To verify that the current law is applicable, consider again the parallel-plate capacitor. In this case, $\mathcal{D} = q_f/A$ and hence the displacement current $A(d\mathcal{D}/dt)$ equals dq_f/dt. Thus the displacement current leaving via the insulating region equals the conduction current entering the capacitor plate.

Displacement current only exists when the electric field is time-varying. Since it is a current, one would expect it to create a magnetic field. Displacement current does indeed produce a magnetic field and hence Ampere's law must be restated to reflect this experimental fact. This has been done in Eq. (2–33). As explained in Sec. 2–4, the most important consequence of displacement current is its key role in electromagnetic wave propagation.

With Ampere's law modified to include the effect of displacement currents, all four of Maxwell's equations may now be stated. They are summarized here in both integral and differential forms.

Maxwell's equations in integral form.

$$\oint \vec{\mathcal{D}} \cdot d\vec{S} = \int_v \rho_v \, dv \quad , \quad \oint \vec{\mathcal{B}} \cdot d\vec{S} = 0 \tag{2–35}$$

$$\oint \vec{\mathcal{E}} \cdot d\vec{l} = -\int_S \frac{\partial \vec{\mathcal{B}}}{\partial t} \cdot d\vec{S} \tag{2–36}$$

$$\oint \vec{\mathcal{H}} \cdot d\vec{l} = \int_S \vec{\mathcal{J}} \cdot d\vec{S} + \int_S \frac{\partial \vec{\mathcal{D}}}{\partial t} \cdot d\vec{S} \tag{2–37}$$

where $\vec{\mathcal{D}} = \epsilon_R \epsilon_0 \vec{\mathcal{E}}$, $\vec{\mathcal{B}} = \mu_R \mu_0 \vec{\mathcal{H}}$ and $\vec{\mathcal{J}} = \sigma \vec{\mathcal{E}} = \frac{I}{A}$

Maxwell's equations in differential form. By applying the rules of vector calculus, Maxwell's equations may be written in differential or point form.[6] Namely,

$$\nabla \cdot \vec{\mathcal{D}} = \rho_v \quad , \quad \nabla \cdot \vec{\mathcal{B}} = 0 \tag{2–38}$$

$$\nabla \times \vec{\mathcal{E}} = -\frac{\partial \vec{\mathcal{B}}}{\partial t} \tag{2–39}$$

$$\nabla \times \vec{\mathcal{H}} = \vec{\mathcal{J}} + \frac{\partial \vec{\mathcal{D}}}{\partial t} \tag{2–40}$$

These relationships represent the fundamental equations of electromagnetism.

2–3b Boundary Conditions for Electric and Magnetic Fields

Most electromagnetic problems involve boundaries between regions having different electric and magnetic properties. The conditions that must exist at these boundaries may be deduced from Maxwell's equations. Table 2–1 lists the conditions for tangential (subscript T) and normal (subscript N) components of the fields.

[6] As explained in the preface, the use of the differential form of Maxwell's equations is limited to a few sections in the text. These sections are indicated in the table of contents by a star (✱).

Thus the direction of $\mathcal{E}$ lines terminating on the surface of a perfect conductor must be perpendicular to that surface.

When the fields are time-varying, $\mathcal{H}$ and $\mathcal{B}$ must also be zero within a perfect conductor. This is a direct consequence of Faraday's law, Eq. (2–36), and the fact that $\mathcal{E} \equiv 0$ in a perfect conductor. Thus for time-varying fields at the boundary of a perfect conductor, the magnetic field conditions become

$$\mathcal{H}_{T_1} = \mathcal{H} \qquad \text{and} \qquad \mathcal{B}_{N_1} = 0 \tag{2–42}$$

where again it is assumed that region 2 is the perfect conductor. If μ_{R_1} is a scalar, $\mathcal{H}_{N_1} = 0$, which means that the magnetic field must be tangential to the surface of a perfect conductor. Furthermore, if region 1 is a perfect insulator ($\tan\delta = 0$), the following boundary condition also holds for time-varying fields.

$$\frac{\partial \mathcal{H}_{T_1}}{\partial n} = 0 \tag{2–43}$$

where n represents a direction perpendicular to the boundary surface.

The analysis of microwave transmission lines and components make extensive use of the boundary conditions derived here. An appreciation of their meaning is also useful in estimating the electromagnetic field pattern in structures constrained by dielectric and metallic boundaries.

2–4 WAVE PROPAGATION IN PERFECT INSULATORS

The existence of self-propagating electromagnetic waves is the single most important consequence of Maxwell's equations. This propagation results from the fact that a time-varying magnetic field produces a time-varying electric field (Faraday's law) which, in turn, produces a time-varying magnetic field (Ampere's law) . . . and so forth. Since one type field produces the other in a plane normal to it, the electric and magnetic fields are always perpendicular to each other and travel at the same speed. This section reviews the important case of electromagnetic wave propagation in unbounded insulators.

The situation in which the fields are a function of only one coordinate represents the simplest application of Maxwell's equations. The resultant analysis describes the phenomena of uniform plane waves, which is of considerable engineering importance.

Consider the coordinate system shown in Fig. 2–10 and assume a time-varying electric field in the x direction that is independent of x and y. That is, $\mathcal{E}_x = f(z, t)$.

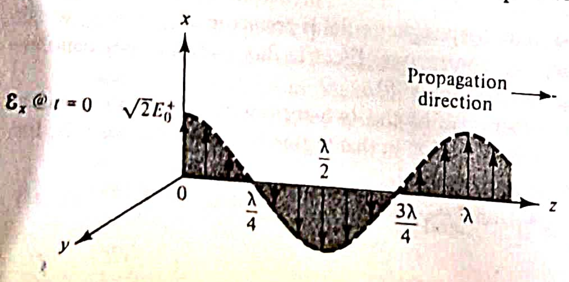


Figure 2–10 $\mathcal{E}_x$ at $t = 0$ as a function of position along the propagation axis.

Also assume that all space consists of a perfect insulator devoid of free charges, which means that both ρ_v and J are zero in Eqs. (2–35) to (2–40). Since $\mathscr{E}$ is not a function of x and y and only has a component in the x direction, Maxwell's equations reduce to[7]

$$\frac{\partial \mathscr{E}_x}{\partial z} = -\mu_R \mu_0 \frac{\partial \mathscr{H}_y}{\partial t} \quad \text{and} \quad -\frac{\partial \mathscr{H}_y}{\partial z} = \epsilon_R \epsilon_0 \frac{\partial \mathscr{E}_x}{\partial t} \tag{2–44}$$

By differentiating the first equation with respect to z and the second one with respect to t, $\mathscr{H}_y$ can be eliminated, resulting in

$$\frac{\partial^2 \mathscr{E}_x}{\partial z^2} = \mu_R \mu_0 \epsilon_R \epsilon_0 \frac{\partial^2 \mathscr{E}_x}{\partial t^2} \tag{2–45}$$

This is the well-known wave equation of mathematical physics. Let us now solve this equation for steady-state sinusoidal excitation using the rms-phasor method.[8]

Rewriting Eqs. (2–44) and (2–45) in phasor form yields

$$\frac{dE_x}{dz} = -j\omega\mu_R \mu_0 H_y \quad , \quad -\frac{dH_y}{dz} = j\omega\epsilon_R \epsilon_0 E_x \tag{2–46}$$

and

$$\frac{d^2 E_x}{dz^2} = -\omega^2 \mu_R \mu_0 \epsilon_R \epsilon_0 E_x \tag{2–47}$$

where each differentiation with respect to time introduced a multiplying factor $j\omega$. Since Eq. (2–47) is a second-order differential equation, it has two independent solutions, which may be written as

$$E_x = E_0^+ e^{-j\beta z} + E_0^- e^{+j\beta z} \tag{2–48}$$

where $\beta \equiv \dfrac{\omega}{v}$, $v = \dfrac{1}{\sqrt{\mu_R \mu_0 \epsilon_R \epsilon_0}}$ and $\omega = 2\pi f = 2\pi/T$.

Substitution into the first of Eqs. (2–46) yields the two associated solutions for H_y.

$$H_y = H_0^+ e^{-j\beta z} - H_0^- e^{+j\beta z} \tag{2–49}$$

where $\dfrac{E_0^+}{H_0^+} = \dfrac{E_0^-}{H_0^-} \equiv \eta = \sqrt{\dfrac{\mu_R \mu_0}{\epsilon_R \epsilon_0}}$.

To reconstruct the instantaneous form of the solutions, multiply the rms-phasor solutions by $\sqrt{2}\, e^{j\omega t}$ and take the real parts thereof. Thus,

$$\mathscr{E}_x = \sqrt{2}\, E_0^+ \cos(\omega t - \beta z) + \sqrt{2}\, E_0^- \cos(\omega t + \beta z) \tag{2–50}$$

and

$$\mathscr{H}_y = \sqrt{2}\, H_0^+ \cos(\omega t - \beta z) - \sqrt{2}\, H_0^- \cos(\omega t + \beta z) \tag{2–51}$$

[7] The reduction of the differential form of Maxwell's equations is found in most books on electromagnetic theory (for example, Chapter 11 in Ref. 2–4). An excellent treatment of the integral form and its simplification is given in Chapter 12 of Ref. 2–1.

[8] The rms-phasor is described in Sec. 1–4.

To understand the nature of the above equations, consider the first term in Eq. (2–50). Since $\beta = \omega/v$, it can be rewritten as $\sqrt{2}\,E_0^+ \cos[\omega(t - z/v)]$. At $z = 0$, the electric field is given by $\sqrt{2}\,E_0^+ \cos \omega t$ and is plotted in Fig. 2–11. This curve does *not* represent a wave, but merely the time variation of $\mathscr{E}_x$ at one place, the $z = 0$ plane. A wave implies the movement of a time function from one place to another. The term $\sqrt{2}\,E_0^+ \cos[\omega(t - z/v)]$ does represent a wave since its cosinusoidal variation at $z = 0$ is repeated at the plane $z = l$ with a time delay l/v. Thus it appears that the $\mathscr{E}_x$ time function has traveled a distance l with a velocity v. The conclusion is that the first term in Eq. (2–50) represents an electric wave traveling in the positive z direction with a velocity v. By a similar argument, the second term describes a wave traveling in the negative z direction. From ac theory, phase delay is merely normalized time delay multiplied by 2π rad (that is, $2\pi t_d/T$). Since $\beta z = (2\pi/T)(z/v)$, it represents a phase delay for the forward traveling wave ($+z$ direction). The quantity β is the phase shift per unit length and is known as the *phase constant* of the wave.

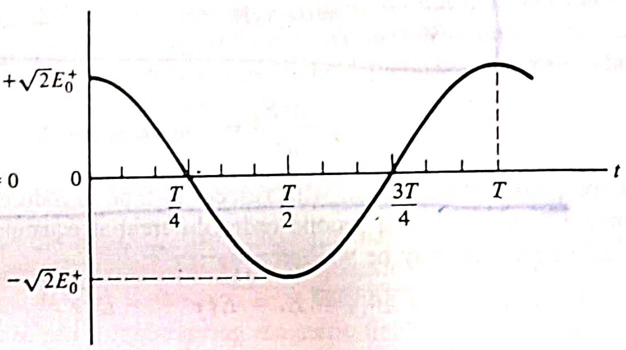


Figure 2–11 $\mathscr{E}_x$ at $z = 0$ as a function of time.

The wavelength λ and period T are related by the wave velocity. Wavelength is defined as the distance one must traverse for a function, periodic in space but fixed in time, to repeat itself. This quantity is described in Fig. 2–10 for the first term in Eq. (2–50) at $t = 0$. Period, on the other hand, is the time required for a function, periodic in time but at a fixed point in space, to repeat itself. It is described in Fig. 2–11. If the field pattern shown in Fig. 2–10 moves in the $+z$ direction with a velocity v, we have a wave. An observer situated at, say, $z = \lambda$ will see a cosinusoidal variation of $\mathscr{E}_x$ with time. When $t = 0$, its value will be a positive maximum. The time required for the next positive maximum, presently located at $z = 0$, to arrive at $z = \lambda$ represents one period. Thus, $T = \lambda/v$. Since $T = 1/f$, this leads to the familiar relationship for all wave propagation

$$f\lambda = v \tag{2–52}$$

For the electromagnetic wave described here, v is given by the expression below Eq. (2–48). With

$$\mu_0 = 4\pi \times 10^{-7} \text{ H/m} \qquad \text{and} \qquad \epsilon_0 \approx \frac{1}{36\pi} \times 10^{-9} \text{ F/m},$$

$$v = \frac{c}{\sqrt{\mu_R \epsilon_R}} \tag{2-53}$$

where $c \approx 3 \times 10^8$ m/s is the velocity of light in free space. If the wavelength in free space is denoted by λ_0,

$$\boxed{\lambda_0 = \frac{c}{f}} \tag{2-54}$$

and hence for any other insulator,

$$\boxed{\lambda = \frac{\lambda_0}{\sqrt{\mu_R \epsilon_R}}} \tag{2-55}$$

Keep in mind that for nonmagnetic insulators $\mu_R = 1.00$. Equations (2–53) and (2–55) show that both the velocity and wavelength of an electric wave in a dielectric are smaller than their values in free space. This effect is useful in reducing the size of microwave components. Also, since the phase constant $\beta = \omega/v$,

$$\beta = \frac{2\pi}{\lambda} = \frac{2\pi}{\lambda_0}\sqrt{\mu_R \epsilon_R} \qquad \text{rad/length} \tag{2-56}$$

Equation (2–51) reveals that the traveling waves also contain magnetic field components having the same velocity as the electric field. This is understandable because the two fields generate each other. Also, the ratio of their magnitudes is a constant. The relationship is given below Eq. (2–49). The ratio η is called the *intrinsic impedance* of the medium. Note that it is a real number since a lossless insulator has been assumed. This means that the traveling electric and magnetic waves are in phase. Substituting in the MKS values of μ_0 and ϵ_0 yields the equation

$$\eta = 120\pi \sqrt{\frac{\mu_R}{\epsilon_R}} = 377 \sqrt{\frac{\mu_R}{\epsilon_R}} \qquad \text{ohms} \tag{2-57}$$

Note that for free space, the intrinsic impedance is 377 ohms.

Figure 2–12 shows a sketch of the forward traveling electromagnetic wave described by the first terms in Eqs. (2–50) and (2–51). The peak values are $\sqrt{2}\,E_0^+$ and $\sqrt{2}\,H_0^+$ and therefore E_0^+ and H_0^+ are the rms values. Keep in mind that since $\mathscr{E}_x$ and $\mathscr{H}_y$ are independent of x and y, the $\mathscr{E}$-$\mathscr{H}$ pattern shown in the figure is the same along any other line parallel to the z axis. For this reason, the wave is called a *uniform plane wave*.

Figure 2–12 shows the electromagnetic wave at two instances in time, namely, $t = 0$ and $t = T/4$. Since it is traveling in the $+z$ direction, the second wave is merely the wave at $t = 0$ displaced one quarter of a wavelength in the positive z direction. Note that the electric and magnetic fields are perpendicular to each other and both lie in a plane transverse to the direction of propagation. For this reason the uniform plane wave is called a *transverse electromagnetic* (TEM) wave. The fact that

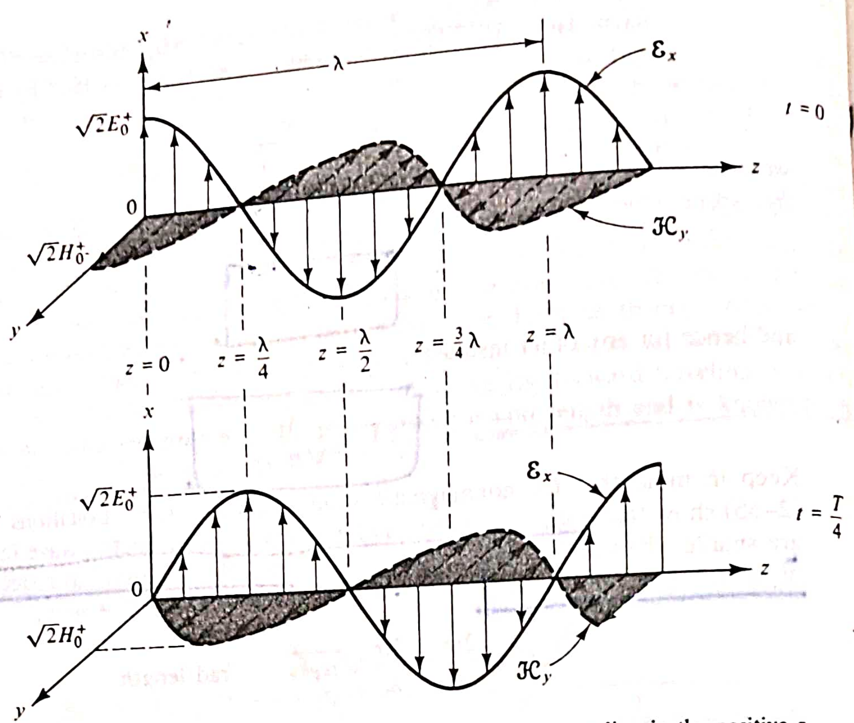


Figure 2–12 Description of an electromagnetic wave traveling in the positive z direction (shown at $t = 0$ and $t = T/4$).

the electric and magnetic fields are perpendicular to each other is a direct consequence of Faraday's and Ampere's laws. To understand this, consider the electric field at $z = \lambda/4$ in Fig. 2–12. Note that when $t = 0$, $\mathcal{E}_x = 0$ at that point. However, as the wave travels to the right, $\mathcal{E}_x$ immediately starts to increase to some finite positive value. This means that its time rate of change is positive and hence the displacement current is in the positive x direction when $t = 0$. In other words, although $\mathcal{E}_x = 0$ at $t = 0$, its time derivative is at a positive maximum. Ampere's law states that a line integral of $\mathcal{H}$ enclosing this displacement current must also be positive. By applying the right-hand rule, the magnetic field when $t = 0$ must be in the positive y direction for z slightly less than $\lambda/4$ and in the negative y direction for z slightly greater than $\lambda/4$. This is indeed the case as shown. If the wave were traveling in the negative z direction, the displacement current would be reversed and hence the direction of the magnetic field would be reversed. A similar argument, utilizing Faraday's law, shows that the movement of the magnetic wave leads to a finite line integral of $\mathcal{E}$ in a plane perpendicular to $\mathcal{H}$. In developing this argument, remember that Faraday's law has a minus sign, while Ampere's law does not.

Let us now consider the power flow associated with the electromagnetic wave propagation we have been describing. Since $\mathcal{E}$ and $\mathcal{H}$ are force fields and contain stored energy, it makes sense that electromagnetic waves involve the propagation of energy. A formal derivation of the energy in an electromagnetic field results in the following vector relationship for power flow.[9]

$$\boxed{\vec{\wp} = \vec{\mathscr{E}} \times \vec{\mathscr{H}}} \tag{2-58}$$

$\vec{\wp}$ is the instantaneous vector power density (W/m^2) and is known as the *Poynting vector*. Its magnitude represents the value of instantaneous power density, while its direction indicates the direction of power flow. Because Eq. (2-58) involves the vector cross product, the direction of $\vec{\wp}$ is always perpendicular to both $\vec{\mathscr{E}}$ and $\vec{\mathscr{H}}$. In our example (Fig. 2-12), $\vec{\mathscr{E}}$ crossed into $\vec{\mathscr{H}}$ is in the positive z direction for any and all values of position and time. To reverse the direction of propagation and hence power flow, either $\vec{\mathscr{E}}$ or $\vec{\mathscr{H}}$ (not both) must be reversed. This agrees with the conclusion arrived at using the displacement current concept in combination with Faraday's and Ampere's laws.

The average power densities for the forward and reverse traveling waves described by Eqs. (2-50) and (2-51) are

$$p_z^+ = E_0^+ H_0^+ \qquad \text{and} \qquad p_z^- = E_0^- H_0^- \tag{2-59}$$

where E_0 and H_0 represent rms values. The average power flow through a surface S perpendicular to the direction of propagation is given by

$$P = \int_S p_z\, dS \qquad \text{where} \qquad p_z = p_z^+ - p_z^- \tag{2-60}$$

For uniform plane waves, p_z is independent of position and hence $P = p_z S$.

The following illustrative example is intended to reinforce the various ideas discussed in this section.

Example 2-3:

A 500 MHz electromagnetic wave is propagating through a perfect nonmagnetic dielectric having $\epsilon_R = 6$.

(a) Calculate the wavelength and the phase constant.

(b) With the wave traveling in the $+z$ direction, the sinusoidal electric field at $z = 80$ cm is delayed relative to the field at $z = 65$ cm. Calculate the time delay (in nanoseconds) and the phase delay (in degrees).

(c) Calculate the average power density in the wave if the peak value of magnetic field is 0.5 A/m.

$\lambda = \frac{\lambda_0}{\sqrt{\mu_R \epsilon_R}}$, $\beta = \frac{2\pi}{\lambda}$

Solution:

(a) $\lambda_0 = c/f = (3 \times 10^8)/(500 \times 10^6) = 0.6$ m.
Since $\mu_R = 1$, $\lambda = 0.6/\sqrt{6} = 0.245$ m
and $\beta = 2\pi/0.245 = 25.65$ rad/m or 1470°/m.

(b) $t_d = \Delta z/v$, where $v = 3 \times 10^8/\sqrt{6} = 1.22 \times 10^8$ m/s.
Thus, $t_d = (0.80 - 0.65)/v = 1.23 \times 10^{-9}$ s or 1.23 ns.
Phase Delay $= \beta \Delta z = 1470(0.15) = 220°$.

(c) With $\mu_R = 1$, $\eta = 377/\sqrt{6} = 154$ ohms.
$p_z^+ = E_0^+ H_0^+ = \eta (H_0^+)^2$, where $H_0^+ = 0.5/\sqrt{2} = 0.354$ A/m.
Therefore, $p_z^+ = 154(0.354)^2 = 19.3$ W/m^2.

[9] See, for example, Chapter 12 in Ref. 2-1 or Chapter 4 in Ref. 2-6.

2-5 WAVE POLARIZATION

There are four types of wave polarization: linear, circular, elliptic, and random. The wave described in Fig. 2–12 is an example of a linear or plane polarized wave. By convention, the direction of electric field is used to denote the polarization. Therefore the wave in Fig. 2–12 is *vertically polarized* because the electric field is always and everywhere in the vertical direction. A horizontally polarized wave would, of course, be one in which the E lines are horizontal. Linear polarized waves are thus characterized by the fact that the orientation of the field is the same everywhere in space and is independent of time.

It is very useful to describe an electromagnetic wave in terms of phasor quantities. Figure 2–13 shows the phasor representation at five points along the propagation axis for the forward traveling wave depicted in Fig. 2–12. The rms-phasor representation of $\mathscr{E}_x$ at $z = 0$ is given by $E_0^+\angle 0$. With the wave traveling in the $+z$ direction, E at $z = l$ will be phase delayed βl. For instance at $z = \lambda/4$, the phase delay is $(2\pi/\lambda)(\lambda/4) = \pi/2$ rad and hence $\mathrm{E} = E_0^+\angle -\pi/2$. For a given propagation direction, knowledge of E at one point defines its value at all other points along the propagation axis. The magnetic field phasors are also shown in the figure. As explained previously, the electric and magnetic fields are perpendicular to each other, have the same velocity and are related by the intrinsic impedance η. Note that at any point, E and H are in phase, which is the case for η real.

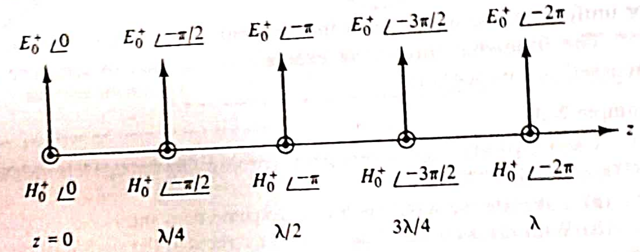


Figure 2–13 Rms-phasor representation of a linear polarized wave.

Since we are dealing with sinusoidal *vector* fields, the quantities in the figure are actually *vector phasors*. They can be expressed mathematically with the aid of unit vectors ($\vec{a}$). The vector-phasor representation for the electric and magnetic fields in Fig. 2–13 are $\vec{E} = \vec{a}_x E$ and $\vec{H} = \vec{a}_y H$, where $\vec{a}_x$ and $\vec{a}_y$ are the unit vectors in the x and y directions. For example, at $z = \lambda/4$

$$\vec{E} = \vec{a}_x E_0^+ \angle -\pi/2 = \vec{a}_x E_0^+ e^{-j\pi/2} \quad \text{and} \quad \vec{H} = \vec{a}_y H_0^+ \angle -\pi/2 = \vec{a}_y H_0^+ e^{-j\pi/2}$$

This method of describing an electromagnetic wave is a very useful analytical tool. In most cases, it is sufficient to specify the electric field and η, since $H = E/\eta$ and, for a given propagation direction, the magnetic field direction can be deduced from Eq. (2–58).

The rules of vector decomposition may be applied to vector phasors as illustrated by the following example.

Chapter

5

Polarization

5.1 Introduction

In the *far field zone* of a transmitting antenna, the radiated wave takes on the characteristics of a *transverse electromagnetic* (TEM) wave. Far field zone refers to distances greater than $2D^2/\lambda$ from the antenna, where D is the largest linear dimension of the antenna and λ is the wavelength. For a parabolic antenna of 3 m diameter transmitting a 6-GHz wave ($\lambda = 5$ cm), the far field zone begins at approximately 360 m. The TEM designation is illustrated in Fig. 5.1, where it can be seen that both the magnetic field **H** and the electric field **E** are transverse to the direction of propagation, denoted by the propagation vector **k**.

E, **H**, and **k** represent vector quantities, and it is important to note their relative directions. When one looks along the direction of propagation, the rotation from **E** to **H** is in the direction of rotation of a right-hand-threaded screw, and the vectors are said to form a *right-hand set*. The wave always retains the directional properties of the right-hand set, even when reflected, for example. One way of remembering how the right-hand set appears is to note that the letter E comes before H in the alphabet and rotation is from **E** to **H** when looking along the direction of propagation.

At great distances from the transmitting antenna, such as are normally encountered in radio systems, the TEM wave can be considered to be plane. This means that the **E** and **H** vectors lie in a plane, which is at right angles to the vector **k**. The vector **k** is said to be normal to the plane. The magnitudes are related by $E = HZ_0$, where $Z_0 = 120\pi\Omega$.

The direction of the line traced out by the tip of the electric field vector determines the *polarization* of the wave. Keep in mind that the electric and magnetic fields are varying as functions of time. The magnetic field varies exactly in phase with the electric field, and its amplitude is proportional to the electric field amplitude, so it is only necessary

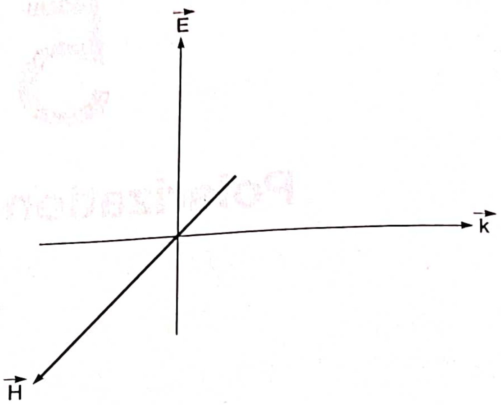


Figure 5.1 Vector diagram for a transverse electromagnetic (TEM) wave.

to consider the electric field in this discussion. The tip of the **E** vector may trace out a straight line, in which case the polarization is referred to as *linear.* Other forms of polarization, specifically elliptical and circular, will be introduced later.

In the early days of radio, there was little chance of ambiguity in specifying the direction of polarization in relation to the surface of the earth. Most transmissions utilized linear polarization and were along terrestrial paths. Thus *vertical polarization* meant that the electric field was perpendicular to the earth's surface, and *horizontal polarization* meant that it was parallel to the earth's surface. Although the terms vertical and horizontal are used with satellite transmissions, the situation is not quite so clear. A linear polarized wave transmitted by a geostationary satellite may be designated vertical if its electric field is parallel to the earth's polar axis, but even so the electric field will be parallel to the earth at the equator. This situation will be clarified shortly.

Suppose for the moment that horizontal and vertical are taken as the x and y axes of a right-hand set, as shown in Fig. 5.2*a*. A vertically polarized electric field can be described as

$$\mathbf{E}_y = \hat{a}_y E_y \sin \omega t \tag{5.1}$$

where $\hat{a}_y$ is the unit vector in the vertical direction and E_y is the *peak value* or *amplitude* of the electric field. Likewise, a horizontally polarized wave could be described by

$$\mathbf{E}_x = \hat{a}_x E_x \sin \omega t \tag{5.2}$$

These two fields would trace out the straight lines shown in Fig. 5.2*b*. Now consider the situation where both fields are present simultaneously.

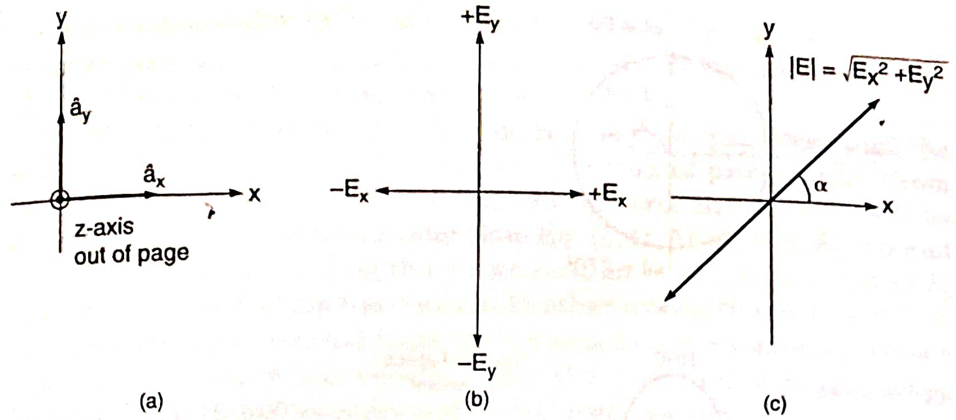


Figure 5.2 Horizontal and vertical components of linear polarization.

These would add vectorially, and the resultant would be a vector **E** (Fig. 5.2*c*) of amplitude $\sqrt{E_x^2 + E_y^2}$, at an angle to the horizontal given by

$$\alpha = \arctan \frac{E_y}{E_x} \tag{5.3}$$

E varies sinusoidally in time in the same manner as the individual components. It is still linearly polarized but cannot be classified as simply horizontal or vertical. Arguing back from this, it is evident that **E** can be resolved into vertical and horizontal components, a fact which is of great importance in practical transmission systems. The power in the resultant wave is proportional to the voltage $\sqrt{E_x^2 + E_y^2}$, squared, which is $E_x^2 + E_y^2$. In other words, the power in the resultant wave is the sum of the powers in the individual waves, which is to be expected.

More formally, $\mathbf{E}_y$ and $\mathbf{E}_x$ are said to be *orthogonal*. The dictionary definition of orthogonal is at *right angles*, but a wider meaning will be attached to the word later.

Consider now the situation where the two fields are equal in amplitude (denoted by E), but one leads the other by 90° in phase. The equations describing these are

$$\mathbf{E}_y = \hat{a}_y E \sin \omega t \tag{5.4a}$$

$$\mathbf{E}_x = \hat{a}_x E \cos \omega t \tag{5.4b}$$

Applying Eq. (5.3) in this case yields $\alpha = \omega t$. The tip of the resultant electric field vector traces out a circle, as shown in Fig. 5.3*a*, and the resultant wave is said to be *circularly polarized*. The amplitude of the resultant vector is E. The resultant field in this case does not go through zero. At $\omega t = 0$, the y component is zero and the x component is E. At

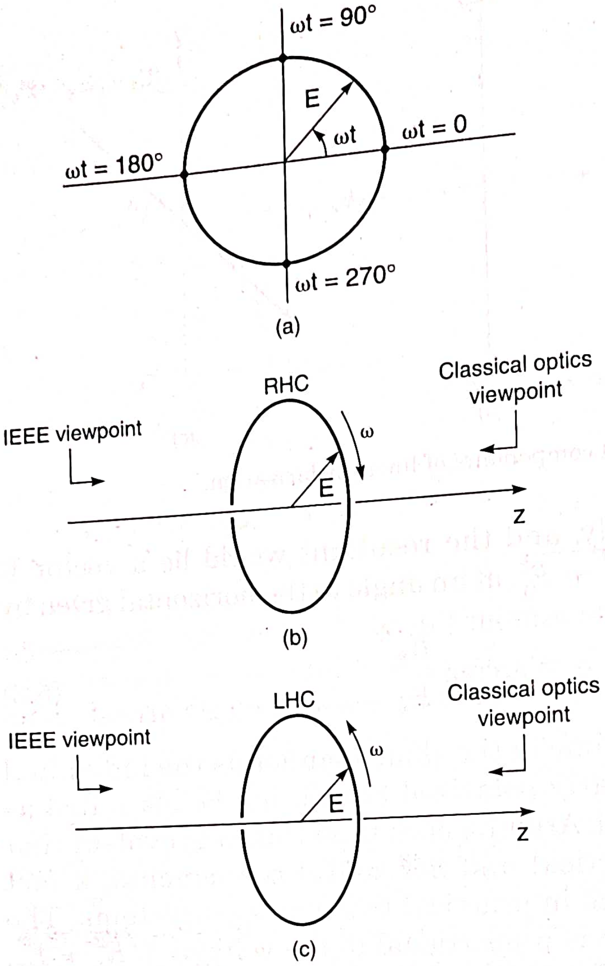


Figure 5.3 Circular polarization.

$\omega t = 90°$, the y component is E and the x component is zero. Compare this with the linear polarized case where at $\omega t = 0$, both the x and y components are zero, and at $\omega t = 90°$, both components are maximum at E. Because the resultant does not vary in time, the power must be found by adding the powers in the two linear polarized, sinusoidal waves. This gives a resultant proportional to $2E^2$.

The direction of circular polarization is defined by the sense of rotation of the electric vector, but this also requires that the way the vector is viewed must be specified. The *Institute of Electrical and Electronics Engineers* (IEEE) defines *right-hand circular* (RHC) *polarization* as a rotation in the clockwise direction when the wave is viewed along the direction of propagation, that is, when viewed from "behind," as shown in Fig. 5.3*b*. *Left-hand circular* (LHC) *polarization* is when the rotation is in the counterclockwise direction when viewed along the direction of propagation, as shown in Fig. 5.3*c*. LHC and RHC polarizations are orthogonal. The direction of propagation is along the $+z$ axis.

As a caution it should be noted that the classical optics definition of circular polarization is just the opposite of the IEEE definition. The IEEE definition will be used throughout this text.

For a right-hand set of axes (Fig. 5.1) and with propagation along the $+z$ axis, then when viewed along the direction of propagation (from "behind") and with the $+y$ axis directed upward, the $+x$ axis will be directed toward the left. Consider now Eq. (5.4). At $\omega t = 0$, E_y is 0 and E_x is a maximum at E along the $+x$ axis. At $\omega t = 90°$, E_x is zero and E_y is a maximum at E along the $+y$ axis. In other words, the resultant field of amplitude E has rotated from the $+x$ axis to the $+y$ axis, which is a clockwise rotation when viewed along the direction of propagation. Equation (5.4) therefore represents RHC polarization.

Given that Eq. (5.4) represents RHC polarization, it is left as an exercise to show that the following equations represent LHC polarization:

$$\mathbf{E}_y = \hat{a}_y E \sin\omega t \tag{5.5a}$$

$$\mathbf{E}_x = -\hat{a}_x E \cos\omega t \tag{5.5b}$$

In the more general case, a wave may be *elliptically polarized.* This occurs when the two linear components are

$$\mathbf{E}_y = \hat{a}_y E_y \sin\omega t \tag{5.6a}$$

$$\mathbf{E}_x = \hat{a}_x E_x \sin(\omega t + \delta) \tag{5.6b}$$

a6

Here, E_y and E_x are not equal in general, and δ is a fixed phase angle. It is left as an exercise for the student to show that when $E_y = 1$, $E_x = 1/3$, and $\delta = 30°$, the polarization ellipse is as shown in Fig. 5.4.

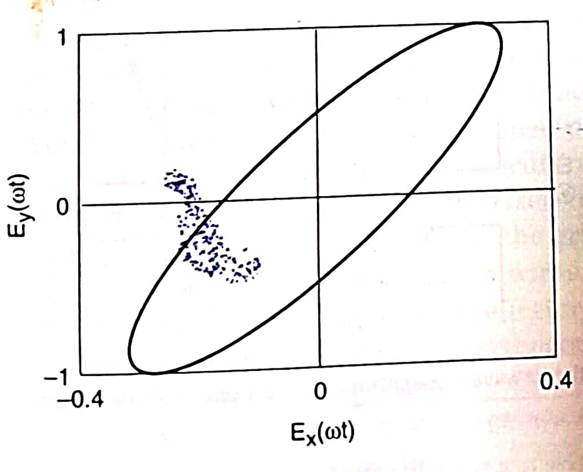


Figure 5.4 Elliptical polarization.

Likewise rewriting (68),

$$H_T = 2H_i \cos \beta x \cos (\omega t) \tag{5-70}$$

Comparison of (69) and (70) shows that E_T and H_T differ in time phase by $\pi/2$ radians or 90 degrees.

5.10 Reflection by a Perfect Conductor—Oblique Incidence. Whenever a wave is incident obliquely on the interface between two media, it is necessary to consider separately two special cases. The first of these is the case in which the electric vector is parallel to the boundary surface or perpendicular to the plane of incidence. (The plane of incidence is the plane containing the incident ray and the normal to the surface.) This case is often termed *horizontal polarization*. In the second case the magnetic vector is parallel to the boundary surface, and the electric vector is parallel to the plane of incidence. This case is often termed *vertical polarization*. The two cases are shown in Fig. 5-8. The terms "horizontally and vertically

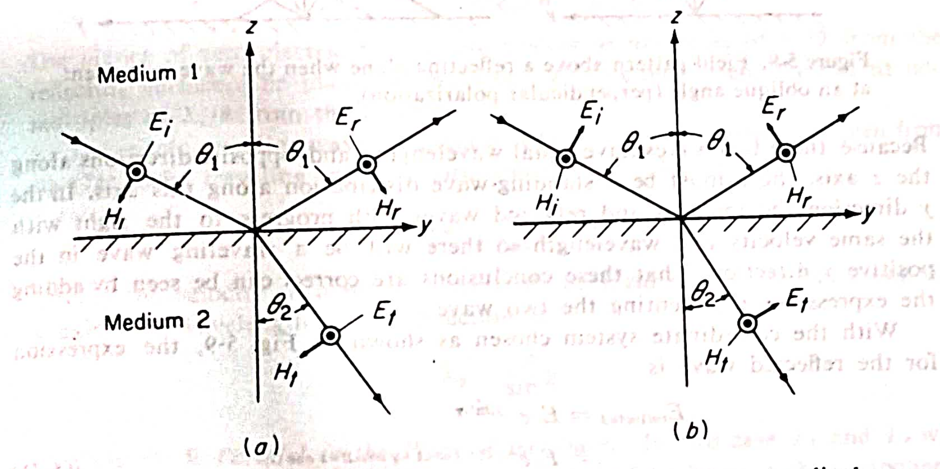


Figure 5-8. Reflection and refraction of waves having (*a*) perpendicular (horizontal) polarization and (*b*) parallel (vertical) polarization.

polarized waves" refer to the fact that waves from horizontal and vertical antennas, respectively, would produce these particular orientations of electric and magnetic vectors in waves striking the surface of the earth. However, it is seen that, whereas the electric vector of a "horizontally" polarized wave is horizontal, the electric vector of a "vertically" polarized wave is not wholly vertical but has some horizontal component. More significant designations are the terms "perpendicular" and "parallel" polarization to indicate that the electric vector is perpendicular or parallel to the plane of incidence. In wave-guide work the terms *transverse electric* (TE) and *transverse magnetic*

15

Radio-wave Propagation

15.1 Introduction

Radio communications use electromagnetic waves propagated through the earth's atmosphere or space to carry information over long distances without the use of wires. Radio waves with frequencies ranging from about 100 Hz in the ELF band to well above 300 GHz in the EHF band have been used for communications purposes, and more recently radiation in and near the visible range (near 1000 THz, or 10^{15} Hz) have also been used. Figure 15.8.1 shows the frequency-band designations in common use.

Some of the basic properties of a *transverse electromagnetic* (TEM) wave are described in Appendix B. Although the electric and magnetic fields exist simultaneously, in practice, antennas are designed to work through one or other of these fields. Antennas are described in Chapter 16. Basically, to launch an electromagnetic wave into space, an electric charge has to be accelerated, which in practice means that the current in the radiator must change with time (for example, be alternating). In this chapter, sinusoidal or cosinusoidal variations will be assumed unless stated otherwise.

15.2 Propagation in Free Space

Mode of Propagation

Consider first an average power P_T, assumed to be radiated equally in all directions (isotropically). This will spread out spherically as it travels away from the source, so that at distance d, the power density in the wave, which is the power per unit area of wavefront, will be power density

$$P_{Di} = \frac{P_T}{4\pi d^2} \quad \text{W/m}^2 \tag{15.2.1}$$

This is so because $4\pi d^2$ is the surface area of the sphere of radius d, centered on the source. P_{Di} stands for isotropic power density.

It is known that all practical antennas have directional characteristics; that is, they radiate more power in some directions at the expense of less in others. The directivity gain is the ratio of actual power density along the main axis of radiation of the antenna to that which would be produced by an isotropic antenna at the same distance fed with the same input power. Let G_T be the *maximum* directivity gain of the transmitting antenna; then the power density along the direction of maximum radiation will be

$G_T = \frac{P_t}{P_{ci}}$

$$\begin{aligned} P_D &= P_{Di} G_T \\ &= \frac{P_T G_T}{4\pi d^2} \end{aligned} \tag{15.2.2}$$

A receiving antenna can be positioned so that it collects maximum power from the wave. When so positioned, let P_R be the power delivered by the antenna to the load (receiver) under matched conditions; then the antenna can be considered as having an effective area (or aperture) A_{eff}, where

$$\begin{aligned} P_R &= P_D A_{eff} \\ &= \frac{P_T G_T}{4\pi d^2} A_{eff} \end{aligned} \tag{15.2.3}$$

It can be shown that for any antenna, the ratio of maximum directivity gain to effective area is

$G_R = \frac{4\pi A_{eff}}{\lambda^2}$

$$\frac{A_{eff}}{G} = \frac{\lambda^2}{4\pi} \tag{15.2.4}$$

Here, λ is the wavelength of the wave being radiated. Letting G_R be the maximum directivity gain of the receiving antenna, we have

$$\frac{P_R}{P_T} = G_T G_R \left(\frac{\lambda}{4\pi d}\right)^2 \tag{15.2.5}$$

This is the fundamental equation for free-space transmission. Usually it is expressed in terms of frequency f, in megahertz, and distance d, in kilometers. As shown in Appendix B, $\lambda f = c$, and on substituting this in Eq. (15.2.5) and doing the arithmetic, which is left as an exercise for the reader, the result obtained is

$$\frac{P_R}{P_T} = G_T G_R \frac{0.57 \times 10^{-3}}{(df)^2} \tag{15.2.6}$$

By expressing the power ratios in decibels, Eq. (15.2.6) can be written as

$$\left(\frac{P_R}{P_T}\right)_{\text{dB}} = (G_T)_{\text{dB}} + (G_R)_{\text{dB}} - (32.5 + 20 \log_{10} d + 20 \log_{10} f) \tag{15.2.7}$$

The third term in parentheses on the right-hand side of Eq. (15.2.7) is the loss, in decibels, resulting from the spreading of the wave as it propagates outward from the source. It is known as the transmission path loss, L. Thus

$$L = (32.5 + 20 \log_{10} d + 20 \log_{10} f)_{\text{dB}} \tag{15.2.8}$$

where d is in kilometers and f in megahertz.

Equation (15.2.7) then becomes

$$\left(\frac{P_R}{P_T}\right)_{\text{dB}} = (G_T)_{\text{dB}} + (G_R)_{\text{dB}} - (L)_{\text{dB}} \tag{15.2.9}$$

EXAMPLE 15.2.1

In a satellite communications system, free-space conditions may be assumed. The satellite is at a height of 36,000 km above earth, the frequency used is 4000 MHz, the transmitting antenna gain is 15 dB, and the receiving antenna gain is 45 dB. Calculate (a) the free-space transmission loss and (b) the received power when the transmitted power is 200 W.

SOLUTION (a) $L = 32.5 + 20 \log_{10} 36{,}000 + 20 \log_{10} 4000$

$= \mathbf{196\ dB}$

(b) $\left(\frac{P_R}{P_T}\right)_{\text{dB}} = 15 + 45 - 196$

$= -136$ dB

This is a power ratio of 0.25×10^{-13}, and since $P_T = 200$ W,

$$P_R = 200 \times 0.25 \times 10^{-13}$$
$$= 5 \times 10^{-12}\ \text{W} = \mathbf{5\ pW}$$

Frequently, it is required to know the electric field strength of the wave at the receiving antenna. In Appendix B, E is given by Eq. (B.12) in terms of power density P_D and wave impedance Z_0 as

$$E = \sqrt{Z_0 P_D} \tag{15.2.10}$$

Also, Z_0 is given by Eq. (B.10) as

$$Z_0 = \sqrt{\frac{\mu}{\varepsilon}} \tag{15.2.11}$$

The free-space values are as follows: $\mu = \mu_0 = 4\pi \times 10^{-7}$ H/m, and $\varepsilon = \varepsilon_0 = 8.854 \times 10^{-12}$ F/m. Substituting these in Eq. (15.2.11) gives

$$Z_0 = 120\pi\ \Omega \tag{15.2.12}$$

The field strength may now be found by substituting Eqs. (15.2.2) and (15.2.12) in Eq. (15.2.10)

$$E = \frac{\sqrt{30 P_T G_T}}{d} \quad V/m \tag{15.2.13}$$

This is the fundamental equation that gives the field strength at the receiving antenna, for free-space propagation conditions. A receiving antenna has an effective length ℓ_{eff} (analogous to effective area) such that the open circuit EMF of the antenna E_s is given by

$$V_s = E\ell_{\text{eff}} \tag{15.2.14}$$

Effective length is discussed in Section 16.9.

EXAMPLE 15.2.2

Calculate the open-circuit voltage induced in a $\frac{1}{2}\lambda$ dipole when 10 W at 150 MHz is radiated from another $\frac{1}{2}\lambda$ dipole 50 km distant. The antennas are positioned for optimum transmission and reception.

SOLUTION

$$\lambda = \frac{3 \times 10^8}{150 \times 10^6}$$

$$= 2 \text{ m}$$

In Chapter 16 it is shown that, for a $\frac{1}{2}\lambda$ dipole, the maximum gain is 1.64 : 1, and the effective length is λ/π. Therefore,

$$V_s = \frac{\sqrt{30 \times 10 \times 1.64}}{50 \times 10^3} \cdot \frac{2}{\pi}$$

$$= \mathbf{282\ \mu V}$$

Equation (15.2.13) is sometimes expressed in terms of the field strength at unit distance, E_0. Thus, at $d = 1$ m,

$$E = E_0 = \frac{\sqrt{30 P_T G_T}}{1} \quad \text{V/m} \tag{15.2.15}$$

and therefore Eq. (15.2.13) can be written as

$$E = \frac{E_0}{d} \quad \text{V/m} \tag{15.2.16}$$

Note that both E and E_0 are in units of V/m, although Eq. (15.2.16) may tend to suggest that E was in units of (V/m)/m, or V/m^2. Equation (15.2.16) really expresses a proportionality, and written in full it would be

$$E = E_0 \times \frac{1\ (\text{m})}{d\ (\text{m})} \quad \text{V/m} \tag{15.2.17}$$

Microwave Systems

Microwave radio systems operating at frequencies above 1 GHz propagate mainly in a line-of-sight or free-space mode, whether they are on the ground or in satellite systems. Since the 1950s, microwave radio systems have become the workhorses of long-distance telephone communications systems. These systems provide the needed transmission bandwidth and reliability to allow the transmission of many thousands of telephone channels as well as several television channels over the same route and using the same facilities. Carrier frequencies in the 3- to 12-GHz range are used. Since microwaves travel only on line-of-sight paths, it is necessary to provide repeater stations at about 50-km intervals. This makes the equipment costs for such a system very large, but this is more than made up for by the increased channel capacity. Transmitter output powers are low (they may be less than 1 W), because highly directional high-gain antennas are used.

Figure 15.2.1(a) shows the equipment needed to provide one channel of a microwave system. It consists of two terminal stations and one or more repeater stations. At the sending terminal, the inputs comprising several hundred telephone channels and/or a television channel are frequency-multiplexed within the baseband band pass of 6 MHz. The baseband frequency modulates a 70-MHz IF signal, which is then up-converted to the microwave output frequency f_1 within the 4-GHz band. This signal is amplified and fed through a directional antenna toward a repeater station some 50 km distant. At the repeater station, the signal at f_1 is received on one antenna pointed toward the originating station, down-converted to the IF, amplified, and up-converted to a new frequency f_2 for retransmission toward the receiving terminal station. When the signal is passed through a chain of several repeaters, alternate links in the chain use alternate frequencies so that retransmitted energy at a repeater station does not feed back into its own receiver.

With two frequencies in use, alternating at each repeater, two channels, one in each direction, can be provided, as illustrated in Fig. 15.2.1(b). In some microwave systems several two-way channel pairs are provided, and a more complex system of frequency switching is used at the repeater stations.

At the receiving terminal station, the signal is down-converted to the IF and then demodulated to recover the baseband signal. This baseband signal

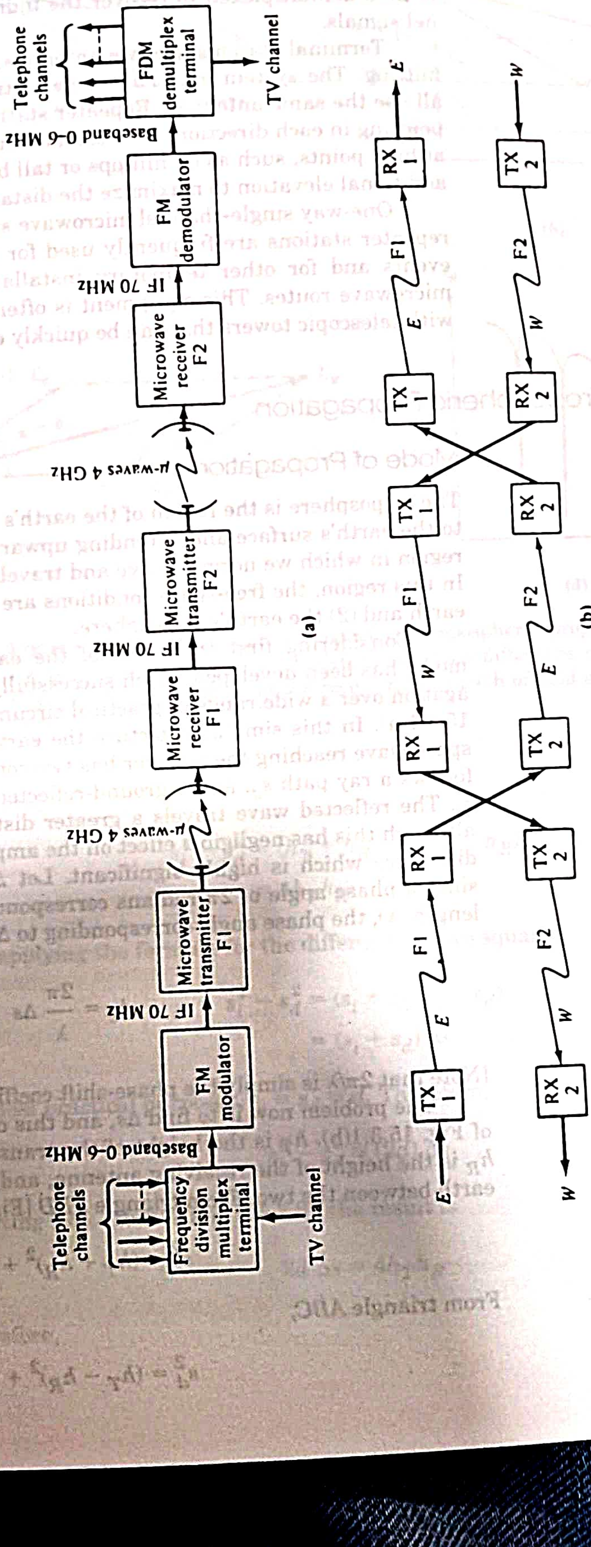


Figure 15.2.1 Microwave relay system. (a) One-way channel showing the equipment used on the route. (b) Two-way channel pair showing frequency interchanges at intermediate repeater stations.

is then demultiplexed to recover the individual telephone or television channel signals.

Terminal stations use two antennas, one for receiving and one for transmitting. The system may have several transmitters and receivers, but they all use the same antennas. Repeater stations are provided with two antennas pointing in each direction, for a total of four. Repeater station sites are chosen at high points, such as on hilltops or tall buildings, and sturdy towers provide additional elevation to maximize the distance between stations.

One-way single-channel microwave systems using portable terminal and repeater stations are frequently used for remote television pickup of special events and for other temporary installations, such as the testing of new microwave routes. This equipment is often mounted in mobile vans outfitted with telescopic towers that can be quickly erected.

15.3 Tropospheric Propagation

Mode of Propagation

The troposphere is the region of the earth's atmosphere immediately adjacent to the earth's surface and extending upward for some tens of kilometers (the region in which we normally live and travel, including high-flying jet planes). In this region, the free-space conditions are modified by (1) the surface of the earth and (2) the earth's atmosphere.

Considering first the effect of the earth's surface, a much-simplified model has been developed which successfully describes electromagnetic propagation over a wide range of practical circumstances and is illustrated in Fig. 15.3.1(a). In this simplified picture the earth is assumed to be flat, and the space wave reaching the receiver has two components: the direct wave, which follows a ray path s_d, and a ground-reflected wave, which follows a ray path s_i. The reflected wave travels a greater distance than the direct wave, and although this has negligible effect on the amplitude, it does introduce a phase difference which is highly significant. Let Δs be the path difference; then, since a phase angle of 2π radians corresponds to a path length of one wavelength (λ), the phase angle corresponding to Δs is

$$\phi_s = \frac{2\pi}{\lambda}\Delta s \qquad (15.3.1)$$

[Note that $2\pi/\lambda$ is simply the phase-shift coefficient introduced in Eq. (13.5.4)].

The problem now is to find Δs, and this can be solved from the geometry of Fig. 15.3.1(b). h_T is the height of the transmitting antenna above ground, h_R is the height of the receiving antenna, and d is the distance along the flat earth between the two. From triangle FBD [Fig.15.3.1(b)],

$$s_i^2 = (h_T + h_R)^2 + d^2$$

From triangle ABC,

$$s_d^2 = (h_T - h_R)^2 + d^2$$

will double the field strength at the receiving point. Note also that the field strength is proportional to E_0, which, in turn, is proportional to the square root of the transmitted power. Doubling the power results in only a $\sqrt{2}$ increase in field strength.

Equations (15.3.9) and (15.3.10) also apply when a transmission takes place from the h_R antenna and is received at the h_T antenna: it is only necessary to change E_0 to the new value determined by the gain of the antenna at h_R and the power transmitted from there.

As the distance d increases, it becomes necessary to take into account the curvature of the earth. Reflection from the curved surface reduces both the amplitude of the reflected wave and the phase difference. These two effects tend to offset each other, and the resultant amplitude does not vary rapidly as a result.

Radio Horizon

The curvature of the earth has a more important effect in that it presents a horizon that limits the range of transmission. This range is greater than the optical range because the effect of the earth's atmosphere is to cause a bending of the radio wave, which carries it beyond the optical horizon. Figure 15.3.2(a) shows a typical radio-wave ray path, and Fig. 15.3.2(b) shows how the path can be considered straight by assigning a greater radius to the earth than it actually has. For standard atmospheric conditions the increase in radius has been worked out at $\frac{4}{3}$, so that

$$a' = \frac{4}{3}a \tag{15.3.11}$$

where a is the earth's actual radius and a' is the fictitious radius that accounts for refraction. From Fig. 15.3.2(b),

$$(a')^2 + d_1^2 = (a' + h_T)^2$$

Therefore,

$$d_1^2 = 2a'h_T + h_T^2 \tag{15.3.12}$$

But, since $a' \gg h_T$,

$$d_1^2 \approx 2a'h_T \tag{15.3.13}$$

Similarly,

$$d_2^2 \approx 2a'h_R \tag{15.3.14}$$

The maximum radio range d_{max} is

$$\begin{aligned} d_{max} &\approx d_1 + d_2 \\ &= \sqrt{2a'h_T} + \sqrt{2a'h_R} \end{aligned} \tag{15.3.15}$$

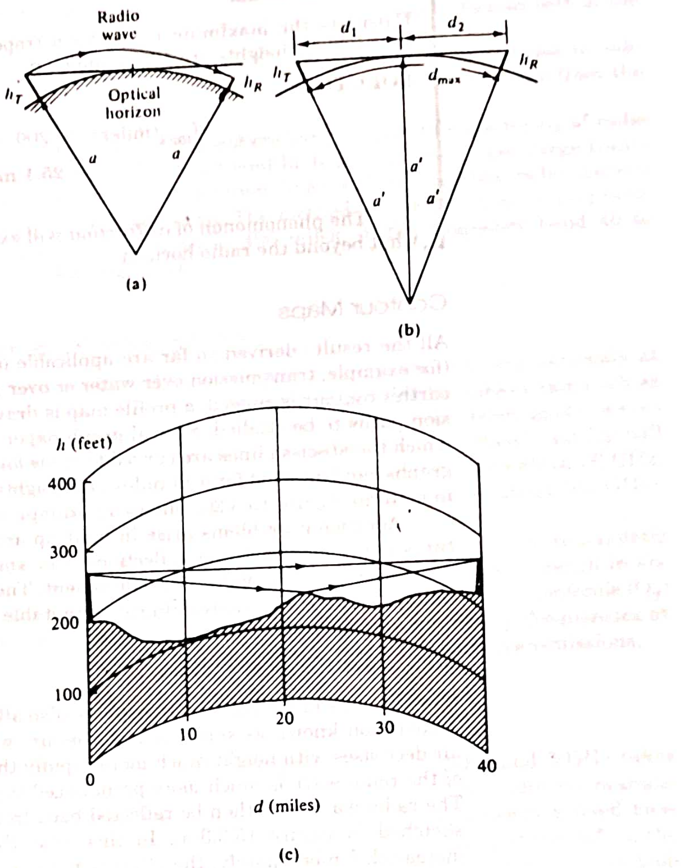


Figure 15.3.2 (a) Curvature of ray path resulting from change of refractive index of the air. (b) Equivalent straight line ray path for fictitious earth radius a'. (c) Example of a contour map for radio path planning.

Substituting in the numerical values, $a' = \frac{4}{3} \times 3960$ miles, and expressing h_T and h_R in feet results in the useful expression

$$d_{max}(\text{miles}) = \sqrt{2h_T\,(\text{ft})} + \sqrt{2h_R\,(\text{ft})} \tag{15.3.16}$$

Alternatively, in metric units,

$$d_{max}(\text{km}) = \sqrt{17h_T\,(m)} + \sqrt{17h_R\,(\text{m})} \tag{15.3.17}$$

EXAMPLE 15.3.2

Calculate the maximum range for a tropospheric transmission for which the antenna heights are 100 ft and 60 ft.

SOLUTION

$$d_{max}(\text{miles}) = \sqrt{200} + \sqrt{120}$$
$$= \mathbf{25.1\ miles}$$

The phenomenon of *diffraction* will extend the range in practice somewhat beyond the radio horizon.

Contour Maps

All the results derived so far are applicable only for smooth earth conditions (for example, transmission over water or over reasonably flat land). Where the earth's contour is rugged, a profile map is drawn to enable proposed transmission paths to be studied. Special graph paper is available for this purpose, in which the abscissa lines are curved to allow for the fictitious radius a', and the graphs can be scaled for d in miles and heights in feet (or d in km and heights in meters). Figure 15.3.2(c) shows an example of a profile map.

Additional problems arise in built-up areas, where buildings and structures can cause multiple reflections and shielding, which are particularly troublesome with mobile radio equipment. The only satisfactory solution is to conduct field trials to determine an acceptable site for the base station.

Super- and Subrefractions

Irregularities in the earth's atmosphere also affect tropospheric transmissions. A condition known as *superrefraction* occurs when the refractive index of the air decreases with height much more rapidly than normal, so that the bending of the radio wave is much more pronounced than that shown in Fig. 15.3.2(a). The radio wave may then be reflected back from the earth to follow a path as sketched in Figure 15.3.3(a). In this way, the radio range is considerably increased. Unfortunately, the effect is not sufficiently reliable for it to be utilized for commercial communications systems, but it does account for some of the abnormally long distance interference that has been observed at VHF.

An increase of temperature with height (known as temperature inversion) gives rise to superrefraction, as does an increase of humidity with height.

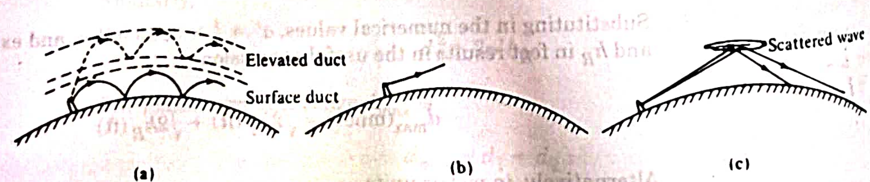


Figure 15.3.3 (a) Superrefraction. (b) Subrefraction. (c) Tropospheric scatter propagation.

4.2 Free Space Propagation Model

The free space propagation model is used to predict received signal strength when the transmitter and receiver have a clear, unobstructed line-of-sight path between them. Satellite communication systems and microwave line-of-sight radio links typically undergo free space propagation. As with most large-scale radio wave propagation models, the free space model predicts that received power decays as a function of the T–R separation distance raised to some power (i.e. a power law function). The free space power received by a receiver antenna which is separated from a radiating transmitter antenna by a distance d, is given by the Friis free space equation,

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{4.1}$$

where P_t is the transmitted power, $P_r(d)$ is the received power which is a function of the T–R separation, G_t is the transmitter antenna gain, G_r is the receiver antenna gain, d is the T–R separation distance in meters, L is the system loss factor not related to propagation ($L \geq 1$), and λ is the wavelength in meters. The gain of an antenna is related to its effective aperture, A_e, by

$$G = \frac{4\pi A_e}{\lambda^2} \tag{4.2}$$

The effective aperture A_e is related to the physical size of the antenna, and λ is related to the carrier frequency by

$$\lambda = \frac{c}{f} = \frac{2\pi c}{\omega_c} \tag{4.3}$$

where f is the carrier frequency in Hertz, ω_c is the carrier frequency in radians per second, and c is the speed of light given in meters/s. The values for P_t and P_r must be expressed in the same units, and G_t and G_r are dimensionless quantities. The miscellaneous losses L ($L \geq 1$) are usually due to transmission line attenuation, filter losses, and antenna losses in the communication system. A value of $L = 1$ indicates no loss in the system hardware.

The Friis free space equation of (4.1) shows that the received power falls off as the square of the T–R separation distance. This implies that the received power decays with distance at a rate of 20 dB/decade.

An *isotropic* radiator is an ideal antenna which radiates power with unit gain uniformly in all directions, and is often used to reference antenna gains in wireless systems. The *effective isotropic radiated power (EIRP)* is defined as

$$EIRP = P_t G_t \tag{4.4}$$

and represents the maximum radiated power available from a transmitter in the direction of maximum antenna gain, as compared to an isotropic radiator.

In practice, *effective radiated power (ERP)* is used instead of EIRP to denote the maximum radiated power as compared to a half-wave dipole antenna (instead of an isotropic antenna). Since a dipole antenna has a gain of 1.64 (2.15 dB above an isotropic), the ERP will be

2.15 dB smaller than the EIRP for the same transmission system. In practice, antenna gains are given in units of dBi (dB gain with respect to an isotropic antenna) or dBd (dB gain with respect to a half-wave dipole) [Stu81].

The *path loss*, which represents signal attenuation as a positive quantity measured in dB, is defined as the difference (in dB) between the effective transmitted power and the received power, and may or may not include the effect of the antenna gains. The path loss for the free space model when antenna gains are included is given by

$$PL(\mathrm{dB}) = 10\log\frac{P_t}{P_r} = -10\log\left[\frac{G_t G_r \lambda^2}{(4\pi)^2 d^2}\right] \tag{4.5}$$

When antenna gains are excluded, the antennas are assumed to have unity gain, and path loss is given by

$$PL(\mathrm{dB}) = 10\log\frac{P_t}{P_r} = -10\log\left[\frac{\lambda^2}{(4\pi)^2 d^2}\right] \tag{4.6}$$

The Friis free space model is only a valid predictor for P_r for values of d which are in the far-field of the transmitting antenna. The far-field, or *Fraunhofer region*, of a transmitting antenna is defined as the region beyond the far-field distance d_f, which is related to the largest linear dimension of the transmitter antenna aperture and the carrier wavelength. The Fraunhofer distance is given by

$$d_f = \frac{2D^2}{\lambda} \tag{4.7.a}$$

where D is the largest physical linear dimension of the antenna. Additionally, to be in the far-field region, d_f must satisfy

$$d_f \gg D \tag{4.7.b}$$

and

$$d_f \gg \lambda \tag{4.7.c}$$

Furthermore, it is clear that Equation (4.1) does not hold for $d = 0$. For this reason, large-scale propagation models use a close-in distance, d_0, as a known received power reference point. The received power, $P_r(d)$, at any distance $d > d_0$, may be related to P_r at d_0. The value $P_r(d_0)$ may be predicted from Equation (4.1), or may be measured in the radio environment by taking the average received power at many points located at a close-in radial distance d_0 from the transmitter. The reference distance must be chosen such that it lies in the far-field region, that is, $d_0 \geq d_f$, and d_0 is chosen to be smaller than any practical distance used in the mobile communication system. Thus, using Equation (4.1), the received power in free space at a distance greater than d_0 is given by

$$P_r(d) = P_r(d_0)\left(\frac{d_0}{d}\right)^2 \qquad d \geq d_0 \geq d_f \tag{4.8}$$

In mobile radio systems, it is not uncommon to find that P_r may change by many orders of magnitude over a typical coverage area of several square kilometers. Because of the large dynamic range of received power levels, often dBm or dBW units are used to express received power levels. Equation (4.8) may be expressed in units of dBm or dBW by simply taking the logarithm of both sides and multiplying by 10. For example, if P_r is in units of dBm, the received power is given by

$$P_r(d)\ \text{dBm} = 10\log\left[\frac{P_r(d_0)}{0.001\ \text{W}}\right] + 20\log\left(\frac{d_0}{d}\right) \qquad d \geq d_0 \geq d_f \tag{4.9}$$

where $P_r(d_0)$ is in units of watts.

The reference distance d_0 for practical systems using low-gain antennas in the 1–2 GHz region is typically chosen to be 1 m in indoor environments and 100 m or 1 km in outdoor environments, so that the numerator in Equations (4.8) and (4.9) is a multiple of 10. This makes path loss computations easy in dB units.

Example 4.1

Find the far-field distance for an antenna with maximum dimension of 1 m and operating frequency of 900 MHz.

Solution

Given:

Largest dimension of antenna, $D = 1$ m

Operating frequency $f = 900$ MHz, $\lambda = c/f = \dfrac{3 \times 10^8\ \text{m/s}}{900 \times 10^6\ \text{Hz}}$ m

Using Equation (4.7.a), far-field distance is obtained as

$$d_f = \frac{2(1)^2}{0.33} = 6\ \text{m}$$

Example 4.2

If a transmitter produces 50 W of power, express the transmit power in units of (a) dBm, and (b) dBW. If 50 W is applied to a unity gain antenna with a 900 MHz carrier frequency, find the received power in dBm at a free space distance of 100 m from the antenna. What is P_r(10 km)? Assume unity gain for the receiver antenna.

Solution

Given:

Transmitter power, $P_t = 50$ W

Carrier frequency, $f_c = 900$ MHz

Using Equation (4.9),

(a) Transmitter power,

$$P_t(\text{dBm}) = 10\log[P_t(\text{mW})/(1\ \text{mW})]$$
$$= 10\log[50 \times 10^3] = 47.0\ \text{dBm}.$$

(b) Transmitter power,

$$P_t(\text{dBW}) = 10\log[P_t(\text{W})/(1\ \text{W})]$$
$$= 10\log[50] = 17.0\ \text{dBW}.$$

The received power can be determined using Equation (4.1)

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} = \frac{50(1)(1)(1/3)^2}{(4\pi)^2(100)^2(1)} = (3.5 \times 10^{-6})\ \text{W} = 3.5 \times 10^{-3}\ \text{mW}$$

$$P_r(\text{dBm}) = 10\log P_r(\text{mW}) = 10\log(3.5 \times 10^{-3}\ \text{mW}) = -24.5\ \text{dBm}.$$

The received power at 10 km can be expressed in terms of dBm using Equation (4.9), where $d_0 = 100$ m and $d = 10$ km

$$P_r(10\ \text{km}) = P_r(100) + 20\log\left[\frac{100}{10000}\right] = -24.5\ \text{dBm} - 40\ \text{dB}$$

$P_r + 20\log\left(\frac{d_0}{d}\right)$

$$= -64.5\ \text{dBm}.$$

4.3 Relating Power to Electric Field

The free space path loss model of Section 4.2 is readily derived from first principles. It can be proven that any radiating structure produces electric and magnetic fields [Gri87], [Kra50]. Consider a small linear radiator of length L, that is placed coincident with the z-axis and has its center at the origin, as shown in Figure 4.2.

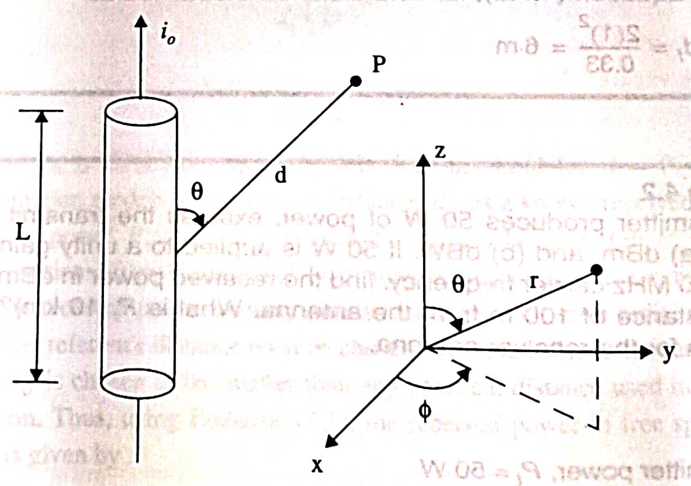


Figure 4.2 Illustration of a linear radiator of length $L (L \ll \lambda)$, carrying a current of amplitude i_0 and making an angle θ with a point, at distance d.

If a current flows through such an antenna, it launches electric and magnetic fields that can be expressed as

$$E_r = \frac{i_0 L\cos\theta}{2\pi\varepsilon_0 c}\left\{\frac{1}{d^2}+\frac{c}{j\omega_c d^3}\right\}e^{j\omega_c(t-d/c)} \tag{4.10}$$

$$E_\theta = \frac{i_0 L\sin\theta}{4\pi\varepsilon_0 c^2}\left\{\frac{j\omega_c}{d}+\frac{c}{d^2}+\frac{c^2}{j\omega_c d^3}\right\}e^{-j\omega_c(t-d/c)} \tag{4.11}$$

$$H_\phi = \frac{i_0 L\sin\theta}{4\pi c}\left\{\frac{j\omega_c}{d}+\frac{c}{d^2}\right\}e^{j\omega_c(t-d/c)} \tag{4.12}$$

with $E_\phi = H_r = H_\theta = 0$. In the above equations, all $1/d$ terms represent the radiation field component, all $1/d^2$ terms represent the induction field component, and all $1/d^3$ terms represent the electrostatic field component. As seen from Equations (4.10) to (4.12), the electrostatic and inductive fields decay much faster with distance than the radiation field. At regions far away from the transmitter (far-field region), the electrostatic and inductive fields become negligible and only the radiated field components of E_θ and H_ϕ need be considered.

In free space, the *power flux density* P_d (expressed in W/m^2) is given by

$$P_d = \frac{EIRP}{4\pi d^2} = \frac{P_t G_t}{4\pi d^2} = \frac{E^2}{R_{fs}} = \frac{E^2}{\eta}\ \text{W/m}^2 \tag{4.13}$$

where R_{fs} is the intrinsic impedance of free space given by $\eta = 120\pi\Omega$ (377 Ω). Thus, the power flux density is

$$P_d = \frac{|E|^2}{377\Omega}\ \text{W/m}^2 \tag{4.14}$$

where $|E|$ represents the magnitude of the radiating portion of the electric field in the far field. Figure 4.3a illustrates how the power flux density disperses in free space from an isotropic point source. P_d may be thought of as the *EIRP* divided by the surface area of a sphere with radius d. The power received at distance d, $P_r(d)$, is given by the power flux density times the effective aperture of the receiver antenna, and can be related to the electric field using Equations (4.1), (4.2), (4.13), and (4.14).

$$P_r(d) = P_d A_e = \frac{|E|^2}{120\pi}A_e = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2} = \frac{|E|^2 G_r \lambda^2}{480\pi}\ \text{W} \tag{4.15}$$

Equation (4.15) relates electric field (with units of V/m) to received power (with units of watts), and is identical to Equation (4.1) with $L = 1$.

Often it is useful to relate the received power level to a receiver input voltage, as well as to an induced E-field at the receiver antenna. If the receiver antenna is modeled as a matched resistive load to the receiver, then the receiver antenna will induce an rms voltage into the receiver

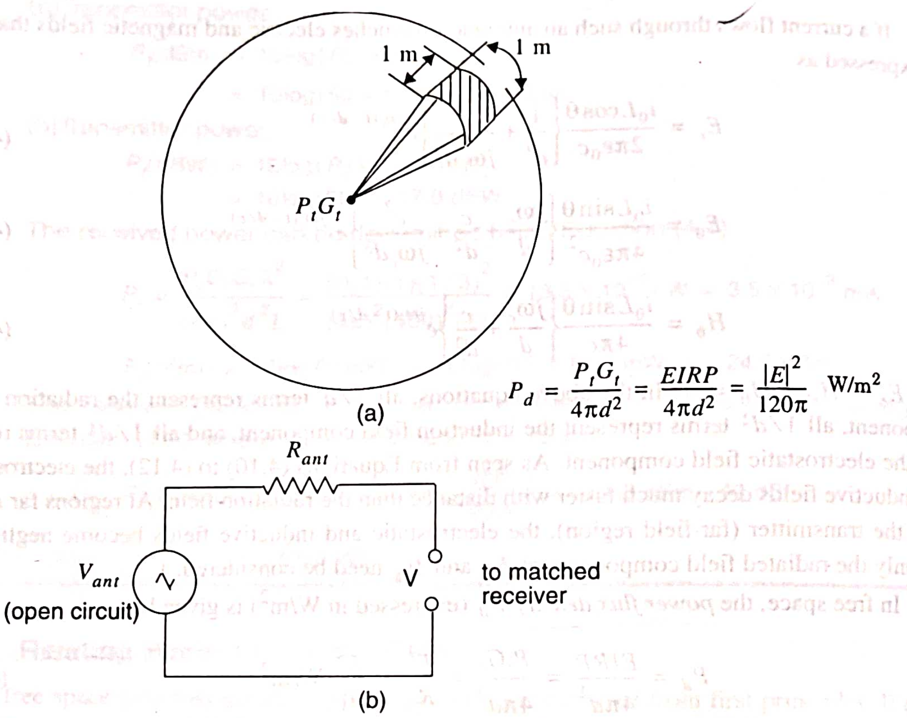


Figure 4.3 (a) Power flux density at a distance *d* from a point source; (b) model for voltage applied to the input of a receiver.

which is half of the open circuit voltage at the antenna. Thus, if V is the rms voltage at the input of a receiver (measured by a high impedance voltmeter), and R_{ant} is the resistance of the matched receiver, the received power is given by

$$P_r(d) = \frac{V^2}{R_{ant}} = \frac{[V_{ant}/2]^2}{R_{ant}} = \frac{V_{ant}^2}{4R_{ant}} \tag{4.16}$$

Through Equations (4.14) to (4.16), it is possible to relate the received power to the received E-field or the open circuit rms voltage at the receiver antenna terminals. Figure 4.3b illustrates an equivalent circuit model. Note $V_{ant} = V$ when there is no load.

Example 4.3

Assume a receiver is located 10 km from a 50 W transmitter. The carrier frequency is 900 MHz, free space propagation is assumed, $G_t = 1$, and $G_r = 2$, find (a) the power at the receiver, (b) the magnitude of the E-field at the receiver antenna, (c) the rms voltage applied to the receiver input assuming that the receiver antenna has a purely real impedance of 50 Ω and is matched to the receiver.

Solution

Given:

Transmitter power, $P_t = 50$ W

Carrier frequency, $f_c = 900$ MHz

Transmitter antenna gain, $G_t = 1$

Receiver antenna gain, $G_r = 2$

Receiver antenna resistance $= 50\ \Omega$

(a) Using Equation (4.5), the power received at distance $d = 10$ km is

$$P_r(d) = 10\log\left(\frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2}\right) = 10\log\left(\frac{50 \times 1 \times 2 \times (1/3)^2}{(4\pi)^2 10000^2}\right)$$

$$= -91.5 \text{ dBW} = -61.5 \text{ dBm}$$

(b) Using Equation (4.15), the magnitude of the received E-field is

$$|E| = \sqrt{\frac{P_r(d)120\pi}{A_e}} = \sqrt{\frac{P_r(d)120\pi}{G_r\lambda^2/4\pi}} = \sqrt{\frac{7 \times 10^{-10} \times 120\pi}{2 \times 0.33^2/(4\pi)}} = 0.0039 \text{ V/m}$$

(c) Using Equation (4.16), the applied rms voltage at the receiver input is

$$V_{ant} = \sqrt{P_r(d) \times R_{ant}} = \sqrt{7 \times 10^{-10} \times 50} = 0.187 \text{ mV}$$

4.4 The Three Basic Propagation Mechanisms

Reflection, diffraction, and scattering are the three basic propagation mechanisms which impact propagation in a mobile communication system. These mechanisms are briefly explained in this section, and propagation models which describe these mechanisms are discussed subsequently in this chapter. Received power (or its reciprocal, path loss) is generally the most important parameter predicted by large-scale propagation models based on the physics of reflection, scattering, and diffraction. Small-scale fading and multipath propagation (discussed in Chapter 5) may also be described by the physics of these three basic propagation mechanisms.

Reflection occurs when a propagating electromagnetic wave impinges upon an object which has very large dimensions when compared to the wavelength of the propagating wave. Reflections occur from the surface of the earth and from buildings and walls.

Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a surface that has sharp irregularities (edges). The secondary waves resulting from the obstructing surface are present throughout the space and even behind the obstacle, giving rise to a bending of waves around the obstacle, even when a line-of-sight path does not exist between transmitter and receiver. At high frequencies, diffraction, like reflection, depends on the geometry of the object, as well as the amplitude, phase, and polarization of the incident wave at the point of diffraction.

Scattering occurs when the medium through which the wave travels consists of objects with dimensions that are small compared to the wavelength, and where the number of obstacles per unit volume is large. Scattered waves are produced by rough surfaces, small objects, or by other irregularities in the channel. In practice, foliage, street signs, and lamp posts induce scattering in a mobile communications system.

4.5 Reflection

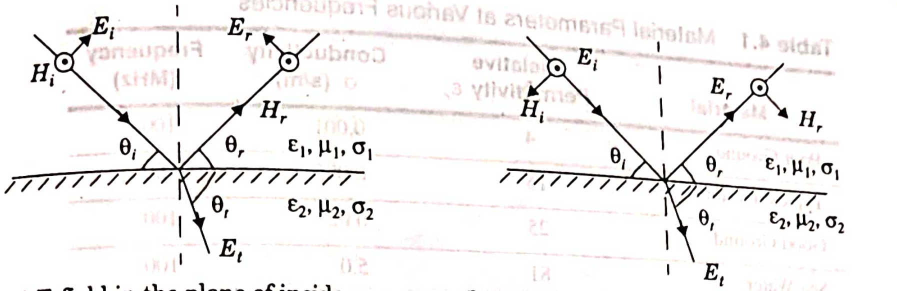
When a radio wave propagating in one medium impinges upon another medium having different electrical properties, the wave is partially reflected and partially transmitted. If the plane wave is incident on a perfect dielectric, part of the energy is transmitted into the second medium and part of the energy is reflected back into the first medium, and there is no loss of energy in absorption. If the second medium is a perfect conductor, then *all* incident energy is reflected back into the first medium without loss of energy. The electric field intensity of the reflected and transmitted waves may be related to the incident wave in the medium of origin through the *Fresnel reflection coefficient* (Γ). The reflection coefficient is a function of the material properties, and generally depends on the wave polarization, angle of incidence, and the frequency of the propagating wave.

In general, electromagnetic waves are *polarized*, meaning they have instantaneous electric field components in orthogonal directions in space. A polarized wave may be mathematically represented as the sum of two spatially orthogonal components, such as vertical and horizontal, or left-hand or right-hand circularly polarized components. For an arbitrary polarization, superposition may be used to compute the reflected fields from a reflecting surface.

4.5.1 Reflection from Dielectrics

Figure 4.4 shows an electromagnetic wave incident at an angle θ_i with the plane of the boundary between two dielectric media. As shown in the figure, part of the energy is reflected back to the first media at an angle θ_r, and part of the energy is transmitted (refracted) into the second media at an angle θ_t. The nature of reflection varies with the direction of polarization of the E-field. The behavior for arbitrary directions of polarization can be studied by considering the two distinct cases shown in Figure 4.4. The *plane of incidence* is defined as the plane containing the incident, reflected, and transmitted rays [Ram65]. In Figure 4.4a, the E-field polarization is parallel with the plane of incidence (that is, the E-field has a vertical polarization, or normal component, with respect to the reflecting surface) and in Figure 4.4b, the E-field polarization is perpendicular to the plane of incidence (that is, the incident E-field is pointing out of the page toward the reader, and is perpendicular to the page and parallel to the reflecting surface).

In Figure 4.4, the subscripts *i, r, t* refer to the incident, reflected, and transmitted fields, respectively. Parameters $\varepsilon_1, \mu_1, \sigma_1$, and $\varepsilon_2, \mu_2, \sigma_2$ represent the permittivity, permeability, and conductance of the two media, respectively. Often, the dielectric constant of a perfect (lossless) dielectric is related to a relative value of permittivity, ε_r, such that $\varepsilon = \varepsilon_0\varepsilon_r$, where ε_0 is a



(a) E-field in the plane of incidence (b) E-field normal to the plane of incidence

Figure 4.4 Geometry for calculating the reflection coefficients between two dielectrics.

constant given by 8.85×10^{-12} F/m. If a dielectric material is lossy, it will absorb power and may be described by a complex dielectric constant given by

$$\varepsilon = \varepsilon_0 \varepsilon_r - j\varepsilon' \tag{4.17}$$

where

$$\varepsilon' = \frac{\sigma}{2\pi f} \tag{4.18}$$

and σ is the conductivity of the material measured in Siemens/meter. The terms ε_r and σ are generally insensitive to operating frequency when the material is a good conductor $(f < \sigma/(\varepsilon_0 \varepsilon_r))$. For lossy dielectrics, ε_0 and ε_r are generally constant with frequency, but σ may be sensitive to the operating frequency, as shown in Table 4.1. Electrical properties of a wide range of materials were characterized over a large frequency range by Von Hipple [Von54].

Because of superposition, only two orthogonal polarizations need be considered to solve general reflection problems. The reflection coefficients for the two cases of parallel and perpendicular E-field polarization at the boundary of two dielectrics are given by

$$\Gamma_{\parallel} = \frac{E_r}{E_i} = \frac{\eta_2 \sin\theta_t - \eta_1 \sin\theta_i}{\eta_2 \sin\theta_t + \eta_1 \sin\theta_i} \quad \text{(E-field in plane of incidence)} \tag{4.19}$$

$$\Gamma_{\perp} = \frac{E_r}{E_i} = \frac{\eta_2 \sin\theta_i - \eta_1 \sin\theta_t}{\eta_2 \sin\theta_i + \eta_1 \sin\theta_t} \quad \text{(E-field normal to the plane of incidence)} \tag{4.20}$$

where η_i is the intrinsic impedance of the i th medium ($i = 1, 2$), and is given by $\sqrt{\mu_i/\varepsilon_i}$, the ratio of electric to magnetic field for a uniform plane wave in the particular medium. The velocity of an electromagnetic wave is given by $1/(\sqrt{\mu\varepsilon})$, and the boundary conditions at the surface of incidence obey Snell's Law which, referring to Figure 4.4, is given by

$$\sqrt{\mu_1 \varepsilon_1} \sin(90 - \theta_i) = \sqrt{\mu_2 \varepsilon_2} \sin(90 - \theta_t) \tag{4.21}$$

Table 4.1 Material Parameters at Various Frequencies

Material	Relative Permittivity ε_r	Conductivity σ (s/m)	Frequency (MHz)
Poor Ground	4	0.001	100
Typical Ground	15	0.005	100
Good Ground	25	0.02	100
Sea Water	81	5.0	100
Fresh Water	81	0.001	100
Brick	4.44	0.001	4000
Limestone	7.51	0.028	4000
Glass, Corning 707	4	0.00000018	1
Glass, Corning 707	4	0.000027	100
Glass, Corning 707	4	0.005	10000

The boundary conditions from Maxwell's equations are used to derive Equations (4.19) and (4.20) as well as Equations (4.22), (4.23.a), and (4.23.b).

$$\theta_i = \theta_r \tag{4.22}$$

and

$$E_r = \Gamma E_i \tag{4.23.a}$$

$$E_t = (1+\Gamma)E_i \tag{4.23.b}$$

where Γ is either $\Gamma_{\parallel}$ or $\Gamma_{\perp}$, depending on whether the E-field is in (vertical) or normal (horizontal) to the plane of incidence.

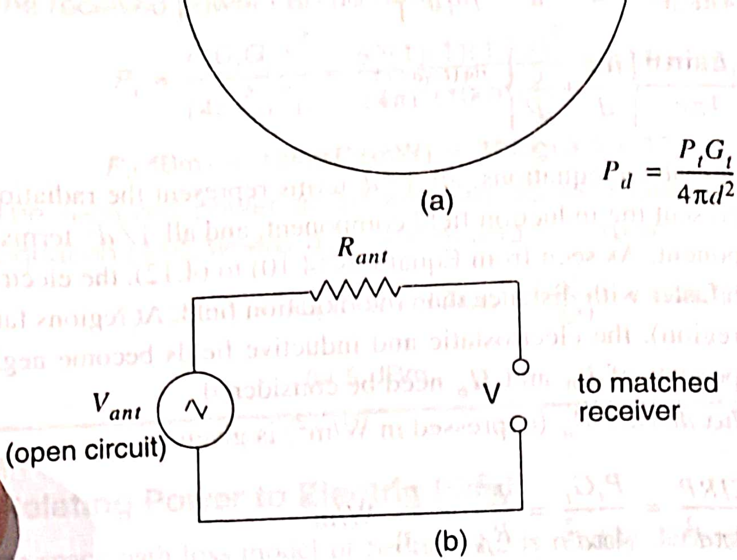
For the case when the first medium is free space and $\mu_1 = \mu_2$, the reflection coefficients for the two cases of vertical and horizontal polarization can be simplified to

$$\Gamma_{\parallel} = \frac{-\varepsilon_r \sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}}{\varepsilon_r \sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}} \tag{4.24}$$

and

$$\Gamma_{\perp} = \frac{\sin\theta_i - \sqrt{\varepsilon_r - \cos^2\theta_i}}{\sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}} \tag{4.25}$$

For the case of elliptical polarized waves, the wave may be broken down (depolarized) into its vertical and horizontal E-field components, and superposition may be applied to determine transmitted and reflected waves. In the general case of reflection or transmission, the horizontal and vertical



$$P_d = \frac{P_t G_t}{4\pi d^2} = \frac{EIRP}{4\pi d^2} = \frac{|E|^2}{120\pi} \ \text{W/m}^2$$

Figure 4.3 (a) Power flux density at a distance d from a point source; (b) model for voltage applied to the input of a receiver.

which is half of the open circuit voltage at the antenna. Thus, if V is the rms voltage at the input of a receiver (measured by a high impedance voltmeter), and R_{ant} is the resistance of the matched receiver, the received power is given by

$$P_r(d) = \frac{V^2}{R_{ant}} = \frac{[V_{ant}/2]^2}{R_{ant}} = \frac{V_{ant}^2}{4R_{ant}} \tag{4.16}$$

Through Equations (4.14) to (4.16), it is possible to relate the received power to the received E-field or the open circuit rms voltage at the receiver antenna terminals. Figure 4.3b illustrates an equivalent circuit model. Note $V_{ant} = V$ when there is no load.

Example 4.3

Assuming free space propagation, a receiver is located 10 km away from a 50 W transmitter. The carrier frequency is 900 MHz, antenna gain at transmitter end receiver is 1 and 2, respectively, calculate

(a) power received at the receiver
(b) the magnitude of the E-field at the receiver antenna

(c) the power flux density
(d) the rms voltage applied to the receiver input. The receiver antenna has 50 Ω impedance and is matched to the receiver.

Solution

Given:

Transmitter power, $P_t = 50$ W
Carrier frequency, $f_c = 900$ MHz
Transmitter antenna gain, $G_t = 1$
Receiver antenna gain, $G_r = 2$
Receiver antenna resistance = 50 Ω

(a) Using Equation (4.5), the power received at distance $d = 10$ km is

$$P_r(d) = 10\log\left[\frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2}\right] = 10\log\left[\frac{50 \times 1 \times 2 \times (1/3)^2}{(4\pi)^2 10000^2}\right]$$

$$= -91.5 \text{ dBW} = -61.5 \text{ dBm}$$

(b) Using Equation (4.15), the magnitude of the received E-field is

$$|E| = \sqrt{\frac{P_r(d)120\pi}{A_e}} = \sqrt{\frac{P_r(d)120\pi}{G_r\lambda^2/4\pi}} = \sqrt{\frac{7\times10^{-10}\times120\pi}{2\times0.33^2/(4\pi)}} = 0.0039 \text{ V/m}$$

(c) Power flux density, $P_d = (P_t G_t)/(4\pi)^2 d^2 = 50 \times 1/(4 \times 3.14)^2 \times 10000 = 0.0316$ mW

(d) Using Equation (4.16), the open circuit rms voltage at the receiver input is

$$V_{ant} = \sqrt{P_r(d) \times 4R_{ant}} = \sqrt{7\times10^{-10} \times 4 \times 50} = 0.374 \text{ mV}$$

4.4 The Three Basic Propagation Mechanisms

Reflection, diffraction, and scattering are the three basic propagation mechanisms which impact propagation in a mobile communication system. These mechanisms are briefly explained in this section, and propagation models which describe these mechanisms are discussed subsequently in this chapter. Received power (or its reciprocal, path loss) is generally the most important parameter predicted by large-scale propagation models based on the physics of reflection, scattering, and diffraction. Small-scale fading and multipath propagation (discussed in Chapter 5) may also be described by the physics of these three basic propagation mechanisms.

Reflection occurs when a propagating electromagnetic wave impinges upon an object which has very large dimensions when compared to the wavelength of the propagating wave. Reflections occur from the surface of the earth and from buildings and walls.

Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a surface that has sharp irregularities (edges). The secondary waves resulting from the obstructing surface are present throughout the space and even behind the obstacle, giving rise to a bending of waves around the obstacle, even when a line-of-sight path does not exist between