr} e^{j\psi/2} \cos(\psi/2) \tag{73}$$

from which we may determine the field amplitude through

$$|E_{\theta P}(r, \theta, \phi)| = \sqrt{E_{\theta P} E_{\theta P}^*} = \frac{2E_0}{r} |F(\theta)|\, |\cos(\psi/2)| \tag{74}$$

Equation (74) demonstrates the important principle of *pattern multiplication* that applies to arrays of identical antennas. Specifically, the total field magnitude consists of the product of the pattern function magnitude, or *element factor* for the individual antennas, $|F(\theta)|$, and the normalized *array factor* magnitude, given by $|\cos(\psi/2)|$. The array factor is often denoted by

$$A(\theta, \phi) = \cos(\psi/2) = \cos\left[\tfrac{1}{2}\left(\xi + kd\, \sin\theta \cos\phi\right)\right] \tag{75}$$

Equation (74) then becomes

$$|E_{\theta P}(r, \theta, \phi)| = \frac{2E_0}{r} |F(\theta)|\, |A(\theta, \phi)| \tag{76}$$

This principle can be extended to arrays of multiple elements by appropriately modifying the array factor, as we will find. The underlying assumption is that the individual array elements are essentially uncoupled; that is, they induce negligible currents in each other. With appreciable coupling, the problem is far more complicated, and pattern multiplication cannot be used.

In the field pattern expressed in (76), the E plane (or θ dependence) is primarily determined by the individual elements, or by $|F(\theta)|$. It is in the H plane where the effect of the array is the strongest. In fact, the main reason for using an array of this configuration is to enable control of the H-plane pattern. In the H plane ($\theta = \pi/2$),

Eqs. (75) and (76) give the field dependence on ϕ as

$$E_{\theta P}(r, \pi/2, \phi) \propto A(\pi/2, \phi) = \cos\left[\tfrac{1}{2}(\xi + kd \cos \phi)\right] \qquad (77)$$

The H-plane pattern depends on the choices of the relative current phase, ξ, and the element spacing, d.

<hr>

EXAMPLE 14.3

Investigate the H-plane pattern when the currents are in phase ($\xi = 0$).

Solution. With $\xi = 0$, Eq. (77) becomes

$$A(\pi/2, \phi) = \cos\left[\frac{kd}{2} \cos \phi\right] = \cos\left[\frac{\pi d}{\lambda} \cos \phi\right]$$

This reaches a maximum at $\phi = \pi/2$ and $3\pi/2$, or along the direction that is normal to the plane of the antennas (the y axis). This occurs regardless of the choice of d, and the array is thus referred to as a *broadside array*. Now, by choosing $d = \lambda/2$, we obtain $A = \cos[(\pi/2) \cos \phi]$, which becomes zero at $\phi = 0$ and π (along the x axis), and we have single main beams along the positive and negative y axis. When d is increased beyond $\lambda/2$, additional maxima (sidelobes) appear as ϕ is varied, but zeros still occur along the x axis if d is set to odd multiples of $\lambda/2$.

<hr>

The broadside array of the previous example can be regarded as the simplest case. More interesting behavior occurs when a nonzero phase difference exists between the two currents, and adjustments can be performed in the phase and element spacing.

<hr>

EXAMPLE 14.4

Determine the necessary conditions to establish an *endfire* array, in which the maximum radiation is directed along the x axis.

Solution. Setting $\phi = 0$ or π in Eq. (77) and requiring the equation to achieve a maximum results in the condition:

$$A = \cos\left[\frac{\xi}{2} \pm \frac{\pi d}{\lambda}\right] = \pm 1$$

or

$$\frac{\xi}{2} \pm \frac{\pi d}{\lambda} = m\pi$$

where m is an integer that includes 0, and where the plus sign in the bracket applies for $\phi = 0$, and the minus sign for $\phi = \pi$. One case of practical interest occurs when $m = 0$, $d = \lambda/4$, and $\xi = -\pi/2$, which satisfies the above condition when the positive sign is chosen. Equation (77) now becomes

$$A(\pi/2, \phi) = \cos\left[\frac{\pi}{4} (\cos \phi - 1)\right]$$

This function maximizes at $\phi = 0$, and reaches zero at $\phi = \pi$. We have thus created an array that radiates a *single* main lobe along the positive x axis. The way this

works can be understood by realizing that the phase lag in current in the element at $x = d$ just compensates for the phase lag that arises from the propagation delay between the element at the origin and the one at $x = d$. The second element radiation is therefore precisely in phase with the radiation from the first element. The two fields, therefore, constructively interfere and propagate together in the forward x direction. In the reverse direction, radiation from the antenna at $x = d$ arrives at the origin to find itself π radians out of phase with the radiation from the $x = 0$ element. The two fields therefore destructively interfere, and no radiation occurs in the negative x direction.

D14.6. In the broadside configuration of Example 14.3, the element spacing is changed to $d = \lambda$. Determine (a) the ratio of the emitted intensities in the $\phi = 0$ and $\phi = 90°$ directions in the H plane, (b) the directions (values of ϕ) of the main beams in the H-plane pattern, and (c) the locations (values of ϕ) of the zeros in the H-plane pattern.

Ans. 1; (0, $\pm 90°$, 180°); ($\pm 45°$, $\pm 135°$)

D14.7. In the endfire configuration of Example 14.4, determine the directions (values of ϕ) for the main beams in the H plane if the wavelength is shortened from $\lambda = 4d$ to (a) $\lambda = 3d$, (b) $\lambda = 2d$, and (c) $\lambda = d$.

Ans. $\pm 41.4°$; $\pm 45.0°$; $\pm 75.5°$

14.6 UNIFORM LINEAR ARRAYS

We next expand our treatment to arrays of more than two elements. By doing this, more options are given to the designer that enable improvement of the directivity and possibly an increase in the bandwidth of the antenna, for example, As might be imagined, a full treatment of this subject would require an entire book. Here, we consider only the case of the uniform linear array to exemplify the analysis methods and to present some of the key results.

The uniform linear array configuration is shown in Figure 14.13. The array is linear because the elements are arranged along a straight line (the x axis in this case). The array is uniform because all elements are identical, have equal spacing, d, and carry the same current amplitude, I_0, and the phase progression in current from element to element is given by a constant value, ξ. The normalized array factor for the two-element array can be expressed using (71) as:

$$|A(\theta, \phi)| = |A_2(\theta, \phi)| = |\cos(\psi/2)| = \frac{1}{2}\left|1 + e^{j\psi}\right| \qquad (78)$$

where the subscript 2 is applied to A to indicate that the function applies to two elements. The array factor for a linear array of n elements as depicted in Figure 14.13
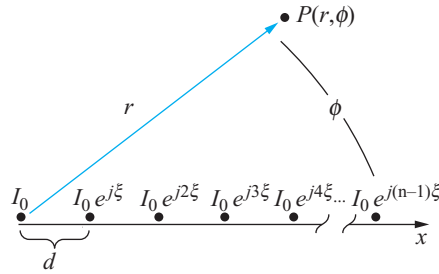
Figure 14.13 H-plane diagram of a uniform linear array of n dipoles, arranged along x, and with individual dipoles oriented along z (out of the page). All elements have equal spacing, d, and carry equal current amplitudes, I_0. Current phase shift ξ occurs between adjacent elements. Fields are evaluated at far-zone point, P, from which the dipoles appear to be grouped at the origin.

is a direct extension of (78), and becomes

$$|A_n(\theta, \phi)| = |A_n(\psi)| = \frac{1}{n} \left|1 + e^{j\psi} + e^{j2\psi} + e^{j3\psi} + e^{4\psi} + \ldots + e^{j(n-1)\psi}\right|$$

(79)

With the elements arranged along the x axis as shown in Figure 14.13, we have $\psi = \xi + kd \sin\theta \cos\phi$, as before. The geometric progression that comprises Eq. (79) can be expressed in closed form to give

$$|A_n(\psi)| = \frac{1}{n} \frac{\left|1 - e^{jn\psi}\right|}{\left|1 - e^{j\psi}\right|} = \frac{1}{n} \frac{\left|e^{jn\psi/2}\left(e^{-jn\psi/2} - e^{jn\psi/2}\right)\right|}{\left|e^{j\psi/2}\left(e^{-j\psi/2} - e^{j\psi/2}\right)\right|}$$

(80)

In the far right side of Eq. (80), we recognize the Euler identities for the sine function in both numerator and denominator, leading finally to

$$|A_n(\psi)| = \frac{1}{n} \frac{|\sin(n\psi/2)|}{|\sin(\psi/2)|}$$

(81)

The electric field in the far zone for an array of n dipoles can now be written in terms of A_n by extending the result in Eq. (76). Writing $|A_n(\psi)| = |A_n(\theta, \phi)|$, we have

$$|E_{\theta P}(r, \theta, \phi)| = \frac{nE_0}{r} |F(\theta)| \, |A_n(\theta, \phi)|$$

(82)

demonstrating again the principle of pattern multiplication, in which we now have a new array function that pertains to the linear array.
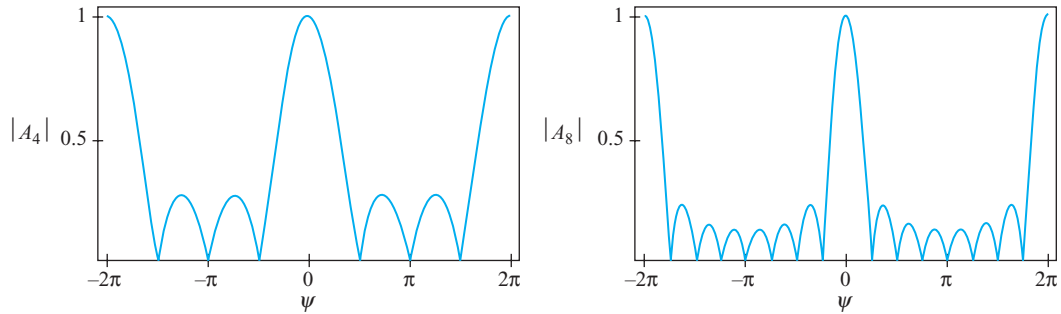
Figure 14.14 $|A_n(\psi)|$ as evaluated from Eq. (81) over the range $-2\pi < \psi < 2\pi$ for cases in which the number of elements, n, is (a) 4, and (b) 8.

Plots of Eq. (81) are shown in Figure 14.14 for the cases in which $n = 4$ and $n = 8$. Note that the functions always maximize to unity when $\psi = 2m\pi$, where m is an integer that includes zero. These principal maxima corrrespond to the main beams of the array pattern. The effect of increasing the number of elements is to narrow the main lobes and to bring in more secondary maxima (sidelobes).

To see how the array pattern is shaped, it is necessary to interpret the array function, Eq. (81), with regard to angular variation in the H plane. In this plane (where $\theta = \pi/2$), we have $\psi = \xi + kd \cos\phi$. Then, knowing that ϕ varies from 0 to 2π radians, $\cos\phi$ varies between ± 1, and we can see that ψ will be within the range

$$\xi - kd \leq \psi \leq \xi + kd \tag{83}$$

Choices of the current phasing, ξ, and the antenna spacing, d, determine the range of ψ values that will appear in the actual array pattern. This could lead, in some cases, to a fairly narrow range in ψ that may or may not include a principal maximum. The current phase determines the central value of ψ and the antenna spacing determines the maximum variation of ψ that occurs about the central value as the azimuth angle ϕ is varied.

As discussed in Section 14.5, a *broadside array* has main beams that occur normal to the array plane (at $\phi = \pi/2,\ 3\pi/2$). The condition for this is that the principal maximum, $\psi = 0$, will occur at these angles. We therefore write

$$\psi = 0 = \xi + kd \cos(\pi/2) = \xi$$

and so we would set $\xi = 0$ to obtain a broadside array. In this case, (83) gives $-kd < \psi < kd$. The central value of ψ is thus zero, and so the principal maximum there is included in the pattern. In the H plane, and with $\xi = 0$, we thus have $\psi = kd \cos\phi$. The $\psi = 0$ point will always occur at $\phi = \pi/2$ and $3\pi/2$, and this will be true regardless of the choice of element spacing d. The effect of increasing d is to enlarge the range of ψ that results when ϕ varies over its range of 0 to 2π. Therefore, for a given number of elements, the main beam will get narrower, but more sidelobes will be present in the pattern when the element spacing is increased.

An *endfire array* requires a principal maximum to occur along the x axis. In the H plane, we may therefore write

$$\psi = 0 = \xi + kd\cos(0) = \xi + kd$$

or $\xi = -kd$ to obtain endfire operation with a maximum occurring along the positive x axis. This may or may not result in a main beam occurring along the negative x axis as well.

For arrays of 4 and 8 elements, select the current phase and element spacing that will give *unidirectional endfire* operation, in which the main beam exists in the $\phi = 0$ direction, whereas no radiation occurs in the direction of $\phi = \pi$, nor in the broadside directions ($\phi = \pm\pi/2$).

Solution. We want $\psi = 0$ when $\phi = 0$. Therefore, from $\psi = \xi + kd\cos\phi$, we would require that $0 = \xi + kd$, or that $\xi = -kd$. Using 4 or 8 elements, we find either from Eq. (81) or from Figure 14.14 that zeros will occur when $\psi = \pm\pi/2$ and $\pm\pi$. Therefore, if we choose $\xi = -\pi/2$ and $d = \lambda/4$, we obtain $\psi = -\pi/2$ at $\phi = \pi/2$, $3\pi/2$, and $\psi = -\pi$ at $\phi = \pi$. We thus have $\psi = -(\pi/2)(1 - \cos\phi)$. Polar plots of the resulting array functions are shown in Figure 14.15a and b. Again, the move from 4 to 8 elements has the effect of decreasing the main beamwidth, while increasing the sidelobe count from 1 to 3, in this case. If an odd number of elements is used with the above choices in phasing and spacing, a small sidelobe will be present in the $\phi = \pi$ direction.

In general, we may choose current phasing and element spacing to establish the main beam in any direction. Choosing the $\psi = 0$ principal maximum, we may write

$$\psi = 0 = \xi + kd\cos\phi_{max} \quad \Rightarrow \quad \cos\phi_{max} = -\frac{\xi}{kd}$$

so that the main beam direction can be changed by varying the current phasing.

(a)   (b)
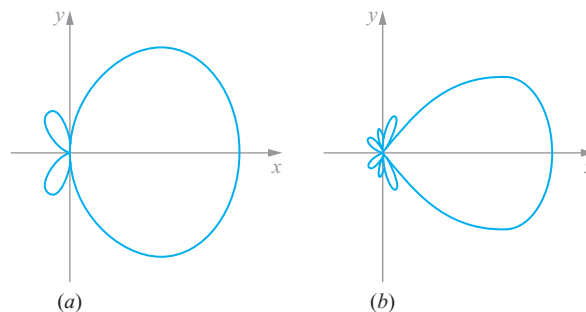
Figure 14.15 H-plane plots of (a) 4-element and (b) 8-element arrays having element spacing of $d = \lambda/4$, and current phasing $\xi = -\pi/2$.

D14.8. In an endfire linear dipole array in which $\xi = -kd$, what minimum element spacing d in wavelengths results in bi-directional operation, in which equal intensities occur in the H plane at $\phi = 0$ and $\phi = \pi$?

Ans. $d = \lambda/2$

D14.9. For a linear dipole array in which the element spacing is $d = \lambda/4$, what current phase ξ will result in a main beam in the direction of a) $\phi = 30°$; b) $\phi = 45°$.

Ans. $-\pi\sqrt{3}/4$; $-\pi\sqrt{2}/4$

14.7 ANTENNAS AS RECEIVERS

We next turn to the other fundamental purpose of an antenna, which is its use as a means to detect, or receive, radiation that originates from a distant source. We will approach this problem through study of a transmit-receive antenna system. This is composed of two antennas, along with their supporting electronics, that play the interchangeable roles of transmitter and detector.

Figure 14.16 shows an example of a transmit-receive arrangement, in which the two coupled antennas together comprise a linear two-port network. Voltage V_1 and current I_1 on the antenna at the left affect the voltage and current (V_2 and I_2) on the antenna at the right—and vice-versa. This coupling is quantified through trans-impedance parameters, Z_{12} and Z_{21}. The governing equations take the form

$$V_1 = Z_{11}I_1 + Z_{12}I_2 \qquad (84a)$$

$$V_2 = Z_{21}I_1 + Z_{22}I_2 \qquad (84b)$$

Z_{11} and Z_{22} are the input impedances to antennas 1 and 2 when either antenna is isolated and is used as a transmitter, or equivalently, if the two antennas are sufficiently far away from each other. The real parts of Z_{11} and Z_{22} will be the associated radiation
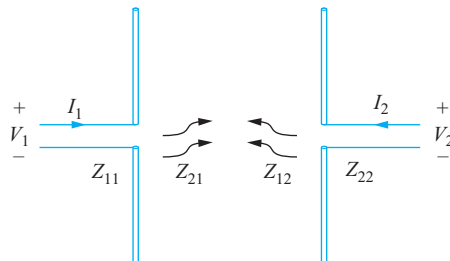
Figure 14.16 A pair of coupled antennas, demonstrating Eqs. (84a) and (84b).

resistances, provided ohmic losses in all conductors and all losses to surrounding objects are reduced to zero. We will assume this, in addition to far-zone operation, to be true here. The trans-impedances, Z_{12} and Z_{21}, depend on the spacing and relative orientation between the antennas, as well as on the characteristics of the surrounding medium. A critical property of the transimpedances in a linear medium is that they are equal. This property is the embodiment of the *reciprocity theorem*. Stated simply,

$$Z_{12} = Z_{21} \tag{85}$$

Further insights can be found by inverting (84*a*) and (84*b*), and invoking the admittance parameters, Y_{ij}:

$$I_1 = Y_{11} V_1 + Y_{12} V_2 \tag{86a}$$
$$I_2 = Y_{21} V_1 + Y_{22} V_2 \tag{86b}$$

where, again, the reciprocity theorem tells us that $Y_{12} = Y_{21}$.

Now, suppose that the terminals of antenna 2 are shorted, so that $V_2 = 0$. In this case, Eq. (86*b*) gives $I_2' = Y_{21} V_1'$, where the single prime denotes the condition of a shorted antenna 2. Instead, we could short antenna 1, resulting in $I_1'' = Y_{12} V_2''$ (with the double prime indicating conditions with antenna 1 shorted). Because reciprocity holds, it follows that

$$\frac{V_2''}{I_1''} = \frac{V_1'}{I_2'} \tag{87}$$

Equation (87) applies regardless of the relative positioning and orientation of the two antennas. We know that in a given direction, each antenna will transmit a power density whose value is determined by the antenna radiation pattern. Furthermore, we would expect to see the current that is set up on the receiving antenna depend on that antenna's orientation; that is, there is a *reception pattern* that the receiving antenna presents to the incoming signal. Now, for a fixed relative orientation between the two antennas, with antenna 1 as the transmitter, and antenna 2 shorted, a certain ratio V_1'/I_2' will be obtained. This ratio will depend on the relative orientation, which in turn will depend on the radiation pattern of antenna 1 and on the reception pattern of antenna 2. If roles are reversed such that the transmitter now becomes the receiver, and with antenna 1 shorted, a ratio V_2''/I_1'' will be obtained that by Eq. (87) is the same as before. The conclusion we must come to is that the extent to which the receiving antenna *accepts* power will be determined by its *radiation* pattern. This means, for example, that the main beam direction in the radiation pattern of the receiving antenna corresponds to the direction from which it is most sensitive to incoming signals. *The radiation and receiving patterns of any antenna are the same.*

We next consider a more general transmission case, in which the receiving antenna is to deliver power to a load. Antenna 1 (Figure 14.16) serves as the transmitter, while antenna 2 is the receiver, at which the load is attached. A primary assumption is that the antennas are far enough away from each other so that only forward coupling (through Z_{21}) will be appreciable. The large separation distance means that the induced current
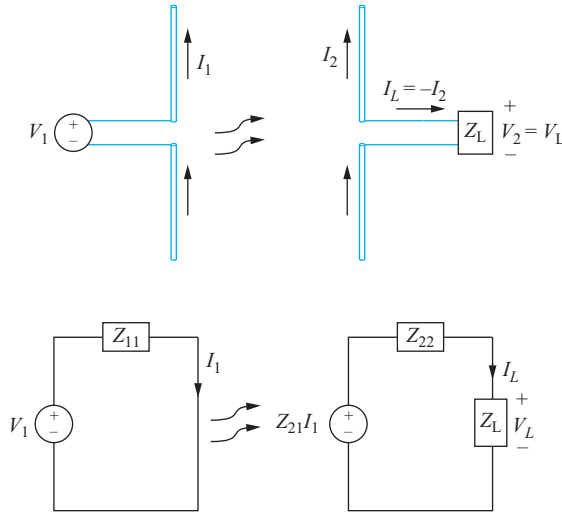
Figure 14.17 Transmitting and receiving antennas, and their equivalent circuits.

I_2 is likely to be much less than I_1. Reverse coupling (through Z_{12}) would involve transmission of the *received* signal in antenna 2 back to antenna 1; specifically, the induced current I_2 further induces a (now very weak) additional current I_1' on antenna 1; that antenna would then carry a net current of $I_1 + I_1'$, where $I_1' \ll I_1$. We therefore assume that the product $Z_{12}I_2$ can be neglected, under which Eq. (84a) gives $V_1 = Z_{11}I_1$. A load impedance, Z_L, is connected across the terminals of antenna 2, as shown in the upper part of Figure 14.17. V_2 is the voltage across this load. Current $I_L = -I_2$ now flows through the load. Taking this current as positive, Eq. (84b) becomes

$$V_2 = V_L = Z_{21}I_1 - Z_{22}I_L \tag{88}$$

This is just the Kirchoff voltage law equation for the right-hand equivalent circuit shown in the lower part of Figure 14.17. The term $Z_{21}I_1$ is interpreted as the source voltage for this circuit, originating from antenna 1. Using (88), along with $V_L = Z_L I_L$, leads to

$$I_L = \frac{Z_{21}I_1}{Z_{22} + Z_L} \tag{89}$$

The time-average power dissipated by Z_L is now

$$P_L = \frac{1}{2}\mathcal{R}e\left\{V_L I_L^*\right\} = \frac{1}{2}|I_L|^2 \mathcal{R}e\left\{Z_L\right\} = \frac{1}{2}|I_1|^2 \left|\frac{Z_{21}}{Z_{22} + Z_L}\right|^2 \mathcal{R}e\left\{Z_L\right\} \tag{90}$$

The maximum power transferred to the load occurs when the load impedance is conjugate-matched to the driving point impedance, or $Z_L = Z_{22}^*$. Making this

substitution in (90), and using the fact that $Z_{22} + Z_{22}^* = 2R_{22}$ gives

$$P_L = \frac{1}{2}|I_1|^2 \left|\frac{Z_{21}}{2R_{22}}\right|^2 \mathcal{R}e\{Z_{22}\} = \frac{|I_1|^2 |Z_{21}|^2}{8R_{22}} \tag{91}$$

The time-average power transmitted by antenna 1 is

$$P_r = \frac{1}{2}\mathcal{R}e\{V_1 I_1^*\} = \frac{1}{2}R_{11}|I_1|^2 \tag{92}$$

By comparing the above result with Eq. (65), R_{11} is interpreted as the radiation resistance of the transmitting antenna if (1) there are no resistive losses, and (2) the current amplitude at the driving point is the maximum amplitude, I_0. As we found earlier, the latter will occur in a dipole if the overall antenna length is an integer multiple of a half-wavelength. Using (91) and (92), we write the ratio of the received and transmitted powers:

$$\frac{P_L}{P_r} = \frac{|Z_{21}|^2}{4R_{11}R_{22}} \tag{93}$$

At this stage, more understanding is needed of the transimpedance, Z_{21} (or Z_{12}). This quantity will depend on the distance and relative orientations of the two antennas, in addition to other parameters. Figure 14.18 shows two dipole antennas, separated
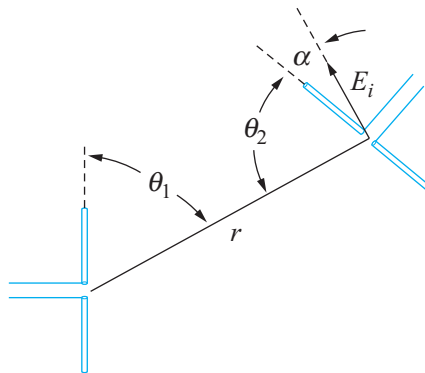
Figure 14.18   A transmit-receive antenna pair, showing relative orientation angles for the case in which the antennas lie in the same plane (in which case the ϕ coordinates are not necessary). Incident electric field, E_i, from antenna 1 is shown arriving at antenna 2, and presenting angle α to the antenna 2 axis. The field is perpendicular to the distance line r, and thus $\alpha = 90° - \theta_2$. Far-zone operation is assumed, so that the two antennas appear as point objects to each other.

by radial distance, r, and with relative orientations that are specified by values of θ, as measured with respect to *each* antenna axis.[5] With antenna 1 serving as the transmitter and antenna 2 serving as the receiver, the radiation pattern of antenna 1 is given as a function of θ_1 and ϕ_1, while the receiving pattern of antenna 2 (equivalent to *its* radiation pattern) is given as a function of θ_2 and ϕ_2.

A convenient way to express the received power in an antenna is through its *effective area*, denoted $A_e(\theta, \phi)$, and expressed in m². Refer to Figure 14.18, and consider the average power density at the receiver (antenna 2) position, originating from the transmitter (antenna 1). As per Eqs. (25) and (26), this will be the magnitude of the Poynting vector at that location, $S_r(r, \theta_1, \phi_1)$ in W/m², where a dependence on ϕ is now necessary to describe all possible relative orientations. The effective area of the receiving antenna is defined such that when the power density is multiplied by the effective area, the power dissipated by a matched load at the receiving antenna is obtained. With antenna 2 as the receiver, we write

$$P_{L2} = S_{r1}(r, \theta_1, \phi_1) \times A_{e2}(\theta_2, \phi_2) \quad [\text{W}] \tag{94}$$

But now, using Eqs. (34) and (38), we may write the power density in terms of the directivity of antenna 1:

$$S_{r1}(r, \theta_1, \phi_1) = \frac{P_{r1}}{4\pi r^2} D_1(\theta_1, \phi_1) \tag{95}$$

Combining Eqs. (94) and (95), we obtain the ratio of the power received by antenna 2 to the power radiated by antenna 1:

$$\frac{P_{L2}}{P_{r1}} = \frac{A_{e2}(\theta_2, \phi_2) D_1(\theta_1, \phi_1)}{4\pi r^2} = \frac{|Z_{21}|^2}{4R_{11}R_{22}} \tag{96}$$

where the second equality repeats Eq. (93). We can solve (96) to find

$$|Z_{21}|^2 = \frac{R_{11}R_{22}A_{e2}(\theta_2, \phi_2)D_1(\theta_1, \phi_1)}{\pi r^2} \tag{97a}$$

We next note that if roles are reversed, in which antenna 2 transmits to antenna 1, we would find

$$|Z_{12}|^2 = \frac{R_{11}R_{22}A_{e1}(\theta_1, \phi_1)D_2(\theta_2, \phi_2)}{\pi r^2} \tag{97b}$$

---

[5] One way to express the relative orientations is to define the z axis along the radial distance line, r. Then angles θ_i and ϕ_i ($i = 1, 2$) are used locally to describe the orientations of the antenna axes, in which the origins of the two spherical coordinate systems are located at each antenna feed. The ϕ coordinate would thus be the angle of rotation about the r axis. For example, in Figure 14.18, with both antennas in the plane of the page, both ϕ coordinates could be assigned values of zero. With antenna 2 rotated about r such that it is normal to the page, ϕ_2 would be 90°, and the antennas would be cross-polarized.

The reciprocity theorem states that $Z_{12} = Z_{21}$. By equating Eqs. (97a) and (97b), it therefore follows that

$$\frac{D_1(\theta_1, \phi_1)}{A_{e1}(\theta_1, \phi_1)} = \frac{D_2(\theta_2, \phi_2)}{A_{e2}(\theta_2, \phi_2)} = \text{Constant} \tag{98}$$

That is, the ratio of directivity to effective area for *any* antenna is a universal constant, independent of the antenna type or the direction in which these parameters are evaluated. To evaluate the constant, we only need to look at one case.

EXAMPLE 14.6

Find the effective area of a Hertzian dipole, and determine the general relation between the directivity and effective area of any antenna.

Solution. With the Hertzian dipole as the receiving antenna, and having length d, its load voltage, V_L, will depend on the electric field that it intercepts from antenna 1. Specifically, we find the projection of the transmitting antenna field along the length of receiving antenna. This projected field, when multiplied by the length of antenna 2, gives the input voltage to the receiving antenna equivalent circuit. Referring to Figure 14.18, the projection angle is α, and thus the voltage that drives the Hertzian dipole will be

$$V_{\text{in}} = E_i \cos \alpha \times d = E_i d \sin \theta_2$$

The equivalent circuit for the Hertzian dipole is now the same as that of the receiving antenna as shown in Figure 14.17, except that we replace the source voltage, $I_1 Z_{21}$, with V_{in} as given above. Assuming a conjugate-matched load ($Z_L = Z_{22}^*$), the current through the load is now

$$I_L = \frac{E_i d \sin \theta_2}{Z_{22} + Z_L} = \frac{E_i d \sin \theta_2}{2 R_{22}}$$

The power delivered to the matched load is then

$$P_{L2} = \frac{1}{2} \mathcal{R}e \left\{ V_L I_L^* \right\} = \frac{1}{2} R_{22} |I_L|^2 = \frac{(E_i d)^2 \sin^2 \theta_2}{8 R_{22}} \tag{99}$$

For the Hertzian dipole, R_{22} is the radiation resistance. This was previously found to be (Eq. (30))

$$R_{22} = R_{\text{rad}} = 80 \pi^2 \left(\frac{d}{\lambda} \right)^2$$

Substituting this into (99), we find

$$P_{L2} = \frac{1}{640} \left(\frac{E_i \lambda \sin \theta_2}{\pi} \right)^2 \quad \text{[Watts]} \tag{100}$$

The average power density that is incident on the receiving antenna is now

$$S_{r1}(r, \theta_1, \phi_1) = \frac{E_i(r, \theta_1, \phi_1)^2}{2 \eta_0} = \frac{E_i^2}{240 \pi} \quad \text{[Watts/m}^2\text{]} \tag{101}$$

Using (100) and (101), the effective area of the Hertzian is

$$A_{e2}(\theta_2) = \frac{P_{L2}}{S_{r1}} = \frac{3}{8\pi}\lambda^2 \sin^2(\theta_2) \quad [\text{m}^2] \tag{102}$$

The directivity for the Hertzian dipole, derived in Example 14.1, is

$$D_2(\theta_2) = \frac{3}{2}\sin^2(\theta_2) \tag{103}$$

Comparing Eqs. (102) and (103), we find the relation that we are looking for: the effective area and directivity for *any* antenna are related through

$$\boxed{D(\theta, \phi) = \frac{4\pi}{\lambda^2} A_e(\theta, \phi)} \tag{104}$$

We can now return to Eq. (96) and use Eq. (104) to rewrite the ratio of the power delivered to the receiving antenna load to the total power radiated by the transmitting antenna: this yields an expression that involves the simple product of the effective areas, known as the *Friis transmission formula*:

$$\boxed{\frac{P_{L2}}{P_{r1}} = \frac{A_{e2}(\theta_2, \phi_2)\, D_1(\theta_1, \phi_1)}{4\pi r^2} = \frac{A_{e1}(\theta_1, \phi_1) A_{e2}(\theta_2, \phi_2)}{\lambda^2 r^2}} \tag{105}$$

The result can also be expressed in terms of the directivities:

$$\boxed{\frac{P_{L2}}{P_{r1}} = \frac{\lambda^2}{(4\pi r)^2} D_1(\theta_1, \phi_1) D_2(\theta_2, \phi_2)} \tag{106}$$

These results provide an effective summary of what was discussed in this section, by way of giving us a very useful design tool for a free space communication link. Again, Eq. (105) assumes lossless antennas in the far zones of each other and gives the power dissipated by a load that is conjugate matched to the receiving antenna impedance.

D14.10. Given: an antenna having a maximum directivity of 6 dB and operating at wavelength, $\lambda = 1$ m. What is the maximum effective area of the antenna?

Ans. $1/\pi$ m^2

D14.11. The power of 1 mW is dissipated by the matched load of a receiving antenna of a 1-m^2 effective area. This antenna is positioned at the center of the main beam of the transmitting antenna, located 1.0 km away. What total power is radiated by the transmitter if its directivity is (*a*) 10dB, (*b*) 7dB?

Ans. 4π kW; 8π kW

REFERENCES

1. C. Balanis, *Antenna Theory: Analysis and Design, 3rd ed.*, Wiley, Hoboken, 2005. A widely used text at the advanced senior or graduate levels, offering much detail.

2. S. Silver, ed., *Microwave Antenna Theory and Design*, Peter Peregrinus, Ltd on behalf of IEE, London, 1984. This is a reprint of volume 9 of the famous MIT Radiation Laboratory series, originally published by McGraw-Hill in 1949. It contains much information from original sources, that later appeared in the modern textbooks.

3. E.C. Jordan and K.G. Balmain, *Electromagnetic Waves and Radiating Systems,* 2nd ed., Prentice-Hall, Inc., Englewood Cliffs, NJ, 1968. A classic text, covering waveguides and antennas.

4. L.V. Blake, *Antennas*, Wiley, New York, 1966. A short, well-written and very readable text at a basic level.

5. G.S. Smith, *Classical Electromagnetic Radiation*, Cambridge, 1997. This excellent graduate-level text provides a unique perspective and rigorous treatment of the radiation problem as related to all types of antennas.

CHAPTER 14   PROBLEMS

14.1 A short dipole-carrying current $I_0 \cos \omega t$ in the $\mathbf{a}_z$ direction is located at the origin in free space. (*a*) If $k = 1$ rad/m, $r = 2$ m, $\theta = 45°$, $\phi = 0$, and $t = 0$, give a unit vector in rectangular components that shows the instantaneous direction of $\mathbf{E}$. (*b*) What fraction of the total average power is radiated in the belt, $80° < \theta < 100°$?

14.2 Prepare a curve, r vs. θ in polar coordinates, showing the locus in the $\phi = 0$ plane where (*a*) the radiation field $|E_{\theta s}|$ is one-half of its value at $r = 10^4$ m, $\theta = \pi/2$; (*b*) average radiated power density $<\mathcal{S}_r>$ is one-half its value at $r = 10^4$ m, $\theta = \pi/2$.

14.3 Two short antennas at the origin in free space carry identical currents of $5 \cos \omega t$ A, one in the $\mathbf{a}_z$ direction, and one in the $\mathbf{a}_y$ direction. Let $\lambda = 2\pi$ m and $d = 0.1$ m. Find $\mathbf{E}_s$ at the distant point where (*a*) ($x = 0$, $y = 1000$, $z = 0$); (*b*) (0, 0, 1000); (*c*) (1000, 0, 0). (*d*) Find $\mathbf{E}$ at (1000, 0, 0) at $t = 0$. (*e*) Find $|\mathbf{E}|$ at (1000, 0, 0) at $t = 0$.

14.4 Write the Hertzian dipole electric field whose components are given in Eqs. (15) and (16) in the near zone in free space where $kr << 1$. In this case, only a single term in each of the two equations survives, and the phases, δ_r and δ_θ, simplify to a single value. Construct the resulting electric field vector and compare your result to the static dipole result (Eq. (36) in Chapter 4). What relation must exist between the static dipole charge, Q, and the current amplitude, I_0, so that the two results are identical?

14.5 Consider the term in Eq. (14) (or in Eq. (10)) that gives the $1/r^2$ dependence in the Hertzian dipole magnetic field. Assuming this term dominates and that $kr << 1$, show that the resulting magnetic field is the same as that found by applying the Biot–Savart law (Eq. (2), Chapter 7) to a

current element of differential length d, oriented along the z axis, and centered at the origin.

14.6 Evaluate the time-average Poynting vector, $<\mathbf{S}> = \left(\frac{1}{2}\right)\mathcal{R}e\left\{\mathbf{E}_s \times \mathbf{H}_s^*\right\}$ for the Hertzian dipole, assuming the general case that involves the field components as given by Eqs. (10), (13a), and (13b). Compare your result to the far-zone case, Eq. (26).

14.7 A short current element has $d = 0.03\lambda$. Calculate the radiation resistance that is obtained for each of the following current distributions: (a) uniform, I_0; (b) linear, $I(z) = I_0(0.5d - |z|)/0.5d$; (c) step, I_0 for $0 < |z| < 0.25d$ and $0.5I_0$ for $0.25d < |z| < 0.5d$.

14.8 Evaluate the time-average Poynting vector, $<\mathbf{S}> = (1/2)\mathcal{R}e\left\{\mathbf{E}_s \times \mathbf{H}_s^*\right\}$ for the magnetic dipole antenna in the far zone, in which all terms of order $1/r^2$ and $1/r^4$ are neglected in Eqs. (48), (49), and (50). Compare your result to the far-zone power density of the Hertzian dipole, Eq. (26). In this comparison, and assuming equal current amplitudes, what relation between loop radius, a, and dipole length, d, would result in equal radiated powers from the two devices?

14.9 A dipole antenna in free space has a linear current distribution with zero current at each end, and with peak current I_0 at the enter. If the length d is 0.02λ, what value of I_0 is required to (a) provide a radiation-field amplitude of 100 mV/m at a distance of 1 mi, at $\theta = 90°$; (b) radiate a total power of 1 W?

14.10 Show that the chord length in the E-plane plot of Figure 14.4 is equal to $b \sin \theta$, where b is the circle diameter.

14.11 A monopole antenna extends vertically over a perfectly conducting plane, and has a linear current distribution. If the length of the antenna is 0.01λ, what value of I_0 is required to (a) provide a radiation-field amplitude of 100 mV/m at a distance of 1 mi, at $\theta = 90°$; (b) radiate a total power of 1 W? Assume free space above the plane.

14.12 Find the zeros in θ for the E-plane pattern of a dipole antenna for which (a) $\ell = \lambda$; (b) $2\ell = 1.3\lambda$. Use Figure 14.8 as a guide.

14.13 The radiation field of a certain short vertical current element is $E_{\theta s} = (20/r) \sin \theta \, e^{-j10\pi r}$ V/m if it is located at the origin in free space. (a) Find $E_{\theta s}$ at $P(r = 100, \theta = 90°, \phi = 30°)$. (b) Find $E_{\theta s}$ at $P(100, 90°, 30°)$ if the vertical element is located at $A(0.1, 90°, 90°)$. (c) Find $E_{\theta s}$ at $P(100, 90°, 30°)$ if identical vertical elements are located at $A(0.1, 90°, 90°)$ and $B(0.1, 90°, 270°)$.

14.14 For a dipole antenna of overall length $2\ell = \lambda$, evaluate the maximum directivity in decibels, and the half-power beamwidth.

14.15 For a dipole antenna of overall length $2\ell = 1.3\lambda$, determine the locations in θ and the peak intensity of the sidelobes, expressed as a fraction of the main

lobe intensity. Express your result as the sidelobe level in decibels, given by S_s [dB] $= 10 \log_{10}(S_{r,\text{main}}/S_{r,\text{sidelobe}})$. Again, use Figure 14.8 as a guide.

14.16 For a dipole antenna of overall length, $2\ell = 1.5\lambda$, (a) evaluate the locations in θ at which the zeros and maxima in the E-plane pattern occur; (b) determine the sidelobe level, as per the definition in Problem 14.14; (c) determine the maximum directivity.

14.17 Consider a lossless half-wave dipole in free space, with radiation resistance, $R_{\text{rad}} = 73$ ohms, and maximum directivity $D_{\text{max}} = 1.64$. If the antenna carries a 1-A current amplitude, (a) how much total power (in watts) is radiated? (b) How much power is intercepted by a 1-m² aperture situated at distance $r = 1$ km away? The aperture is on the equatorial plane and squarely faces the antenna. Assume uniform power density over the aperture.

14.18 Repeat Problem 14.17, but with a full-wave antenna ($2\ell = \lambda$). Numerical integrals may be necessary.

14.19 Design a two-element dipole array that will radiate equal intensities in the $\phi = 0, \pi/2, \pi$, and $3\pi/2$ directions in the H plane. Specify the smallest relative current phasing, ξ, and the smallest element spacing, d.

14.20 A two-element dipole array is configured to provide zero radiation in the broadside ($\phi = \pm 90°$) and endfire ($\phi = 0, 180°$) directions, but with maxima occurring at angles in between. Consider such a set-up with $\psi = \pi$ at $\phi = 0$ and $\psi = -3\pi$ at $\phi = \pi$, with both values determined in the H-plane. a) Verify that these values give zero broadside and endfire radiation. b) Determine the required relative current phase, ξ. b) Determine the required element spacing, d. c) Determine the values of ϕ at which maxima in the radiation pattern occur.

14.21 In the two-element endfire array of Example 14.4, consider the effect of varying the operating frequency, f, away from the original design frequency, f_0, while maintaining the original current phasing, $\xi = -\pi/2$. Determine the values of ϕ at which the maxima occur when the frequency is changed to (a) $f = 1.5 f_0$; (b) $f = 2 f_0$.

14.22 Revisit Problem 14.21, but with the current phase allowed to vary with frequency (this will automatically occur if the phase difference is established by a simple time delay between the feed currents). Now, the current phase difference will be $\xi' = \xi \, f/f_0$, where f_0 is the original (design) frequency. Under this condition, radiation will maximize in the $\phi = 0$ direction regardless of frequency (show this). Backward radiation (along $\phi = \pi$) will develop, however, as the frequency is tuned away from f_0. Derive an expression for the *front-to-back ratio*, defined as the ratio of the radiation intensities at $\phi = 0$ and $\phi = \pi$, expressed in decibels. Express this result as a function of the frequency ratio f/f_0. Evaluate the front-to-back ratio for (a) $f = 1.5 f_0$; (b) $f = 2 f_0$ and; (c) $f = 0.75 f_0$.

14.23 A *turnstile* antenna consists of two crossed dipole antennas, positioned in this case in the xy plane. The dipoles are identical, lie along the x and y axes, and are both fed at the origin. Assume that equal currents are supplied to each antenna and that a zero phase reference is applied to the x-directed antenna. Determine the relative phase, ξ, of the y-directed antenna so that the net radiated electric field as measured on the positive z axis is (a) left circularly polarized; (b) linearly polarized along the $45°$ axis between x and y.

14.24 Consider a linear endfire array, designed for maximum radiation intensity at $\phi = 0$, using ξ and d values as suggested in Example 14.5. Determine an expression for the front-to-back ratio (defined in Problem 14.22) as a function of the number of elements, n, if n is an odd number.

14.25 A six-element linear dipole array has element spacing $d = \lambda/2$. (a) Select the appropriate current phasing, ξ, to achieve maximum radiation along $\phi = \pm 60°$. (b) With the phase set as in part (a), evaluate the intensities (relative to the maximum) in the broadside and endfire directions.

14.26 In a linear endfire array of n elements, a choice of current phasing that improves the directivity is given by the Hansen–Woodyard condition:

$$\xi = \pm \left(\frac{2\pi d}{\lambda} + \frac{\pi}{n} \right)$$

where the plus or minus sign choices give maximum radiation along $\phi = 180°$ and $0°$, respectively. Applying this phasing may not necessarily lead to unidirectional endfire operation (zero backward radiation), but it will do so with the proper choice of element spacing, d. (a) Determine this required spacing as a function of n and λ. (b) Show that the spacing as found in part (a) approaches $\lambda/4$ for a large number of elements. (c) Show that an even number of elements is required.

14.27 Consider an n-element broadside linear array. Increasing the number of elements has the effect of narrowing the main beam. Demonstrate this by evaluating the separation in ϕ between the zeros on either side of the principal maximum at $\phi = 90°$. Show that for large n this separation is approximated by $\Delta\phi \doteq 2\lambda/L$, where $L \doteq nd$ is the overall length of the array.

14.28 A large ground-based transmitter radiates 10 kW and communicates with a mobile receiving station that dissipates 1mW on the matched load of its antenna. The receiver (not having moved) now transmits back to the ground station. If the mobile unit radiates 100 W, what power is received (at a matched load) by the ground station?

14.29 Signals are transmitted at a 1-m carrier wavelength between two identical half-wave dipole antennas spaced by 1 km. The antennas are oriented such that they are exactly parallel to each other. (a) If the transmitting antenna

radiates 100 watts, how much power is dissipated by a matched load at the receiving antenna? (*b*) Suppose the receiving antenna is rotated by 45° while the two antennas remain in the same plane. What is the received power in this case?

14.30 A half-wave dipole antenna is known to have a maximum effective area, given as A_{max}. (*a*) Write the maximum directivity of this antenna in terms of A_{max} and wavelength λ. (*b*) Express the current amplitude, I_0, needed to radiate total power, P_r, in terms of P_r, A_{max}, and λ. (*c*) At what values of θ and ϕ will the antenna effective area be equal to A_{max}?

Vector Analysis

A.1 GENERAL CURVILINEAR COORDINATES

Let us consider a general orthogonal coordinate system in which a point is located by the intersection of three mutually perpendicular surfaces (of unspecified form or shape),

$$u = \text{constant}$$
$$v = \text{constant}$$
$$w = \text{constant}$$

where u, v, and w are the variables of the coordinate system. If each variable is increased by a differential amount and three more mutually perpendicular surfaces are drawn corresponding to these new values, a differential volume is formed which approximates a rectangular parallelepiped. Because u, v, and w need not be measures of length, such as, the angle variables of the cylindrical and spherical coordinate systems, each must be multiplied by a general function of u, v, and w in order to obtain the differential sides of the parallelepiped. Thus we define the scale factors h_1, h_2, and h_3 each as a function of the three variables u, v, and w and write the lengths of the sides of the differential volume as

$$dL_1 = h_1 du$$
$$dL_2 = h_2 dv$$
$$dL_3 = h_3 dw$$

In the three coordinate systems discussed in Chapter 1, it is apparent that the variables and scale factors are

$$
\begin{array}{llll}
\text{Rectangular:} & u = x & v = y & w = z \\
 & h_1 = 1 & h_2 = 1 & h_3 = 1 \\
\text{Cylindrical:} & u = \rho & v = \phi & w = z \\
 & h_1 = 1 & h_2 = \rho & h_3 = 1 \\
\text{Spherical:} & u = r & v = \theta & w = \phi \\
 & h_1 = 1 & h_2 = r & h_3 = r \sin \theta
\end{array}
\tag{A.1}
$$

553

The choice of u, v, and w has been made so that $\mathbf{a}_u \times \mathbf{a}_v = \mathbf{a}_w$ in all cases. More involved expressions for h_1, h_2, and h_3 are to be expected in other less familiar coordinate systems.[1]

A.2 DIVERGENCE, GRADIENT, AND CURL IN GENERAL CURVILINEAR COORDINATES

If the method used to develop divergence in Sections 3.4 and 3.5 is applied to the general curvilinear coordinate system, the flux of the vector $\mathbf{D}$ passing through the surface of the parallelepiped whose unit normal is $\mathbf{a}_u$ is

$$D_{u0}dL_2dL_3 + \frac{1}{2}\frac{\partial}{\partial u}(D_u dL_2 dL_3)du$$

or

$$D_{u0}h_2h_3 dv\,dw + \frac{1}{2}\frac{\partial}{\partial u}(D_u h_2 h_3 dv\,dw)du$$

and for the opposite face it is

$$-D_{u0}h_2h_3 dv\,dw + \frac{1}{2}\frac{\partial}{\partial u}(D_u h_2 h_3 dv\,dw)du$$

giving a total for these two faces of

$$\frac{\partial}{\partial u}(D_u h_2 h_3 dv\,dw)du$$

Because u, v, and w are independent variables, this last expression may be written as

$$\frac{\partial}{\partial u}(h_2 h_3 D_u)du\,dv\,dw$$

and the other two corresponding expressions obtained by a simple permutation of the subscripts and of u, v, and w. Thus the total flux leaving the differential volume is

$$\left[\frac{\partial}{\partial u}(h_2 h_3 D_u) + \frac{\partial}{\partial v}(h_3 h_1 D_v) + \frac{\partial}{\partial w}(h_1 h_2 D_w)\right] du\,dv\,dw$$

and the divergence of $\mathbf{D}$ is found by dividing by the differential volume

$$\nabla \cdot \mathbf{D} = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial u}(h_2 h_3 D_u) + \frac{\partial}{\partial v}(h_3 h_1 D_v) + \frac{\partial}{\partial w}(h_1 h_2 D_w)\right] \quad\text{(A.2)}$$

The components of the gradient of a scalar V may be obtained (following the methods of Section 4.6) by expressing the total differential of V,

$$dV = \frac{\partial V}{\partial u}du + \frac{\partial V}{\partial v}dv + \frac{\partial V}{\partial w}dw$$

---

[1] The variables and scale factors are given for nine orthogonal coordinate systems on pp. 50–59 in J. A. Stratton, *Electromagnetic Theory*. New York: McGraw-Hill, 1941. Each system is also described briefly.

in terms of the component differential lengths, $h_1 du$, $h_2 dv$, and $h_3 dw$,

$$dV = \frac{1}{h_1}\frac{\partial V}{\partial u}h_1 du + \frac{1}{h_2}\frac{\partial V}{\partial v}h_2 dv + \frac{1}{h_3}\frac{\partial V}{\partial w}h_3 dw$$

Then, because

$$d\mathbf{L} = h_1 du\,\mathbf{a}_u + h_2 dv\,\mathbf{a}_v + h_3 dw\,\mathbf{a}_w \qquad \text{and} \qquad dV = \nabla V \cdot d\mathbf{L}$$

we see that

$$\nabla V = \frac{1}{h_1}\frac{\partial V}{\partial u}\mathbf{a}_u + \frac{1}{h_2}\frac{\partial V}{\partial v}\mathbf{a}_v + \frac{1}{h_3}\frac{\partial V}{\partial w}\mathbf{a}_w \qquad (A.3)$$

The components of the curl of a vector $\mathbf{H}$ are obtained by considering a differential path first in a $u = $ constant surface and finding the circulation of $\mathbf{H}$ about that path, as discussed for rectangular coordinates in Section 7.3. The contribution along the segment in the $\mathbf{a}_v$ direction is

$$H_{v0}h_2 dv - \frac{1}{2}\frac{\partial}{\partial w}(H_v h_2 dv)dw$$

and that from the oppositely directed segment is

$$-H_{v0}h_2 dv - \frac{1}{2}\frac{\partial}{\partial w}(H_v h_2 dv)dw$$

The sum of these two parts is

$$-\frac{\partial}{\partial w}(H_v h_2 dv)dw$$

or

$$-\frac{\partial}{\partial w}(h_2 H_v)dv\,dw$$

and the sum of the contributions from the other two sides of the path is

$$\frac{\partial}{\partial v}(h_3 H_w)dv\,dw$$

Adding these two terms and dividing the sum by the enclosed area, $h_2 h_3 dv\,dw$, we see that the $\mathbf{a}_u$ component of curl $\mathbf{H}$ is

$$(\nabla \times \mathbf{H})_u = \frac{1}{h_2 h_3}\left[\frac{\partial}{\partial v}(h_3 H_w) - \frac{\partial}{\partial w}(h_2 H_v)\right]$$

and the other two components may be obtained by cyclic permutation. The result is expressible as a determinant,

$$\nabla \times \mathbf{H} = \begin{vmatrix} \dfrac{\mathbf{a}_u}{h_2 h_3} & \dfrac{\mathbf{a}_v}{h_3 h_1} & \dfrac{\mathbf{a}_w}{h_1 h_2} \\[2mm] \dfrac{\partial}{\partial u} & \dfrac{\partial}{\partial v} & \dfrac{\partial}{\partial w} \\[2mm] h_1 H_u & h_2 H_v & h_3 H_w \end{vmatrix} \qquad (A.4)$$

The Laplacian of a scalar is found by using (A.2) and (A.3):

$$\nabla^2 V = \nabla \cdot \nabla V = \frac{1}{h_1 h_2 h_3} \left[\frac{\partial}{\partial u} \left(\frac{h_2 h_3}{h_1} \frac{\partial v}{\partial u} \right) + \frac{\partial}{\partial v} \left(\frac{h_3 h_1}{h_2} \frac{\partial V}{\partial v} \right) \right.$$

$$\left. + \frac{\partial}{\partial w} \left(\frac{h_1 h_2}{h_3} \frac{\partial V}{\partial w} \right) \right] \tag{A.5}$$

Equations (A.2) to (A.5) may be used to find the divergence, gradient, curl, and Laplacian in any orthogonal coordinate system for which h_1, h_2, and h_3 are known.

Expressions for $\nabla \cdot \mathbf{D}$, ∇V, $\nabla \times \mathbf{H}$, and $\nabla^2 V$ are given in rectangular, circular cylindrical, and spherical coordinate systems inside the back cover.

A.3   VECTOR IDENTITIES

The vector identities that follow may be proved by expansion in rectangular (or general curvilinear) coordinates. The first two identities involve the scalar and vector triple products, the next three are concerned with operations on sums, the following three apply to operations when the argument is multiplied by a scalar function, the next three apply to operations on scalar or vector products, and the last four concern the second-order operations.

$$(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C} \equiv (\mathbf{B} \times \mathbf{C}) \cdot \mathbf{A} \equiv (\mathbf{C} \times \mathbf{A}) \cdot \mathbf{B} \tag{A.6}$$

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) \equiv (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C} \tag{A.7}$$

$$\nabla \cdot (\mathbf{A} + \mathbf{B}) \equiv \nabla \cdot \mathbf{A} + \nabla \cdot \mathbf{B} \tag{A.8}$$

$$\nabla(V + W) \equiv \nabla V + \nabla W \tag{A.9}$$

$$\nabla \times (\mathbf{A} + \mathbf{B}) \equiv \nabla \times \mathbf{A} + \nabla \times \mathbf{B} \tag{A.10}$$

$$\nabla \cdot (V\mathbf{A}) \equiv \mathbf{A} \cdot \nabla V + V \nabla \cdot \mathbf{A} \tag{A.11}$$

$$\nabla(VW) \equiv V \nabla W + W \nabla V \tag{A.12}$$

$$\nabla \times (V\mathbf{A}) \equiv \nabla V \times \mathbf{A} + V \nabla \times \mathbf{A} \tag{A.13}$$

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) \equiv \mathbf{B} \cdot \nabla \times \mathbf{A} - \mathbf{A} \cdot \nabla \times \mathbf{B} \tag{A.14}$$

$$\nabla(\mathbf{A} \cdot \mathbf{B}) \equiv (\mathbf{A} \cdot \nabla)\mathbf{B} + (\mathbf{B} \cdot \nabla)\mathbf{A} + \mathbf{A} \times (\nabla \times \mathbf{B})$$

$$+ \mathbf{B} \times (\nabla \times \mathbf{A}) \tag{A.15}$$

$$\nabla \times (\mathbf{A} \times \mathbf{B}) \equiv \mathbf{A}\nabla \cdot \mathbf{B} - \mathbf{B}\nabla \cdot \mathbf{A} + (\mathbf{B} \cdot \nabla)\mathbf{A} - (\mathbf{A} \cdot \nabla)\mathbf{B} \tag{A.16}$$

$$\nabla \cdot \nabla V \equiv \nabla^2 V \tag{A.17}$$

$$\nabla \cdot \nabla \times \mathbf{A} \equiv 0 \tag{A.18}$$

$$\nabla \times \nabla V \equiv 0 \tag{A.19}$$

$$\nabla \times \nabla \times \mathbf{A} \equiv \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \tag{A.20}$$

Units

We describe first the International System (abbreviated SI, for Système International d'Unités), which is used in this book and is now standard in electrical engineering and much of physics. It has also been officially adopted as the international system of units by many countries, including the United States.[1]

The fundamental unit of length is the meter, which was defined in the latter part of the nineteenth century as the distance between two marks on a certain platinum-iridium bar. The definition was improved in 1960 by relating the meter to the wavelength of the radiation emitted by the rare gas isotope krypton 86 under certain specified conditions. This so-called krypton meter was accurate to four parts per billion, a value leading to negligible uncertainties in constructing skyscrapers or building highways, but capable of causing an error greater than one meter in determining the distance to the moon. The meter was redefined in 1983 in terms of the velocity of light. At that time the velocity of light was specified to be an auxiliary constant with an *exact* value of $299,792,458$ meters per second. As a result, the latest definition of the meter is the distance light travels in a vacuum in $1/299,792,458$ of a second. If greater accuracy is achieved in measuring c, that value will remain $299,792,458$ m/s, but the length of the meter will change.

It is evident that our definition of the meter is expressed in terms of the second, the fundamental unit of time. The second is defined as $9,192,631,770$ periods of the transition frequency between the hyperfine levels $F = 4$, $m_F = 0$, and $F = 3$, $m_F = 0$ of the ground state $^2s_{1/2}$ of the atom of cesium 133, unperturbed by external

---

[1] The International System of Units was adopted by the Eleventh General Conference on Weights and Measures in Paris in 1960, and it was officially adopted for scientific usage by the National Bureau of Standards in 1964. It is a metric system, which interestingly enough is the only system which has ever received specific sanction from Congress. This occurred first in 1866 and then again in 1975 with the Metric Conversion Act, which provides for "voluntary conversion" to the metric system. No specific time was specified, however, and we can assume that it will still be a few years before the bathroom scale reads mass in kilograms and Miss America is a 90–60–90.

fields. This definition of the second, complex though it may be, permits time to be measured with an accuracy better than one part in 10^{13}.

The standard mass of one kilogram is defined as the mass of an international standard in the form of a platinum-iridium cylinder at the International Bureau of Weights and Measures at Sèvres, France.

The unit of temperature is the kelvin, defined by placing the triple-point temperature of water at 273.16 kelvins.

A fifth unit is the candela, defined as the luminous intensity of an omnidirectional radiator at the freezing temperature of platinum (2042 K) having an area of $1/600,000$ square meter and under a pressure of $101,325$ newtons per square meter.

The last of the fundamental units is the ampere. Before explicitly defining the ampere, we must first define the newton. It is defined in terms of the other fundamental units from Newton's third law as the force required to produce an acceleration of one meter per second per second on a one-kilogram mass. We now may define the ampere as that constant current, flowing in opposite directions in two straight parallel conductors of infinite length and negligible cross section, separated one meter in vacuum, that produces a repulsive force of 2×10^{-7} newton per meter length between the two conductors. The force between the two parallel conductors is known to be

$$F = \mu_0 \frac{I^2}{2\pi d}$$

and thus

$$2 \times 10^{-7} = \mu_0 \frac{1}{2\pi}$$

or

$$\mu_0 = 4\pi \times 10^{-7} \qquad (\text{kg} \cdot \text{m/A}^2 \cdot \text{s}^2, \text{ or H/m})$$

We thus find that our definition of the ampere has been formulated in such a way as to assign an exact, simple, numerical value to the permeability of free space.

Returning to the International System, the units in which the other electric and magnetic quantities are measured are given in the body of the text at the time each quantity is defined, and all of them can be related to the basic units already defined. For example, our work with the plane wave in Chapter 11 shows that the velocity with which an electromagnetic wave propagates in free space is

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$$

and thus

$$\epsilon_0 = \frac{1}{\mu_0 c^2} = \frac{1}{4\pi\,10^{-7} c^2} = 8.854\,187\,817 \times 10^{-12} \text{ F/m}$$

It is evident that the numerical value of ϵ_0 depends upon the defined value of the velocity of light in vacuum, $299,792,458$ m/s.

The units are also given in Table B.1 for easy reference. They are listed in the same order in which they are defined in the text.

Table B.1 Names and units of the electric and magnetic quantities in the International System (in the order in which they appear in the text)

Symbol	Name	Unit	Abbreviation
ν	Velocity	meter/second	m/s
F	Force	newton	N
Q	Charge	coulomb	C
r, R	Distance	meter	m
ϵ_0, ϵ	Permittivity	farad/meter	F/m
E	Electric field intensity	volt/meter	V/m
ρ_v	Volume charge density	coulomb/meter^3	C/m^3
ν	Volume	meter^3	m^3
ρ_L	Linear charge density	coulomb/meter	C/m
ρ_S	Surface charge density	coulomb/meter^2	C/m^2
Ψ	Electric flux	coulomb	C
D	Electric flux density	coulomb/meter^2	C/m^2
S	Area	meter^2	m^2
W	Work, energy	joule	J
L	Length	meter	m
V	Potential	volt	V
p	Dipole moment	coulomb-meter	C·m
I	Current	ampere	A
J	Current density	ampere/meter^2	A/m^2
μ_e, μ_h	Mobility	meter^2/volt-second	m^2/V·s
e	Electronic charge	coulomb	C
σ	Conductivity	siemens/meter	S/m
R	Resistance	ohm	Ω
P	Polarization	coulomb/meter^2	C/m^2
$\chi_{e,m}$	Susceptibility		
C	Capacitance	farad	F
R_s	Sheet resistance	ohm per square	Ω
H	Magnetic field intensity	ampere/meter	A/m
K	Surface current density	ampere/meter	A/m
B	Magnetic flux density	tesla (or weber/meter^2)	T (or Wb/m^2)
μ_0, μ	Permeability	henry/meter	H/m
Φ	Magnetic flux	weber	Wb
V_m	Magnetic scalar potential	ampere	A
A	Vector magnetic potential	weber/meter	Wb/m
T	Torque	newton-meter	N·m
m	Magnetic moment	ampere-meter^2	A·m^2
M	Magnetization	ampere/meter	A/m
$\mathcal{R}$	Reluctance	ampere-turn/weber	A·t/Wb
L	Inductance	henry	H
M	Mutual inductance	henry	H

(*Continued*)

Table B.1 *(Continued)*

Symbol	Name	Unit	Abbreviation
ω	Radian frequency	radian/second	rad/s
c	Velocity of light	meter/second	m/s
λ	Wavelength	meter	m
η	Intrinsic impedance	ohm	Ω
k	Wave number	meter^{-1}	m^{-1}
α	Attenuation constant	neper/meter	Np/m
β	Phase constant	radian/meter	rad/m
f	Frequency	hertz	Hz
S	Poynting vector	watt/meter^2	W/m^2
P	Power	watt	W
δ	Skin depth	meter	m
Γ	Reflection coefficient		
s	Standing wave ratio		
γ	Propagation constant	meter^{-1}	m^{-1}
G	Conductance	siemen	S
Z	Impedance	ohm	Ω
Y	Admittance	siemen	S
Q	Quality factor		

Finally, other systems of units have been used in electricity and magnetism. In the electrostatic system of units (esu), Coulomb's law is written for free space,

$$F = \frac{Q_1 Q_2}{R^2} \qquad \text{(esu)}$$

The permittivity of free space is assigned the value of unity. The gram and centimeter are the fundamental units of mass and distance, and the esu system is therefore a cgs system. Units bearing the prefix *stat-* belong to the electrostatic system of units.

In a similar manner, the electromagnetic system of units (emu) is based on Coulomb's law for magnetic poles, and the permeability of free space is unity. The prefix *ab-* identifies emu units. When electric quantities are expressed in esu units, magnetic quantities are expressed in emu units, and both appear in the same equation (such as Maxwell's curl equations), the velocity of light appears explicitly. This follows from noting that in esu $\epsilon_0 = 1$, but $\mu_0 \epsilon_0 = 1/c^2$, and therefore $\mu_0 = 1/c^2$, and in emu $\mu_0 = 1$, and hence $\epsilon_0 = 1/c^2$. Thus, in this intermixed system known as the Gaussian system of units,

$$\nabla \times \mathbf{H} = 4\pi \mathbf{J} + \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} \qquad \text{(Gaussian)}$$

Other systems include the factor 4π explicitly in Coulomb's law, and it then does not appear in Maxwell's equations. When this is done, the system is said to be rationalized. Hence the Gaussian system is an unrationalized cgs system (when rationalized it is known as the Heaviside–Lorentz system), and the International System we have used throughout this book is a rationalized mks system.

Table B.2   Conversion of International to Gaussian and other units
(use $c = 2.997\ 924\ 58 \times 10^8$)

Quantity	1 mks unit	= Gaussian units	= Other units
d	1 m	10^2 cm	39.37 in.
F	1 N	10^5 dyne	0.2248 lb$_f$
W	1 J	10^7 erg	0.7376 ft-lb$_f$
Q	1 C	$10c$ statC	0.1 abC
ρ_v	1 C/m^3	$10^{-5}c$ statC/cm^3	10^{-7} abC/cm^3
D	1 C/m^2	$4\pi\,10^{-3}c$ (esu)	$4\pi\,10^{-5}$ (emu)
E	1 V/m	$10^4/c$ statV/cm	10^6 abV/cm
V	1 V	$10^6/c$ statV	10^8 abV
I	1 A	0.1 abA	$10c$ statA
H	1 A/m	$4\pi\,10^{-3}$ oersted	$0.4\pi c$ (esu)
V_m	1 A·t	0.4π gilbert	$40\pi c$ (esu)
B	1 T	10^4 gauss	$100/c$ (esu)
Φ	1 Wb	10^8 maxwell	$10^6/c$ (esu)
A	1 Wb/m	10^6 maxwell/cm	
R	1 Ω	10^9 abΩ	$10^5/c^2$ statΩ
L	1 H	10^9 abH	$10^5/c^2$ statH
C	1 F	$10^{-5}c^2$ statF	10^{-9} abF
σ	1 S/m	10^{-11} abS/cm	$10^{-7}c^2$ statS/cm
μ	1 H/m	$10^7/4\pi$ (emu)	$10^3/4\pi c^2$ (esu)
ϵ	1 F/m	$4\pi\,10^{-7}c^2$ (esu)	$4\pi\,10^{-11}$ (emu)

   Table B.2 gives the conversion factors between the more important units of the International System (or rationalized mks system) and the Gaussian system, and several other assorted units.

   Table B.3 lists the prefixes used with any of the SI units, their abbreviations, and the power of ten each represents. Those checked are widely used. Both the prefixes and their abbreviations are written without hyphens, and therefore 10^{-6} F $=$ 1 microfarad $= 1\mu$F $= 1000$ nanofarads $= 1000$ nF, and so forth.

Table B.3   Standard prefixes used with SI units

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

C

A P P E N D I X

Material Constants

Table C.1 lists typical values of the relative permittivity ϵ_r' or dielectric constant for common insulating and dielectric materials, along with representative values for the loss tangent. The values should only be considered representative for each material, and they apply to normal temperature and humidity conditions and to very low audio frequencies. Most of them have been taken from *Reference Data for Radio Engineers*,[1] *The Standard Handbook for Electrical Engineers*,[2] and von Hippel,[3] and these volumes may be referred to for further information on these and other materials.

Table C.2 gives the conductivity for a number of metallic conductors, for a few insulating materials, and for several other materials of general interest. The values have been taken from the references listed previously, and they apply at zero frequency and at room temperature. The listing is in the order of decreasing conductivity.

Some representative values of the relative permeability for various diamagnetic, paramagnetic, ferrimagnetic, and ferromagnetic materials are listed in Table C.3. They have been extracted from the references listed previously, and the data for the ferromagnetic materials is only valid for very low magnetic flux densities. Maximum permeabilities may be an order of magnitude higher.

Values are given in Table C.4 for the charge and rest mass of an electron, the permittivity and permeability of free space, and the velocity of light.[4]

---

[1] International Telephone and Telegraph Co., Inc.: *Reference Data for Radio Engineers*, 7th ed., Howard W. Sams & Co., Indianapolis, IN, 1985.

[2] See References for Chapter 5.

[3] von Hippel, A. R. *Dielectric Materials and Applications*. Cambridge, Mass. and New York: The Technology Press of the Massachusetts Institute of Technology and John Wiley & Sons, 1954.

[4] Cohen, E. R., and B. N. Taylor. *The 1986 Adjustment of the Fundamental Physical Constants.* Elmsford, N.Y.: Pergamon Press, 1986.

Table C.1 ϵ_r' and ϵ''/ϵ'

Material	ϵ_r'	ϵ''/ϵ'
Air	1.0005	
Alcohol, ethyl	25	0.1
Aluminum oxide	8.8	0.000 6
Amber	2.7	0.002
Bakelite	4.74	0.022
Barium titanate	1200	0.013
Carbon dioxide	1.001	
Ferrite (NiZn)	12.4	0.000 25
Germanium	16	
Glass	4–7	0.002
Ice	4.2	0.05
Mica	5.4	0.000 6
Neoprene	6.6	0.011
Nylon	3.5	0.02
Paper	3	0.008
Plexiglas	3.45	0.03
Polyethylene	2.26	0.000 2
Polypropylene	2.25	0.000 3
Polystyrene	2.56	0.000 05
Porcelain (dry process)	6	0.014
Pyranol	4.4	0.000 5
Pyrex glass	4	0.000 6
Quartz (fused)	3.8	0.000 75
Rubber	2.5–3	0.002
Silica or SiO_2 (fused)	3.8	0.000 75
Silicon	11.8	
Snow	3.3	0.5
Sodium chloride	5.9	0.000 1
Soil (dry)	2.8	0.05
Steatite	5.8	0.003
Styrofoam	1.03	0.000 1
Teflon	2.1	0.000 3
Titanium dioxide	100	0.001 5
Water (distilled)	80	0.04
Water (sea)		4
Water (dehydrated)	1	0
Wood (dry)	1.5–4	0.01

Table C.2 σ

Material	σ, S/m	Material	σ, S/m
Silver	6.17×10^7	Nichrome	0.1×10^7
Copper	5.80×10^7	Graphite	7×10^4
Gold	4.10×10^7	Silicon	2300
Aluminum	3.82×10^7	Ferrite (typical)	100
Tungsten	1.82×10^7	Water (sea)	5
Zinc	1.67×10^7	Limestone	10^{-2}
Brass	1.5×10^7	Clay	5×10^{-3}
Nickel	1.45×10^7	Water (fresh)	10^{-3}
Iron	1.03×10^7	Water (distilled)	10^{-4}
Phosphor bronze	1×10^7	Soil (sandy)	10^{-5}
Solder	0.7×10^7	Granite	10^{-6}
Carbon steel	0.6×10^7	Marble	10^{-8}
German silver	0.3×10^7	Bakelite	10^{-9}
Manganin	0.227×10^7	Porcelain (dry process)	10^{-10}
Constantan	0.226×10^7	Diamond	2×10^{-13}
Germanium	0.22×10^7	Polystyrene	10^{-16}
Stainless steel	0.11×10^7	Quartz	10^{-17}

Table C.3 μ_r

Material	μ_r	Material	μ_r
Bismuth	0.999 998 6	Powdered iron	100
Paraffin	0.999 999 42	Machine steel	300
Wood	0.999 999 5	Ferrite (typical)	1000
Silver	0.999 999 81	Permalloy 45	2500
Aluminum	1.000 000 65	Transformer iron	3000
Beryllium	1.000 000 79	Silicon iron	3500
Nickel chloride	1.000 04	Iron (pure)	4000
Manganese sulfate	1.000 1	Mumetal	20 000
Nickel	50	Sendust	30 000
Cast iron	60	Supermalloy	100 000
Cobalt	60		

Table C.4   Physical constants

Quantity	Value
Electron charge	$e = (1.602\,177\,33 \pm 0.000\,000\,46) \times 10^{-19}$ C
Electron mass	$m = (9.109\,389\,7 \pm 0.000\,005\,4) \times 10^{-31}$ kg
Permittivity of free space	$\epsilon_0 = 8.854\,187\,817 \times 10^{-12}$ F/m
Permeability of free space	$\mu_0 = 4\pi 10^{-7}$ H/m
Velocity of light	$c = 2.997\,924\,58 \times 10^8$ m/s

The Uniqueness Theorem

Let us assume that we have two solutions of Laplace's equation, V_1 and V_2, both general functions of the coordinates used. Therefore

$$\nabla^2 V_1 = 0$$

and

$$\nabla^2 V_2 = 0$$

from which

$$\nabla^2(V_1 - V_2) = 0$$

Each solution must also satisfy the boundary conditions, and if we represent the given potential values on the boundaries by V_b, then the value of V_1 on the boundary V_{1b} and the value of V_2 on the boundary V_{2b} must both be identical to V_b,

$$V_{1b} = V_{2b} = V_b$$

or

$$V_{1b} - V_{2b} = 0$$

In Section 4.8, Eq. (43), we made use of a vector identity,

$$\nabla \cdot (V\mathbf{D}) \equiv V(\nabla \cdot \mathbf{D}) + \mathbf{D} \cdot (\nabla V)$$

which holds for any scalar V and any vector $\mathbf{D}$. For the present application we shall select $V_1 - V_2$ as the scalar and $\nabla(V_1 - V_2)$ as the vector, giving

$$\nabla \cdot [(V_1 - V_2)\nabla(V_1 - V_2)] \equiv (V_1 - V_2)[\nabla \cdot \nabla(V_1 - V_2)]$$
$$+ \nabla(V_1 - V_2) \cdot \nabla(V_1 - V_2)$$

which we shall integrate throughout the volume *enclosed* by the boundary surfaces specified:

$$\int_{\text{vol}} \nabla \cdot [(V_1 - V_2)\nabla(V_1 - V_2)]\, dv$$

$$\equiv \int_{\text{vol}} (V_1 - V_2)[\nabla \cdot \nabla(V_1 - V_2)]\, dv + \int_{\text{vol}} [\nabla(V_1 - V_2)]^2\, dv \qquad (D.1)$$

The divergence theorem allows us to replace the volume integral on the left side of the equation with the closed surface integral over the surface surrounding the volume. This surface consists of the boundaries already specified on which $V_{1b} = V_{2b}$, and therefore

$$\int_{\text{vol}} \nabla \cdot [(V_1 - V_2)\nabla(V_1 - V_2)]\, dv = \oint_S [(V_{1b} - V_{2b})\nabla(V_{1b} - V_{2b})] \cdot d\mathbf{S} = 0$$

One of the factors of the first integral on the right side of (D.1) is $\nabla \cdot \nabla(V_1 - V_2)$, or $\nabla^2(V_1 - V_2)$, which is zero by hypothesis, and therefore that integral is zero. Hence the remaining volume integral must be zero:

$$\int_{\text{vol}} [\nabla(V_1 - V_2)]^2\, dv = 0$$

There are two reasons why an integral may be zero: either the integrand (the quantity under the integral sign) is everywhere zero, or the integrand is positive in some regions and negative in others, and the contributions cancel algebraically. In this case the first reason must hold because $[\nabla(V_1 - V_2)]^2$ cannot be negative. Therefore

$$[\nabla(V_1 - V_2)]^2 = 0$$

and

$$\nabla(V_1 - V_2) = 0$$

Finally, if the gradient of $V_1 - V_2$ is everywhere zero, then $V_1 - V_2$ cannot change with any coordinates, and

$$V_1 - V_2 = \text{constant}$$

If we can show that this constant is zero, we shall have accomplished our proof. The constant is easily evaluated by considering a point on the boundary. Here $V_1 - V_2 = V_{1b} - V_{2b} = 0$, and we see that the constant is indeed zero, and therefore

$$V_1 = V_2$$

giving two identical solutions.

The uniqueness theorem also applies to Poisson's equation, for if $\nabla^2 V_1 = -\rho_v/\epsilon$ and $\nabla^2 V_2 = -\rho_v/\epsilon$, then $\nabla^2(V_1 - V_2) = 0$ as before. Boundary conditions still require that $V_{1b} - V_{2b} = 0$, and the proof is identical from this point.

This constitutes the proof of the uniqueness theorem. Viewed as the answer to a question, "How do two solutions of Laplace's or Poisson's equation compare if they both satisfy the same boundary conditions?" the uniqueness theorem should please us by its ensurance that the answers are identical. Once we can find any method of solving Laplace's or Poisson's equation subject to given boundary conditions, we have solved our problem once and for all. No other method can ever give a different answer.

Origins of the Complex Permittivity

As we learned in Chapter 5, a dielectric can be modeled as an arrangement of atoms and molecules in free space, which can be polarized by an electric field. The field forces positive and negative bound charges to separate against their Coulomb attractive forces, thus producing an array of microscopic dipoles. The molecules can be arranged in an ordered and predictable manner (such as in a crystal) or may exhibit random positioning and orientation, as would occur in an amorphous material or a liquid. The molecules may or may not exhibit permanent dipole moments (existing before the field is applied), and if they do, they will usually have random orientations throughout the material volume. As discussed in Section 5.7, the displacement of charges in a regular manner, as induced by an electric field, gives rise to a macroscopic polarization, $\mathbf{P}$, defined as the dipole moment per unit volume:

$$\mathbf{P} = \lim_{\Delta v \to 0} \frac{1}{\Delta v} \sum_{i=1}^{N \Delta v} \mathbf{p_i} \tag{E.1}$$

where N is the number of dipoles per unit volume and $\mathbf{p_i}$ is the dipole moment of the ith atom or molecule, found through

$$\mathbf{p_i} = Q_i \mathbf{d}_i \tag{E.2}$$

Q_i is the positive one of the two bound charges composing dipole i, and $\mathbf{d}_i$ is the distance between charges, expressed as a vector from the negative to the positive charge. Again, borrowing from Section 5.7, the electric field and the polarization are related through

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E} \tag{E.3}$$

where the electric susceptibility, χ_e, forms the more interesting part of the dielectric constant:

$$\epsilon_r = 1 + \chi_e \tag{E.4}$$

Therefore, to understand the nature of ϵ_r, we need to understand χ_e, which in turn means that we need to explore the behavior of the polarization, **P**.

Here, we consider the added complications of how the dipoles respond to a time-harmonic field that propagates as a wave through the material. The result of applying such a forcing function is that *oscillating* dipole moments are set up, and *these in turn establish a polarization wave that propagates through the material*. The effect is to produce a polarization function, $\mathbf{P}(z, t)$, having the same functional form as the driving field, $\mathbf{E}(z, t)$. The molecules themselves do not move through the material, but their oscillating dipole moments collectively exhibit wave motion, just as waves in pools of water are formed by the up and down motion of the water. From here, the description of the process gets complicated and in many ways beyond the scope of our present discussion. We can form a basic qualitative understanding, however, by considering the classical description of the process, which is that the dipoles, once oscillating, behave as microscopic antennas, re-radiating fields that in turn co-propagate with the applied field. Depending on the frequency, there will be some phase difference between the incident field and the radiated field at a given dipole location. This results in a net field (formed through the superposition of the two) that now interacts with the next dipole. Radiation from this dipole adds to the previous field as before, and the process repeats from dipole to dipole. The accumulated phase shifts at each location are manifested as a net slowing down of the phase velocity of the resultant wave. Attenuation of the field may also occur which, in this classical model, can be accounted for by partial phase cancellation between incident and radiated fields.

For our classical description, we use the Lorentz model, in which the medium is considered as an ensemble of identical fixed electron oscillators in free space. The Coulomb binding forces on the electrons are modeled by springs that attach the electrons to the positive nuclei. We consider electrons for simplicity, but similar models can be used for any bound charged particle. Figure E.1 shows a single oscillator, located at position z in the material, and oriented along x. A uniform plane wave, assumed to be linearly polarized along x, propagates through the material in the z direction. The electric field in the wave displaces the electron of the oscillator in the x direction through a distance represented by the vector **d**; a dipole moment is thus established,

$$\mathbf{p}(z, t) = -e\mathbf{d}(z, t) \tag{E.5}$$
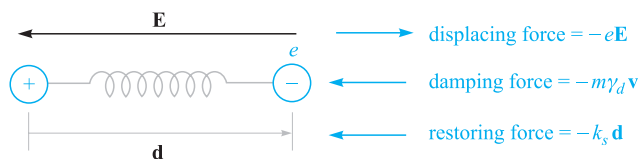
Figure E.1 Atomic dipole model, with the Coulomb force between positive and negative charge modeled by that of a spring having spring constant k_s. An applied electric field displaces the electron through distance d, resulting in dipole moment $\mathbf{p} = -e\mathbf{d}$.

where the electron charge, e, is treated as a positive quantity. The applied force is

$$\mathbf{F}_a(z, t) = -e\mathbf{E}(z, t) \tag{E.6}$$

We need to remember that $\mathbf{E}(z, t)$ at a given oscillator location is the *net* field, composed of the original applied field plus the radiated fields from all other oscillators. The relative phasing between oscillators is precisely determined by the spatial and temporal behavior of $\mathbf{E}(z, t)$.

The restoring force on the electron, $\mathbf{F}_r$, is that produced by the spring which is assumed to obey Hooke's law:

$$\mathbf{F}_r(z, t) = -k_s\mathbf{d}(z, t) \tag{E.7}$$

where k_s is the spring constant (not to be confused with the propagation constant). If the field is turned off, the electron is released and will oscillate about the nucleus at the *resonant frequency,* given by

$$\omega_0 = \sqrt{k_s/m} \tag{E.8}$$

where m is the mass of the electron. The oscillation, however, will be damped since the electron will experience forces and collisions from neighboring oscillators. We model these as a velocity-dependent damping force:

$$\mathbf{F}_d(z, t) = -m\gamma_d\mathbf{v}(z, t) \tag{E.9}$$

where $\mathbf{v}(z, t)$ is the electron velocity. Associated with this damping is the *dephasing* process among the electron oscillators in the system. Their relative phasing, once fixed by the applied sinusoidal field, is destroyed through collisions and dies away exponentially until a state of totally random phase exists between oscillators. The $1/e$ point in this process occurs at the *dephasing time* of the system, which is inversely proportional to the damping coefficient, γ_d (in fact it is $2/\gamma_d$). We are, of course, driving this damped resonant system with an electric field at frequency ω. We can therefore expect the response of the oscillators, measured through the magnitude of $\mathbf{d}$, to be frequency-dependent in much the same way as an RLC circuit is when driven by a sinusoidal voltage.

We can now use Newton's second law and write down the forces acting on the single oscillator of Figure E.1. To simplify the process a little we can use the complex form of the electric field:

$$\mathbf{E}_c = \mathbf{E}_0 e^{-jkz} e^{j\omega t} \tag{E.10}$$

Defining $\mathbf{a}$ as the acceleration vector of the electron, we have

$$m\mathbf{a} = \mathbf{F}_a + \mathbf{F}_r + \mathbf{F}_d$$

or

$$m\frac{\partial^2 \mathbf{d}_c}{\partial t^2} + m\gamma_d \frac{\partial \mathbf{d}_c}{\partial t} + k_s\mathbf{d}_c = -e\mathbf{E}_c \tag{E.11}$$

Note that since we are driving the system with the complex field, $\mathbf{E}_c$, we anticipate a displacement wave, $\mathbf{d}_c$, of the form

$$\mathbf{d}_c = \mathbf{d}_0 e^{-jkz} e^{-j\omega t} \tag{E.12}$$

With the waves in this form, time differentiation produces a factor of $j\omega$. Consequently (E.11) can be simplified and rewritten in phasor form:

$$-\omega^2 \mathbf{d}_s + j\omega\gamma_d \mathbf{d}_s + \omega_0^2 \mathbf{d}_s = -\frac{e}{m}\mathbf{E}_s \tag{E.13}$$

where (E.4) has been used. We now solve (E.13) for $\mathbf{d}_s$, obtaining

$$\mathbf{d}_s = \frac{-(e/m)\mathbf{E}_s}{(\omega_0^2 - \omega^2) + j\omega\gamma_d} \tag{E.14}$$

The dipole moment associated with displacement $\mathbf{d}_s$ is

$$\mathbf{p}_s = -e\mathbf{d}_s \tag{E.15}$$

The polarization of the medium is then found, assuming that all dipoles are identical. Eq. (E.1) thus becomes

$$\mathbf{P}_s = N\mathbf{p}_s$$

which, when using (E.14) and (E.15), becomes

$$\mathbf{P}_s = \frac{Ne^2/m}{(\omega_0^2 - \omega^2) + j\omega\gamma_d}\mathbf{E}_s \tag{E.16}$$

Now, using (E.3) we identify the susceptibility associated with the resonance as

$$\chi_{res} = \frac{Ne^2}{\epsilon_0 m}\frac{1}{(\omega_0^2 - \omega^2) + j\omega\gamma_d} = \chi'_{res} - j\chi''_{res} \tag{E.17}$$

The real and imaginary parts of the permittivity are now found through the real and imaginary parts of χ_{res}: Knowing that

$$\epsilon = \epsilon_0(1 + \chi_{res}) = \epsilon' - j\epsilon''$$

we find

$$\epsilon' = \epsilon_0(1 + \chi'_{res}) \tag{E.18}$$

and

$$\epsilon'' = \epsilon_0\chi''_{res} \tag{E.19}$$

The preceding expressions can now be used in Eqs. (44) and (45) in Chapter 11 to evaluate the attenuation coefficient, α, and phase constant, β, for the plane wave as it propagates through our resonant medium.

The real and imaginary parts of χ_{res} as functions of frequency are shown in Figure E.2 for the special case in which $\omega \doteq \omega_0$. Eq. (E.17) in this instance becomes

$$\chi_{res} \doteq -\frac{Ne^2}{\epsilon_0 m\omega_0\gamma_d}\left(\frac{j + \delta_n}{1 + \delta_n^2}\right) \tag{E.20}$$

where the *normalized detuning* parameter, δ_n, is

$$\frac{2}{\gamma_d}(\omega - \omega_0) \tag{E.21}$$
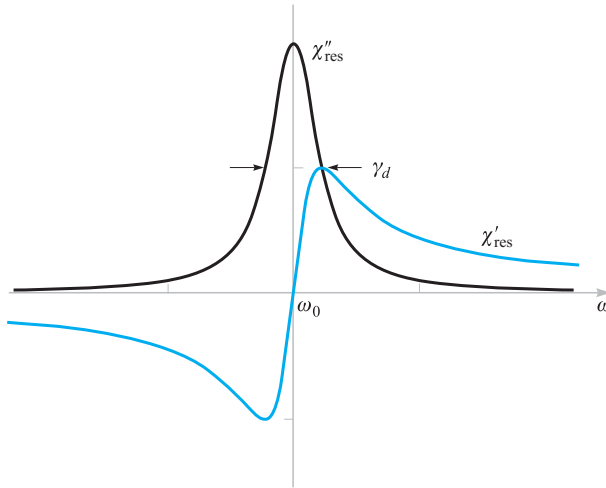
Figure E.2 Plots of the real and imaginary parts of the resonant susceptibility, χ_{res}, as given by Eq. (E.20). The full-width at half-maximum of the imaginary part, χ''_{res}, is equal to the damping coefficient, γ_d.

Key features to note in Figure E.2 include the symmetric χ''_e function, whose full-width at its half-maximum amplitude is γ_d. Near the resonant frequency, where χ''_{res} maximizes, wave attenuation maximizes as seen from Eq. (44), Chapter 11. Additionally, we see that away from resonance, attenuation is relatively weak, and the material becomes transparent. As Figure E.2 shows, there is still significant variation of χ'_{res} with frequency away from resonance, which leads to a frequency-dependent refractive index; this is expressed approximately as

$$n \doteq \sqrt{1 + \chi'_{res}} \qquad \text{(away from resonance)} \qquad \text{(E.22)}$$

This frequency-dependent n, arising from the material resonance, leads to phase and group velocities that also depend on frequency. Thus, group dispersion, leading to pulse-broadening effects as discussed in Chapter 12, can be directly attributable to material resonances.

Somewhat surprisingly, the classical "spring model" described here can provide very accurate predictions on dielectric constant behavior with frequency (particularly off-resonance) and can be used to a certain extent to model absorption properties. The model is insufficient, however, when attempting to describe the more salient features of materials; specifically, it assumes that the oscillating electron can assume any one of a continuum of energy states, when, in fact, energy states in any atomic system are quantized. As a result, the important effects arising from transitions between discrete energy levels, such as spontaneous and stimulated absorption and emission, are not included in our classical spring system. Quantum mechanical models must be used
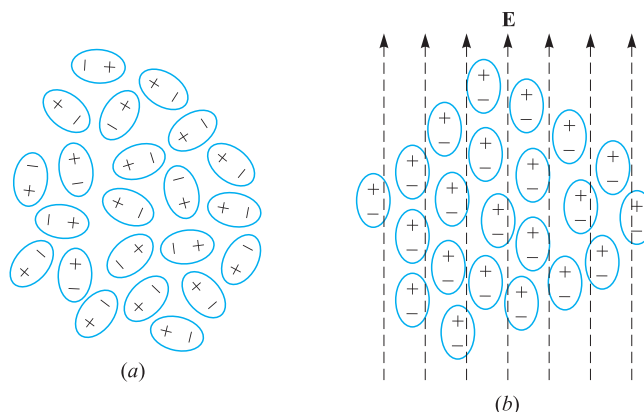
Figure E.3 Idealized sketches of ensembles of polar molecules under conditions of (a) random orientation of the dipole moments, and (b) dipole moments aligned under the influence of an applied electric field. Conditions in (b) are greatly exaggerated, since typically only a very small percentage of the dipoles align themselves with the field. But still enough alignment occurs to produce measurable changes in the material properties.

to fully describe the medium polarization properties, but the results of such studies often reduce to those of the spring model when field amplitudes are very low.

Another way that a dielectric can respond to an electric field is through the orientation of molecules that possess permanent dipole moments. In such cases, the molecules must be free to move or rotate, and so the material is typically a liquid or a gas. Figure E.3 shows an arrangement of polar molecules in a liquid (such as water) in which there is no applied field (Figure E.3a) and where an electric field is present (Figure E.3b). Applying the field causes the dipole moments, previously having random orientations, to line up, and so a net material polarization, **P**, results. Associated with this, of course, is a susceptibility function, χ_e, through which **P** relates to **E**.

Some interesting developments occur when the applied field is time-harmonic. With field periodically reversing direction, the dipoles are forced to follow, but they do so against their natural propensity to randomize, owing to thermal motion. Thermal motion thus acts as a "restoring" force, effectively opposing the applied field. We can also think of the thermal effects as viscous forces that introduce some difficulty in "pushing" the dipoles back and forth. One might expect (correctly) that polarizations of greater amplitude in each direction can be attained at lower frequencies, because enough time is given during each cycle for the dipoles to achieve complete alignment. The polarization amplitude will weaken as the frequency increases because there is no longer enough time for complete alignment during each cycle. This is the basic description of the *dipole relaxation* mechanism for the complex permittivity. There is no resonant frequency associated with the process.

The complex susceptibility associated with dipole relaxation is essentially that of an "overdamped" oscillator, and is given by

$$\chi_{\text{rel}} = \frac{Np^2/\epsilon_0}{3k_B T(1 + j\omega\tau)} \tag{E.23}$$

where p is the permanent dipole moment magnitude of each molecule, k_B is Boltzmann's constant, and T is the temperature in degees Kelvin. τ is the thermal randomization time, defined as the time for the polarization, $\mathbf{P}$, to relax to $1/e$ of its original value when the field is turned off. χ_{rel} is complex, and so it will possess absorptive and dispersive components (imaginary and real parts) as we found in the resonant case. The form of Eq. (E.23) is identical to that of the response of a series RC circuit driven by a sinusoidal voltage (where τ becomes RC).

Microwave absorption in water occurs through the relaxation mechanism in polar water molecules, and is the primary means by which microwave cooking is done, as discussed in Chapter 11. Frequencies near 2.5 GHz are typically used, since these provide the optimum penetration depth. The peak water absorption arising from dipole relaxation occurs at much higher frequencies, however.

A given material may possess more than one resonance and may have a dipole relaxation response as well. In such cases, the net susceptibility is found in frequency domain by the direct sum of all component susceptibilities. In general, we may write:

$$\chi_e = \chi_{\text{rel}} + \sum_{i=1}^{n} \chi_{\text{res}}^i \tag{E.24}$$

where χ_{res}^i is the susceptibility associated with the ith resonant frequency, and n is the number of resonances in the material. The reader is referred to the references for Chapter 11 for further reading on resonance and relaxation effects in dielectrics.

F
APPENDIX

Answers to Odd-Numbered Problems

Chapter 1

1.1 (a) $0.92\mathbf{a}_x + 0.36\mathbf{a}_y + 0.4\mathbf{a}_z$
(b) 48.6 (c) $-580.5\mathbf{a}_x + 3193\mathbf{a}_y - 2902\mathbf{a}_z$

1.3 $(7.8, -7.8, 3.9)$

1.5 (a) $48\mathbf{a}_x + 36\mathbf{a}_y + 18\mathbf{a}_z$
(b) $-0.26\mathbf{a}_x + 0.39\mathbf{a}_y + 0.88\mathbf{a}_z$
(c) $0.59\mathbf{a}_x + 0.20\mathbf{a}_y - 0.78\mathbf{a}_z$
(d) $100 = 16x^2y^2 + 4x^4 + 16x^2 + 16 + 9z^4$

1.7 (a) (1) the plane $z = 0$, with $|x| < 2$, $|y| < 2$;
(2) the plane $y = 0$ with $|x| < 2$, $|z| < 2$; (3) the
plane $x = 0$, with $|y| < 2$, $|z| < 2$; (4) the plane
$x = \pi/2$, with $|y| < 2$, $|z| < 2$ (b) the plane
$2z = y$, with $|x| < 2$, $|y| < 2$, $|z| < 1$
(c) the plane $y = 0$, with $|x| < 2$, $|z| < 2$

1.9 (a) $0.6\mathbf{a}_x + 0.8\mathbf{a}_y$ (b) $53°$ (c) 26

1.11 (a) $(-0.3, 0.3, 0.4)$ (b) 0.05 (c) 0.12 (d) $78°$

1.13 (a) $(0.93, 1.86, 2.79)$ (b) $(9.07, -7.86, 2.21)$
(c) $(0.02, 0.25, 0.26)$

1.15 (a) $(0.08, 0.41, 0.91)$ (b) $(0.30, 0.81, 0.50)$
(c) 30.3 (d) 32.0

1.17 (a) $(0.664, -0.379, 0.645)$
(b) $(-0.550, 0.832, 0.077)$
(c) $(0.168, 0.915, 0.367)$

1.19 (a) $(1/\rho)\mathbf{a}_\rho$ (b) $0.5\mathbf{a}_\rho$, or $0.41\mathbf{a}_x + 0.29\mathbf{a}_y$

1.21 (a) $-6.66\mathbf{a}_\rho - 2.77\mathbf{a}_\phi + 9\mathbf{a}_z$
(b) $-0.59\mathbf{a}_\rho + 0.21\mathbf{a}_\phi - 0.78\mathbf{a}_z$
(c) $-0.90\mathbf{a}_\rho - 0.44\mathbf{a}_z$

1.23 (a) 6.28 (b) 20.7 (c) 22.4 (d) 3.21

1.25 (a) $1.10\mathbf{a}_r + 2.21\mathbf{a}_\phi$ (b) 2.47 (c) $0.45\mathbf{a}_r + 0.89\mathbf{a}_\phi$

1.27 (a) 2.91 (b) 12.61 (c) 17.49 (d) 2.53

1.29 (a) $0.59\mathbf{a}_r + 0.38\mathbf{a}_\theta - 0.72\mathbf{a}_\phi$
(b) $0.80\mathbf{a}_r - 0.22\mathbf{a}_\theta - 0.55\mathbf{a}_\phi$
(c) $0.66\mathbf{a}_r + 0.39\mathbf{a}_\theta - 0.64\mathbf{a}_\phi$

Chapter 2

2.1 $(10/\sqrt{6}, -10/\sqrt{6})$

2.3 $21.5\mathbf{a}_x$ μN

2.5 (a) $4.58\mathbf{a}_x - 0.15\mathbf{a}_y + 5.51\mathbf{a}_z$
(b) -6.89 or -22.11

2.7 $159.7\mathbf{a}_\rho + 27.4\mathbf{a}_\phi - 49.4\mathbf{a}_z$

2.9 (a) $(x+1) = 0.56\,[(x+1)^2 + (y-1)^2 + (z-3)^2]^{1.5}$
(b) 1.69 or 0.31

2.11 (a) -1.63 μC
(b) $-30.11\mathbf{a}_x - 180.63\mathbf{a}_y - 150.53\mathbf{a}_z$
(c) $-183.12\mathbf{a}_\rho - 150.53\mathbf{a}_z$ (d) -237.1

2.13 (a) 82.1 pC (b) 4.24 cm

2.15 (a) 3.35 pC (b) 124 μC/m^3

2.17 (a) $57.5\mathbf{a}_y - 28.8\mathbf{a}_z$ V/m (b) $23\mathbf{a}_y - 46\mathbf{a}_z$

2.19 (a) $7.2\mathbf{a}_x + 14.4\mathbf{a}_y$ kV/m
(b) $4.9\mathbf{a}_x + 9.8\mathbf{a}_y + 4.9\mathbf{a}_z$ kV/m

2.21 $126\mathbf{a}_y$ μN/m

2.23 (a) 8.1 kV/m (b) -8.1 kV/m

2.25 $-3.9\mathbf{a}_x - 12.4\mathbf{a}_y - 2.5\mathbf{a}_z$ V/m

2.27 (a) $y^2 - x^2 = 4xy - 19$ (b) $0.99\mathbf{a}_x + 0.12\mathbf{a}_y$

2.29 (a) 12.2 (b) $-0.87\mathbf{a}_x - 0.50\mathbf{a}_y$
(c) $y = (1/5)\ln\cos 5x + 0.13$

Chapter 3

3.1 (a) $\bar{F} = [Q_1 Q_2/4\pi\epsilon_0 R^2]$ because $\mathbf{a}_r$
(b) Same as part (a)
(c) 0
(d) Force will become attractive!

3.3 (a) 0.25 nC (b) 9.45 pC

3.5 360 nC

3.7 (a) 4.0×10^{-9} nC (b) 3.2×10^{-4} nC/m^2

3.9 (a) 164 pC (b) 130 nC/m^2 (c) 32.5 nC/m^2

3.11 $\mathbf{D} = 0$ $(\rho < 1$ mm$)$;

$D_\rho = \frac{10^{-15}}{2\pi^2\rho}[\sin(2000\pi\rho) + 2\pi[1 -$

$10^3\rho\cos(2000\pi\rho)]]$C/m^2 $(1$ mm $< \rho < 1.5$ mm$)$;
$D_\rho = \frac{2.5\times10^{-15}}{\pi\rho}$ C/m^2 $(\rho > 1.5$ mm$)$

3.13 (a) $D_r(r < 2) = 0$; $D_r(r = 3) = 8.9 \times 10^{-9}$ C/m^2; $D_r(r = 5) = 6.4 \times 10^{-10}$ C/m^2
(b) $\rho_{s0} = -(4/9) \times 10^{-9}$ C/m^2

3.15 (a) $[(8\pi L)/3][\rho_1^3 - 10^{-9}]\mu$C where ρ_1
is in meters (b) $4(\rho_1^3 - 10^{-9})/(3\rho_1)\mu$C/m^2
where ρ_1 is in meters
(c) $D_\rho(0.8$ mm$) = 0$; $D_\rho(1.6$ mm$) = 3.6 \times 10^{-6}\mu$C/m^2; $D_\rho(2.4$ mm$) = 3.9 \times 10^{-6}$ μC/m^2

3.17 (a) 0.1028 C (b) 12.83 (c) 0.1026 C

3.19 113 nC

3.21 (a) 8.96 (b) 71.67 (c) -2

3.23 (b) $\rho_{v0} = 3Q/(4\pi a^3)$ $(0 < r < a)$;
$D_r = Qr/4\pi a^3$ and $\nabla \cdot \mathbf{D} = 3Q/(4\pi a^3)$
$(0 < r < a)$; $D_r = Q/(4\pi r^2)$ and
$\nabla \cdot \mathbf{D} = 0$ $(r > a)$

3.25 (a) 17.50 C/m^3 (b) $5\mathbf{a}_r$ C/m^2 (c) 320π C
(d) 320π C

3.27 (a) 1.20 mC/m^3 (b) 0 (c) -32μC/m^2

3.29 (a) 3.47 C (b) 3.47 C

3.31 -3.91 C

Chapter 4

4.1 (a) -12 nJ (b) 24 nJ (c) -36 nJ (d) -44.9 nJ
(e) -41.8 nJ

4.3 (a) 3.1 μJ (b) 3.1 μJ

4.5 (a) 2 (b) -2

4.7 (a) 90 (b) 82

4.9 (a) 8.14 V (b) 1.36 V

4.11 1.98 kV

4.13 576 pJ

4.15 -68.4 V

4.17 (a) -3.026 V (b) -9.678 V

4.19 .081 V

4.21 (a) -15.0 V (b) 15.0 V
(c) $7.1\mathbf{a}_x + 22.8\mathbf{a}_y - 71.1\mathbf{a}_z$ V/m
(d) 75.0 V/m
(e) $-0.095\mathbf{a}_x - 0.304\mathbf{a}_y + 0.948\mathbf{a}_z$
(f) $62.8\mathbf{a}_x + 202\mathbf{a}_y - 629\mathbf{a}_z$ pC/m^2

4.23 (a) $-48\rho^{-4}$ V/m (b) -673 pC/m^3 (c) -1.96 nC

4.25 (a) $V_p = 279.9$ V, $\mathbf{E}_p = -179.9\mathbf{a}_\rho - 75.0\mathbf{a}_\phi$ V/m,
$\mathbf{D}_p = -1.59\mathbf{a}_\rho - .664\mathbf{a}_\phi$ nC/m^2, $\rho_{vp} = -443$ pC/m^3 (b) -5.56 nC

4.27 (a) 5.78 V (b) 25.2 V/m (c) 5.76 V

4.29 1.31 V

4.31 (a) 387 pJ (b) 207 pJ

4.33 (a) $(5 \times 10^{-6})/(4\pi r^2)\mathbf{a}_r$ C/m^2
(b) 2.81 J (c) 4.45 pF

4.35 (a) 0.779 μJ (b) 1.59 μJ

Chapter 5

5.1 (a) -1.23 MA (b) 0 (c) 0, as expected

5.3 (a) 77.4 A (b) $53.0\mathbf{a}_r$ A/m^2

5.5 (a) -178.0 A (b) 0 (c) 0

5.7 (a) mass flux density in (kg/m^2 $-$ s) and mass
density in (kg/m^3) (b) -550 g/m^3 $-$ s

5.9 (a) 0.28 mm (b) 6.0×10^7 A/m^2

5.11 (a) $\mathbf{E} = [(9.55)/\rho l]\mathbf{a}_\rho$ V/m, $V = (4.88)/l$ V and
$R = (1.63)/l$ Ω, where l is the cylinder length
(not given) (b) $14.64/l$ W

5.13 (a) 0.147 V (b) 0.144 V

5.15 (a) $(\rho + 1)z^2\cos\phi = 2$
(b) $\rho = 0.10$, $\mathbf{E}(.10, .2\pi, 1.5) = -18.2\mathbf{a}_\rho + 145\mathbf{a}_\phi - 26.7\mathbf{a}_z$ V/m (c) 1.32 nC/m^2

5.17 (a) $\mathbf{D}(z = 0) = -(100\epsilon_0 x)/(x^2 + 4)\mathbf{a}_z$ C/m^2
(c) -0.92 nC

5.19 (a) At 0 V: $2x^2y - z = 0$. At 60 V:
$2x^2y - z = 6/z$ (b) 1.04 nC/m^2
(c) $-[0.60\mathbf{a}_x + 0.68\mathbf{a}_y + 0.43\mathbf{a}_z]$

5.21 (a) 1.20 kV (b) $\mathbf{E}_p = 723\mathbf{a}_x - 18.9\mathbf{a}_y$ V/m

5.23 (a) 289.5 V (b) $z/[(x-1)^2 + y^2 + z^2]^{1.5} -$
$z/[(x+1)^2 + y^2 + z^2]^{1.5} = 0.222$

5.25 (a) 4.7×10^{-5} S/m (b) 1.1×10^{-3} S/m
(c) 1.2×10^{-2} S/m

5.27 (a) 6.26 pC/m^2 (b) 1.000176

5.29 (a) $\mathbf{E} = [(144.9)/\rho]\mathbf{a}_\rho$ V/m, $\mathbf{D} =$
$(3.28\mathbf{a}_\rho)/\rho$ nC/m^2 (b) $V_{ab} = 192$ V, $\chi_e = 1.56$
(c) $[(5.0 \times 10^{-29})/\rho]\mathbf{a}_\rho$ C · m

5.31 (a) 80 V/m (b) $-60\mathbf{a}_y - 30\mathbf{a}_z$ V/m (c) 67.1 V/m
(d) 104.4 V/m (e) 40.0° (f) 2.12 nC/m^2 (g) 2.97
nC/m^2 (h) $2.12\mathbf{a}_x - 2.66\mathbf{a}_y - 1.33\mathbf{a}_z$ nC/m^2
(i) $1.70\mathbf{a}_x - 2.13\mathbf{a}_y - 1.06\mathbf{a}_z$ nC/m^2 (j) 54.5°

5.33 $125\mathbf{a}_x + 175\mathbf{a}_y$ V/m

5.35 (a) $\mathbf{E}_2 = \mathbf{E}_1$ (b) $W_{E1} = 45.1$ μJ, $W_{E2} = 338$ μJ

Chapter 6

6.1 $b/a = \exp(2\pi d/W)$

6.3 barium titanate

6.5 451 pF

6.7 (a) 3.05 nF (b) 5.21 nF (c) 6.32 nF (d) 9.83 nF

6.9 (a) 143 pF (b) 101 pF

6.11 (a) 53.3 pF (b) 41.7 pF

6.13 $K_1 = 23.0$, $\rho_L = 8.87$ nC/m, $a = 13.8$ m,
$C = 35.5$ pF

6.15 (a) 47.3 nC/m^2 (b) -15.8 nC/m^2 (c) 24.3 pF/m

6.17 Exact value: 57 pF/m

6.19 Exact value: $11\epsilon_0$ F/m

6.21 (b) $C \approx 110$ pF/m (c) Result would not change.

6.23 (a) 3.64 nC/m (b) 206 mA

6.25 (a) -8 V (b) $8\mathbf{a}_x - 8\mathbf{a}_y - 24\mathbf{a}_z$ V/m
(c) $-4xz(z^2 + 3y^2)$ C/m^3
(d) $xy^2z^3 = -4$ (e) $y^2 - 2x^2 = 2$ and
$3x^2 - z^2 = 2$ (f) No

6.27 $f(x, y) = -4e^{2x} + 3x^2$, $V(x, y) = 3(x^2 - y^2)$

6.29 (b) $A = 112.5$, $B = -12.5$ or
$A = -12.5$, $B = 112.5$

6.31 (a) -106 pC/m^3 (b) ±0.399 pC/m^2 (depending on
which side of the surface is considered)

6.33 (a) yes, yes, yes, no (b) At the 100 V surface, no
for all. At the 0 V surfaces, yes, except for $V_1 + 3$.
(c) Only V$_2$ is

6.35 (a) 33.33 V (b) $[(100)/3]\mathbf{a}_z + 50\mathbf{a}_y$ V/m

6.37 (a) 1.01 cm (b) 22.8 kV/m (c) 3.15

6.39 (a) $(-2.00 \times 10^4)\phi + 3.78 \times 10^3$ V
(b) $[(2.00 \times 10^4)/\rho]\mathbf{a}_\phi$ V/m
(c) $(2.00 \times 10^4\epsilon_0/\rho)\mathbf{a}_\phi$ C/m^2
(d) $[(2.00 \times 10^4)/\rho]$ C/m^2
(e) 84.7 nC

(f) $V(\phi) = 28.7\phi + 194.9$ V, $\mathbf{E} = -(28.7)/\rho\mathbf{a}_\phi$
V/m, $\mathbf{D} = -(28.7\epsilon_0)/\rho\mathbf{a}_\phi$ C/m^2, $\rho_s = (28.7\epsilon_0)/$
ρ C/m^2, $Q_b = 122$ pC (g) 471 pF

6.41 (a) 12.5 mm (b) 26.7 kV/m
(c) 4.23 (with given $\rho_s = 1.0\,\mu$C/m^2)

6.43 (a) $\alpha_A = 26.57°$, $\alpha_B = 56.31°$ (b) 23.3 V

6.45 (a) $833.3r^{-.4}$ V (b) $833.3r^{-.4}$ V

Chapter 7

7.1 (a) $-294\mathbf{a}_x + 196\mathbf{a}_y$ μA/m
(b) $-127\mathbf{a}_x + 382\mathbf{a}_y$ μA/m
(c) $-421\mathbf{a}_x + 578\mathbf{a}_y$ μA/m

7.3 (a)

$$H = \frac{I}{2\pi\rho}\left[1 - \frac{a}{\sqrt{\rho^2 + a^2}}\right]\mathbf{a}_\phi \text{ A/m}$$

(b) $1/\sqrt{3}$

7.5

$$|\mathbf{H}| = \frac{I}{2\pi}\left[\left(\frac{2}{y^2 + 2y + 5} - \frac{2}{y^2 - 2y + 5}\right)^2 \right.$$
$$\left. + \left(\frac{(y-1)}{y^2 - 2y + 5} - \frac{(y+1)}{y^2 + 2y + 5}\right)^2\right]^{1/2}$$

7.7 (a) $\mathbf{H} = I/(2\pi^2 z)(\mathbf{a}_x - \mathbf{a}_y)$ A/m
(b) 0

7.9 $-1.50\mathbf{a}_y$ A/m

7.11 2.0 A/m, 933 mA/m, 360 mA/m, 0

7.13 (e) $H_z(a < \rho < b) = k_b$; $H_z(\rho > b) = 0$

7.15 (a) $45e^{-150\rho}\mathbf{a}_z$ kA/m^2
(b) $12.6[1 - (1 + 150\rho_0)e^{-150\rho_0}]$ A
(c) $\frac{2.00}{\rho}[1 - (1 + 150\rho)e^{-150\rho}]$ A/m

7.17 (a) $2.2 \times 10^{-1}\mathbf{a}_\phi$ A/m (just inside), $2.3 \times$
$10^{-2}\mathbf{a}_\phi$ A/m (just outside)
(b) $3.4 \times 10^{-1}\mathbf{a}_\phi$ A/m
(c) $1.3 \times 10^{-1}\mathbf{a}_\phi$ A/m (d) $-1.3 \times 10^{-1}\mathbf{a}_z$ A/m

7.19 (a) $\mathbf{K} = -I\mathbf{a}_r/2\pi r$ A/m ($\theta = \pi/2$)
(b) $\mathbf{J} = I\mathbf{a}_r/[2\pi r^2(1 - 1/\sqrt{2})]$ A/m^2($\theta < \pi/4$)
(c) $\mathbf{H} = I\mathbf{a}_\phi/[2\pi r \sin\theta]$ A/m ($\pi/4 < \theta < \pi/2$)
(d) $\mathbf{H} = I(1 - \cos\theta)\mathbf{a}_\phi/[2\pi r \sin\theta(1 - 1/\sqrt{2})]$
A/m ($\theta < \pi/4$)

7.21 (a) $\mathbf{I} = 2\pi ba^3/3$ A (b) $\mathbf{H}_{\text{in}} = b\rho^2/3\mathbf{a}_\phi$ A/m
(c) $\mathbf{H}_{\text{out}} = ba^3/3\rho$ $\mathbf{a}_\phi$ A/m

7.23 (a) $60\rho\mathbf{a}_z$ A/m^2 (b) 40π A (c) 40π A

7.25 (a) -259 A (b) -259 A

7.27 (a) $2(x + 2y)/z^3\mathbf{a}_x + 1/z^2\mathbf{a}_z$ A/m
(b) same as part (a) (c) 1/8 A

7.29 (a) $1.59 \times 10^7 \mathbf{a}_z$ A/m^2 (b) $7.96 \times 10^6 \rho \mathbf{a}_\phi$ A/m,
$10\rho \mathbf{a}_\phi$ Wb/m^2 (c) as expected (d) $1/(\pi\rho)\mathbf{a}_\phi$ A/m,
$\mu_0/(\pi\rho)\mathbf{a}_\phi$ Wb/m^2 (e) as expected

7.31 (a) $0.392\ \mu$Wb (b) $1.49\ \mu$Wb (c) $27\ \mu$Wb

7.35 (a) -40ϕ A $\quad (2 < \rho < 4)$, $0\ (\rho > 4)$
(b) $40\mu_0 \ln(3/\rho)\mathbf{a}_z$ Wb/m

7.37 $[120 - (400/\pi)\phi]$ A $\ (0 < \phi < 2\pi)$

7.39 (a) $-30\mathbf{a}_y$ A/m (b) $30y - 6$ A
(c) $-30\mu_0\mathbf{a}_y$ Wb/m^2 (d) $\mu_0(30x - 3)\mathbf{a}_z$ Wb/m

7.41 (a) $-100\rho/\mu_0\mathbf{a}_\phi$ A/m, $-100\rho\mathbf{a}_\phi$ Wb/m^2
(b) $-\frac{200}{\mu_0}\mathbf{a}_z$ A/m^2 (c) -500 MA (d) -500 MA

7.43

$$A_z = \frac{\mu_0 I}{96\pi}\left[\left(\frac{\rho^2}{a^2} - 25\right) + 98 \ln\left(\frac{5a}{\rho}\right)\right] \text{Wb/m}$$

Chapter 8

8.1 (a) $(.90, 0, -.135)$ (b) $3 \times 10^5\mathbf{a}_x - 9 \times 10^4\mathbf{a}_z$ m/s
(c) 1.5×10^{-5} J

8.3 (a) $.70\mathbf{a}_x + .70\mathbf{a}_y - .12\mathbf{a}_z$ (b) 7.25 fJ

8.5 (a) $-18\mathbf{a}_x$ nN (b) $19.8\mathbf{a}_z$ nN (c) $36\mathbf{a}_x$ nN

8.7 (a) $-35.2\mathbf{a}_y$ nN/m (b) 0 (c) 0

8.9 $4\pi \times 10^{-5}$ N/m

8.13 (a) $-1.8 \times 10^{-4}\mathbf{a}_y$ N $\cdot$ m
(b) $-1.8 \times 10^{-4}\mathbf{a}_y$ N $\cdot$ m
(c) $-1.5 \times 10^{-5}\mathbf{a}_y$ N $\cdot$ m

8.15 $(6 \times 10^{-6})[b - 2\tan^{-1}(b/2)]\mathbf{a}_y$ N $\cdot$ m

8.17 $\Delta w/w = \Delta m/m = 1.3 \times 10^{-6}$

8.19 (a) $77.6y\mathbf{a}_z$ kA/m (b) 5.15×10^{-6} H/m
(c) 4.1 (d) $241y\mathbf{a}_z$ kA/m (e) $77.6\mathbf{a}_x$ kA/m^2
(f) $241\mathbf{a}_x$ kA/m^2 (g) $318\mathbf{a}_x$ kA/m^2

8.21 (Use $\chi_m = .003$) (a) 47.7 A/m (b) 6.0 A/m
(c) 0.288 A/m

8.23 (a) 637 A/m, 1.91×10^{-3} Wb/m^2, 884 A/m
(b) 478 A/m, 2.39×10^{-3} Wb/m^2, 1.42×10^3 A/m
(c) 382 A/m, 3.82×10^{-3} Wb/m^2, 2.66×10^3 A/m

8.25 (a) $1.91/\rho$ A/m $\quad (0 < \rho < \infty)$
(b) $(2.4 \times 10^{-6}/\rho)\mathbf{a}_\phi$ T $(\rho < .01)$,
$(1.4 \times 10^{-5}/\rho)\mathbf{a}_\phi$ T $(.01 < \rho < .02)$,
$(2.4 \times 10^{-6}/\rho)\mathbf{a}_\phi$ T $(\rho > .02)$ $(\rho$ in meters)

8.27 (a) $-4.83\mathbf{a}_x - 7.24\mathbf{a}_y + 9.66\mathbf{a}_z$ A/m
(b) $54.83\mathbf{a}_x - 22.76\mathbf{a}_y + 10.34\mathbf{a}_z$ A/m
(c) $54.83\mathbf{a}_x - 22.76\mathbf{a}_y + 10.34\mathbf{a}_z$ A/m
(d) $-1.93\mathbf{a}_x - 2.90\mathbf{a}_y + 3.86\mathbf{a}_z$ A/m
(e) $102°$ (f) $95°$

8.29 10.5 mA

8.31 (a) 2.8×10^{-4} Wb (b) 2.1×10^{-4} Wb
(c) $\approx 2.5 \times 10^{-4}$ Wb

8.33 (a) $23.9/\rho$ A/m (b) $3.0 \times 10^{-4}/\rho$ Wb/m^2
(c) 5.0×10^{-7} Wb
(d) $23.9/\rho$ A/m, $6.0 \times 10^{-4}/\rho$ Wb/m^2, 1.0×10^{-6}
Wb (e) 1.5×10^{-6} Wb

8.35 (a) $20/(\pi r \sin\theta)\mathbf{a}_\phi$ A/m (b) 1.35×10^{-4} J

8.37 $0.17\ \mu$H

8.39 (a) $(1/2)wd\mu_0 K_0^2$ J/m (b) $\mu_0 d/w$ H/m
(c) $\Phi = \mu_0 d K_0$ Wb

8.41 (a) $33\ \mu$H (b) $24\ \mu$H

8.43 (b)

$$L_{\text{int}} = \frac{2W_H}{I^2}$$

$$= \frac{\mu_0}{8\pi}\left[\frac{d^4 - 4a^2c^2 + 3c^4 + 4c^4\ln(a/c)}{(a^2 - c^2)^2}\right]\text{H/m}$$

Chapter 9

9.1 (a) $-5.33\sin 120\pi t$ V (b) $21.3\sin(120\pi t)$ mA

9.3 (a) $-1.13 \times 10^5[\cos(3 \times 10^8 t - 1)$
$-\cos(3 \times 10^8 t)]$ V (b) 0

9.5 (a) -4.32 V (b) -0.293 V

9.7 (a) $(-1.44)/(9.1 + 39.6t)$ A
(b) $-1.44[\frac{1}{61.9 - 39.6t} + \frac{1}{9.1 + 39.6t}]$ A

9.9 $2.9 \times 10^3[\cos(1.5 \times 10^8 t - 0.13x) -$
$\cos(1.5 \times 10^8 t)]$ W

9.11 (a) $\left(\frac{10}{\rho}\right)\cos(10^5 t)\mathbf{a}_\rho$ A/m^2 (b) $8\pi\cos(10^5 t)$ A
(c) $-0.8\pi\sin(10^5 t)$ A (d) 0.1

9.13 (a) $\mathbf{D} = 1.33 \times 10^{-13}\sin(1.5 \times 10^8 t -$
$bx)\mathbf{a}_y$ C/m^2, $\mathbf{E} = 3.0 \times 10^{-3}\sin(1.5 \times$
$10^8 t - bx)\mathbf{a}_y$ V/m
(b) $\mathbf{B} = (2.0)b \times 10^{-11}\sin(1.5 \times 10^8 t - bx)\mathbf{a}_z$T,
$\mathbf{H} = (4.0 \times 10^{-6})b\sin(1.5 \times 10^8 t - bx)\mathbf{a}_z$ A/m
(c) $4.0 \times 10^{-6}b^2\cos(1.5 \times 10^8 t - bx)\mathbf{a}_y$ A/m^2
(d) $\sqrt{5.0}$ m^{-1}

9.15 $\mathbf{B} = 6 \times 10^{-5}\cos(10^{10}t - \beta x)\mathbf{a}_z$T, $\mathbf{D} =$
$-(2\beta \times 10^{-10})\cos(10^{10}t - \beta x)\mathbf{a}_y$ C/m^2,
$\mathbf{E} = -1.67\beta\cos(10^{10}t - \beta x)\mathbf{a}_y$ V/m, $\beta =$
± 600 rad/m

9.17 $a = 66$ m^{-1}

9.21 (a) $\pi \times 10^9$ sec^{-1}
(b) $\frac{500}{\rho}\sin(10\pi z)\sin(\omega t)\mathbf{a}_\rho$ V/m

9.23 (a) $\mathbf{E}_{N1} = 10\cos(10^9 t)\mathbf{a}_z$ V/m $\mathbf{E}_{t1} =$ $(30\mathbf{a}_x + 20\mathbf{a}_y)\cos(10^9 t)$ V/m $\mathbf{D}_{N1} = 200\cos(10^9 t)\mathbf{a}_z$ pC/m^2 $\mathbf{D}_{t1} =$ $(600\mathbf{a}_x + 400\mathbf{a}_y)\cos(10^9 t)$ pC/m^2 (b) $\mathbf{J}_{N1} = 40\cos(10^9 t)\mathbf{a}_z$ mA/m^2 $\mathbf{J}_{t1} =$ $(120\mathbf{a}_x + 80\mathbf{a}_y)\cos(10^9 t)$ mA/m^2 (c) $\mathbf{E}_{t2} = (30\mathbf{a}_x + 20\mathbf{a}_y)\cos(10^9 t)$ V/m $\mathbf{D}_{t2} =$ $(300\mathbf{a}_x + 200\mathbf{a}_y)\cos(10^9 t)$ pC/m^2 $\mathbf{J}_{t2} = (30\mathbf{a}_x + 20\mathbf{a}_y)\cos(10^9 t)$ mA/m^2 (d) $\mathbf{E}_{N2} = 20.3\cos(10^9 t + 5.6°)\mathbf{a}_z$ V/m $\mathbf{D}_{N2} =$ $203\cos(10^9 t + 5.6°)\mathbf{a}_z$ pC/m^2 $\mathbf{J}_{N2} =$ $20.3\cos(10^9 t + 5.6°)\mathbf{a}_z$ mA/m^2

9.25 (b) $\mathbf{B} = \left(t - \frac{z}{c}\right)\mathbf{a}_y$ T $\mathbf{H} = \frac{1}{\mu_0}\left(t - \frac{z}{c}\right)\mathbf{a}_y$ A/m $\mathbf{E} = (ct - z)\mathbf{a}_x$ V/m $\mathbf{D} = \epsilon_0(ct - z)\mathbf{a}_x$ C/m^2

Chapter 10

10.1 $\gamma = 0.094 + j2.25$ $\alpha = 0.094$ Np/m $\beta = 2.25$ rad/m $\lambda = 2.8$ m $Z_0 = 93.6 - j3.64$ Ω

10.3 (a) 96 pF/m (b) 1.44×10^8 m/s (c) 3.5 rad/m (d) $\Gamma = -0.09$, $s = 1.2$

10.5 (a) 83.3 nH/m, 33.3 pF/m (b) 65 cm

10.7 7.9 mW

10.9 (a) $\lambda/8$ (b) $\lambda/8 + m\lambda/2$

10.11 (a) V_0^2/R_L (b) $R_L V_0^2/(R\ell + R_L)^2$ (c) V_0^2/R_L (d) $(V_0^2/R_L)\exp(-2\ell\sqrt{RG})$

10.13 (a) 6.28×10^8 rad/s (b) $4\cos(\omega t - \pi z)$A (c) $0.287\angle 1.28$ rad (d) $57.5\exp[j(\pi z + 1.28)]$ V (e) $257.5\angle 36°$ V

10.15 (a) 104 V (b) $52.6 - j123$ V

10.17 $P_{25} = 2.28$ W, $P_{100} = 1.16$ W

10.19 16.5 W

10.21 (a) $s = 2.62$ (b) $Z_L = 1.04 \times 10^3 + j69.8$ Ω (c) $z_{\max} = -7.2$ mm

10.23 (a) 0.037λ or 0.74 m (b) 2.61 (c) 2.61 (d) 0.463λ or 9.26 m

10.25 (a) $495 + j290$ Ω (b) $j98$ Ω

10.27 (a) 2.6 (b) $11 - j7.0$ mS (c) 0.213λ

10.29 $47.8 + j49.3$ Ω

10.31 (a) 3.8 cm (b)14.2 cm

10.33 (a) $d_1 = 7.6$ cm, $d = 17.3$ cm (b) $d_1 = 1.8$ cm, $d = 6.9$ cm

10.35 (a) 39.6 cm (b) 24 pF

10.37 $V_L = (1/3)V_0$ $(l/v < t < \infty)$ and is zero for $t < l/v$. $I_B = (V_0/100)$ A for $0 < t < 2l/v$ and is $(V_0/75)$ for $t > 2l/v$

10.39

$$\frac{l}{v} < t < \frac{5l}{4v}: \quad V_1 = 0.44\ V_0$$

$$\frac{3l}{v} < t < \frac{13l}{4v}: \quad V_2 = -0.15\ V_0$$

$$\frac{5l}{v} < t < \frac{21l}{4v}: \quad V_3 = 0.049\ V_0$$

$$\frac{7l}{v} < t < \frac{29l}{4v}: \quad V_4 = -0.017\ V_0$$

Voltages in between these times are zero.

10.41

$$0 < t < \frac{l}{2v}: \quad V_L = 0$$

$$\frac{l}{2v} < t < \frac{3l}{2v}: \quad V_L = \frac{V_0}{2}$$

$$t > \frac{3l}{2v}: \quad V_L = V_0$$

10.43

$$0 < t < 2l/v: \quad V_{R_L} = V_0/2$$

$$t > 2l/v: \quad V_{R_L} = 3V_0/4$$

$$0 < t < l/v: \quad V_{R_g} = 0, I_B = 0$$

$$t > l/v: \quad V_{R_g} = V_0/4, I_B = 3V_0/4Z_0$$

Chapter 11

11.3 (a) 0.33 rad/m (b) 18.9 m (c) $-3.76 \times 10^3 \mathbf{a}_z$ V/m

11.5 (a) $\omega = 3\pi \times 10^8$ sec^{-1}, $\lambda = 2$ m, and $\beta = \pi$ rad/m (b) $-8.5\mathbf{a}_x - 9.9\mathbf{a}_y$ A/m (c) 9.08 kV/m

11.7 $\beta = 25$m^{-1}, $\eta = 278.5$ Ω, $\lambda = 25$ cm, $v_p = 1.01 \times 10^8$ m/s, $\epsilon_R = 4.01$, $\mu_R = 2.19$, and $\mathbf{H}(x, y, z, t) = 2\cos(8\pi \times 10^8 t - 25x)\mathbf{a}_y + 5\sin(8\pi \times 10^8 t - 25x)\mathbf{a}_z$ A/m

11.9 (a) $\beta = 0.4\pi$ rad/m, $\lambda = 5$ m, $v_p = 5 \times 10^7$ m/s, and $\eta = 251$ Ω (b) $-403\cos(2\pi \times 10^7 t)$ V/m (c) $1.61\cos(2\pi \times 10^{-7} t)$ A/m

11.11 (a) 0.74 kV/m (b) -3.0 A/m

11.13 $\mu = 2.28 \times 10^{-6}$ H/m, $\epsilon' = 1.07 \times 10^{-11}$ F/m, and $\epsilon'' = 2.90 \times 10^{-12}$ F/m

11.15 (a) $\lambda = 3$ cm, $\alpha = 0$ (b) $\lambda = 2.95$ cm, $\alpha = 9.24 \times 10^{-2}$ Np/m (c) $\lambda = 1.33$ cm, $\alpha = 335$ Np/m

11.17 $\langle \mathcal{S}_z \rangle(z = 0) = 315\mathbf{a}_z$ W/m^2, $\langle \mathcal{S}_z \rangle(z = 0.6) = 248\mathbf{a}_z$ W/m^2

11.19 (a) $\omega = 4 \times 10^8$ rad/s (b) $\mathbf{H}(\rho, z, t) = (4.0/\rho) \cos(4 \times 10^8 t - 4z) \mathbf{a}_\phi$ A/m
(c) $\langle \mathcal{S} \rangle = (2.0 \times 10^{-3}/\rho^2) \cos^2(4 \times 10^8 t - 4z) \mathbf{a}_z$ W/m² (d) P = 5.7 kW

11.21 (a) $H_{\phi 1}(\rho) = (54.5/\rho)(10^4 \rho^2 - 1)$ A/m
$(.01 < \rho < .012)$, $H_{\phi 2}(\rho) = (24/\rho)$ A/m
$(\rho > .012)$, $H_\phi = 0$ $(\rho < .01\text{m})$
(b) $\mathbf{E} = 1.09 \mathbf{a}_z$ V/m
(c) $\langle \mathcal{S} \rangle = -(59.4/\rho)(10^4 \rho^2 - 1) \mathbf{a}_\rho$ W/m²
$(.01 < \rho < .012$ m$)$, $-(26/\rho) \mathbf{a}_\rho$ W/m²
$(\rho > 0.12$ m$)$

11.23 (a) $1.4 \times 10^{-3} \Omega/\text{m}$ (b) 4.1×10^{-2} Ω/m
(c) $4.1 \times 10^{-1} \Omega/\text{m}$

11.25 $f = 1$ GHz, $\sigma = 1.1 \times 10^5$ S/m

11.27 (a) 4.7×10^{-8} (b) 3.2×10^3 (c) 3.2×10^3

11.29 (a) $\mathbf{H}_s = (E_0/\eta_0)(\mathbf{a}_y - j\mathbf{a}_x)e^{-j\beta z}$
(b) $\langle \mathcal{S} \rangle = (E_0^2/\eta_0) \mathbf{a}_z$ W/m² (assuming E_0 is real)

11.31 (a) $L = 14.6\ \lambda$ (b) Left

11.33 (a) $\mathbf{H}_s = (1/\eta)[-18 e^{j\phi} \mathbf{a}_x + 15 \mathbf{a}_y]e^{-j\beta z}$ A/m
(b) $\langle \mathcal{S} \rangle = 275$ Re $\{(1/\eta^*)\}$ W/m²

Chapter 12

12.1 0.01%

12.3 0.056 and 17.9

12.5 (a) 4.7×10^8 Hz (b) $691 + j177\ \Omega$ (c) -1.7 cm

12.7 (a) $s_1 = 1.96$, $s_2 = 2$, $s_3 = 1$ (b) -0.81 m

12.9 (a) 6.25×10^{-2} (b) 0.938 (c) 1.67

12.11 $641 + j501\ \Omega$

12.13 Reflected wave: left circular polarization; power fraction = 0.09. Transmitted wave: right circular polarization; power fraction = 0.91

12.15 (a) 2.55 (b) 2.14 (c) 0.845

12.17 2.41

12.19 (a) $d_1 = d_2 = d_3 = 0$ or $d_1 = d_3 = 0$, $d_2 = \lambda/2$
(b) $d_1 = d_2 = d_3 = \lambda/4$

12.21 (a) Reflected power: 15%. Transmitted power: 85% (b) Reflected wave: s-polarized.
Transmitted wave: Right elliptically polarized.

12.23 $n_0 = (n_1/n_2)\sqrt{n_1^2 - n_2^2}$

12.25 $0.76(-1.19$ dB$)$

12.27 2

12.29 4.3 km

Chapter 13

13.1 (a) 1.14 mm (b) 1.14 mm (c) 1.47 mm

13.3 14.2 pF/m, 0.786 μH/m, 0, 0.023 Ω/m

13.5 (a) 1.23 (b) 1.99 (c) 1.33

13.7 (a) 2.8 (b) 5.85×10^{-2}

13.9 (a) 4.9 (b) 1.33

13.11 9

13.13 9

13.15 1.5 ns

13.17 (a) 12.8 GHz (b) Yes

13.19 (a) 2.5 GHz $< f <$ 3.75 GHz (air-filled)
(b) 3.75 GHs $< f <$ 4.5 GHz (air-filled)

13.21 $a = 1.1$ cm, $b = 0.90$ cm

13.25 72 cm

13.27 3.32

13.29 (a) $\theta_{\min} = \sin^{-1}(n_3/n_1)$ (b) $v_{p,\max} = c/n_3$

13.31 greater than

Chapter 14

14.1 (a) $-0.284 \mathbf{a}_x - 0.959 \mathbf{a}_z$ (b) 0.258

14.3 (a) $-j(1.5 \times 10^{-2})e^{-j1000} \mathbf{a}_z$ V/m
(b) $-j(1.5 \times 10^{-2})e^{-j1000} \mathbf{a}_y$ V/m
(c) $-j(1.5 \times 10^{-2})(\mathbf{a}_y + \mathbf{a}_z)$ V/m
(d) $-(1.24 \times 10^{-2})(\mathbf{a}_y + \mathbf{a}_z)$ V/m
(e) 1.75×10^{-2} V/m

14.7 (a) 0.711 Ω (b) 0.178 Ω (c) 0.400 Ω

14.9 (a) 85.4 A (b) 5.03 A

14.11 (a) 85.4 A (b) 7.1 A

14.13 (a) $0.2 e^{-j1000\pi}$ V/m (b) $0.2 e^{-j1000\pi} e^{j0.5\pi}$ V/m
(c) 0

14.15 Primary maxima: $\theta = \pm 90°$, relative magnitude 1.00. Secondary maxima: $\theta = \pm 33.8°$ and $\theta = \pm 146.2°$, relative magnitude 0.186. $S_s = 7.3$ dB

14.17 (a) 36.5 W (b) 4.8 μW

14.19 $\xi = 0$, $d = \lambda$

14.21 (a) $\pm 48.2°$ (b) $\pm 60°$

14.23 (a) $+\pi/2$ (b) 0

14.25 (a) $\xi = -\pi/2$ (b) 5.6% of maximum (12.6 dB down)

14.29 (a) 1.7μW (b) 672 nW

INDEX