CHAPTER

20 Antennas and radio wave propagation

While discussing transmitter and receiver circuitry we did not have to know much about antennas or propagation. It sufficed to know only that a voltage applied to the terminals of a transmitting antenna causes a proportional voltage to appear very shortly thereafter at the terminals of a receiving antenna. To be more exact, it was sufficient to know that everything between the terminals of the two antennas is equivalent to a linear two-port network. Here we will consider the transmission through this propagation link.

When an ac source (transmitter) is connected to an antenna (practically any metal structure) the resulting current has a component that is in phase with the applied voltage. The impedance of the antenna therefore has a real part, a resistance, and draws power from the source. If the antenna is efficient, most of the power flows away from the antenna in the form of (energy-bearing) electromagnetic waves and only a small fraction of the power will be dissipated by ohmic heating of the antenna itself. The impedance will also generally have a nonzero imaginary part, a reactance. If the reactance is zero at the operating frequency the antenna is said to be resonant, just as an *RLC* circuit is purely resistive at its resonant frequency. An external tuning network (an *antenna tuner*) can be used to cancel the reactance and also transform the resistance to a value that matches a receiver's input impedance or to a value that draws a desired amount of power from a transmitter.

20.1 Electromagnetic waves

As an electromagnetic wave propagates away from the transmitting antenna, it takes on a spherical wavefront. By the time it reaches a distant receiving antenna, the wavefront has a very large radius of curvature and is essentially a plane wave. The ***E*** and ***H*** vectors (electric and magnetic fields, measured respectively in volts/m and amperes/m) both lie in the plane of the wavefront, i.e., they are transverse to the direction of propagation as shown in Figure 20.1. The fields are in phase with each other; they rise and fall together in space and time.

https://doi.org/10.1017/CBO9780511626951.021 Published online by Cambridge University Press

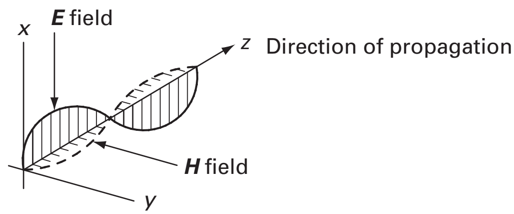


Figure 20.1. An electromagnetic wave – the ***E*** and ***H*** fields are transverse to the direction of propagation.

The ratio of the electric field strength to the magnetic field strength in free space is given by $E/H = (\mu_0/\varepsilon_0)^{1/2} = 120\pi = 377$ ohms. This ratio is known as the "impedance of free space." These fields are perpendicular to each other and their vector cross-product, $\boldsymbol{E} \times \boldsymbol{H}$, points in the direction of propagation. A continuous monochromatic wave has sinusoidal spatial components, $E_0 e^{-jkz}$ and $H_0 e^{-jkz}$, as shown in the figure. The wavenumber, k, is defined as $2\pi/\lambda$ radians per meter. The phase velocity, ω/k, is given by $c = (\mu_0\varepsilon_0)^{-1/2} = 3 \times 10^8$ m/s (the speed of light).

20.1.1 Propagation in a vacuum

Static electric and magnetic fields are always associated with sources, i.e., electric field lines terminate on electric charges and magnetic field lines encircle current filaments. But when an antenna launches an electromagnetic wave, the field lines break away from the sources, reconnecting into closed loops, and the wave becomes autonomous. For a vacuum, Maxwell's curl equations are

$$\nabla \times \boldsymbol{E} = -\mu_0 j\omega \boldsymbol{H}, \tag{20.1a}$$

$$\nabla \times \boldsymbol{H} = \varepsilon_0 j\omega \boldsymbol{E}. \tag{20.1b}$$

The right-hand sides of these equations are like sources; the E-field is produced by the changing H-field and the H-field is produced by the changing E-field. You can verify the statements made above about plane waves by writing the E-field shown in Figure 20.1 as $\boldsymbol{E} = \hat{x} E_0 e^{j(\omega t - kz)}$ and substituting it into Equation (20.1a) to get $\boldsymbol{H}$. You can then verify that $\boldsymbol{H}$ satisfies Equation (20.1b).

Electromagnetic waves are produced by electric charges undergoing acceleration. Time-varying currents contain accelerating charges. A sinusoidally time-varying current distribution on an antenna launches sinusoidal electromagnetic waves. The time-averaged power density or "energy flux" of the waves is given by $S = ½(E \times H) = ½E^2/377$ watts/m^2, where E and H are the (peak) field amplitudes and the factor ½ is the time average of $\cos^2(\omega t)$. At a receiving antenna, the fields from an incident wave produce currents that result in a voltage at the antenna terminals.

20.2 Radiation from a current element

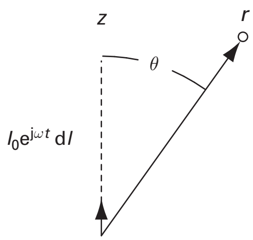


Figure 20.2. An elemental sinusoidal ac current element.

Just as Maxwell's equations can be manipulated to yield back Coulomb's law (which gives the E-field from an elementary point charge) and Ampere's law (which gives the H-field from a constant current element), they can also be stirred together to yield the E and H fields produced by a *time-varying* current element. The E and H fields produced by an antenna are superpositions of the E and H fields from every elemental ac current element on the antenna. Consider an ac current element in the z-direction, as shown in Figure 20.2.

This element has length δl and carries a current $I_0 \cos(\omega t)$ which, as usual, we express as $I_0 e^{j\omega t}$. At the observation point, $\boldsymbol{r}$, the magnetic field is given by

$$\delta H = \frac{I_0 \delta l}{4\pi}\left(\frac{jk\sin(\theta)e^{j(\omega t - kr)}}{r} + \frac{\sin(\theta)e^{j(\omega t - kr)}}{r^2}\right), \tag{20.2}$$

where $k = \omega/c$ and θ is the angle between the current element $\boldsymbol{\delta l}$ and the vector $\boldsymbol{r}$. Equation (20.2) is Ampere's law, generalized for an elemental current with a sinusoidal time variation. The direction of $\boldsymbol{\delta H}$ is given by $\boldsymbol{\delta l} \times \boldsymbol{r}$ (into the page). The first term in the bracket falls off slowly as $1/r$ and is the radiation term. The second term falls off quickly as $1/r^2$ and corresponds to near-field stored energy; j is just a phase factor since $j = e^{j\pi/2}$. The normally implicit $e^{j\omega t}$ term is included here to highlight the wave term, $e^{j\omega t - kr}$. If we let the frequency approach zero, the radiation term vanishes (since k goes to zero), the second term becomes $\sin(\theta)/r^2$ and Equation (20.2) reduces to Ampere's law (also known as the Biot – Savart law) for calculating the magnetic field produced by a dc current element.

Figure 20.3 shows the radiation pattern (*antenna pattern*) obtained from the radiation term in Equation (20.2) (the first term in the bracket). This is a surface plot where the distance from the origin to any point on the surface is proportional to the power radiated in that direction, i.e., proportional to $|\boldsymbol{H}|^2$ at a large constant value of r. We do not have to work out the E-field separately, since we already know that, in the far field, $\boldsymbol{E}$ is perpendicular to both $\boldsymbol{H}$ and $\boldsymbol{k}$ and that its magnitude (in volts/m) is 377 times the magnitude of $\boldsymbol{H}$ (in amps/m).[1]

In this chapter we are interested in radiation far from the antenna, so we can ignore the near-field (second) term in Equation (20.2). At any observation point, the H-field is the integral of the contributions from every current element in any given antenna. For a wire antenna, we would evaluate a line integral. For a reflector antenna, we would have a surface integral, summing up the contributions from the current flowing on each element of the surface.

[1] If we do need to know E in the $1/r^2$(near field) region, we can calculate the curl of H from Equation (20.3) and plug that into Equation (20.2) to get E.

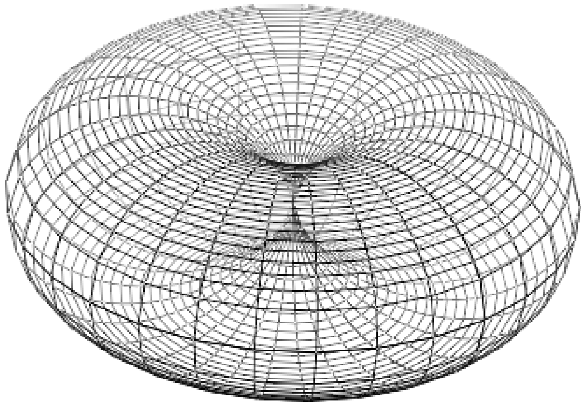


Figure 20.3. Elementary dipole radiation pattern (surface of constant field strength).

20.3 Dipole antenna

A short dipole antenna (shorter than, say, $\lambda/20$) will have the same radiation pattern as the elemental dipole since we can consider it to be a string of elemental dipoles, short enough that the phase paths from every one to the observation point are essentially the same. Thus, the string is equivalent to a single elemental dipole of strength $I\delta\, l = \sum I_i \delta l_i$. We can calculate the power radiated by a single elemental dipole by integrating the average energy flux density, ½EH, over the surface of a bounding sphere to include the power radiated in every direction. Using the radiation term in Equation (20.2) for H, we have

$$\delta H = \frac{I_0 \delta l}{4\pi}\left(\frac{jk\sin(\theta)\mathrm{e}^{\mathrm{j}(\omega t - kr)}}{r}\right), \tag{20.3}$$

$$\begin{aligned}\text{Total Pwr} &= \int (1/2)EH\,\mathrm{d}S = \int (1/2)(\mu_0/\epsilon_0)^{1/2} H^2 r^2 \mathrm{d}\boldsymbol{\Omega} \\ &= (1/2)(\mu_0/\epsilon_0)^{1/2}(I_0 k\delta l/(4\pi))^2 \int \frac{\sin^2(\theta)}{r^2} r^2 \mathrm{d}\boldsymbol{\Omega} \\ &= (1/2)(\mu_0/\epsilon_0)^{1/2}(I_0 k\delta l/(4\pi))^2 \int_0^{\pi} 2\pi \sin^3(\theta)\mathrm{d}\theta \\ &= (1/2)120\pi(I_0(2\pi/\lambda)\delta l/(4\pi))^2 8\pi/3 \\ &= 40\pi^2 (I_0\delta l/\lambda)^2.\end{aligned} \tag{20.4}$$

It turns out that the current distribution for a short dipole is a triangle function (actually the almost linear ramp portion of a sine function near zero), as shown in Figure 20.4.

https://doi.org/10.1017/CBO9780511626951.021 Published online by Cambridge University Press

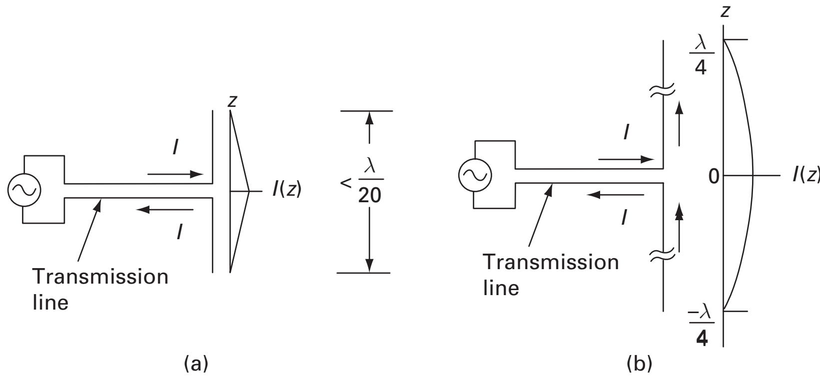


Figure 20.4. (a) A short dipole has a triangular current distribution. (b) A half-wave dipole has a sinusoidal current distribution.

The average current for a short dipole is therefore half the input current, so the power radiated is

$$P_{\text{short dipole}} = 40\pi^2 \left(\frac{I_{\text{in}}}{2} l/\lambda\right)^2, \tag{20.5}$$

where I_{in} is the peak input current and l is the length of the dipole. The input power must be equal to ½ $I_{\text{in}}{}^2 R$, where R is the real part of the antenna's input impedance, known as the *radiation resistance*. For the short dipole, we can therefore write

$$R_{\text{short dipole}} = 2P_{\text{short dipole}}/I_{\text{in}}{}^2 = 20\pi^2 (l/\lambda)^2. \tag{20.6}$$

For a short dipole with $l = 0.05\lambda$, $R = 0.49$ ohms. The imaginary part of the input impedance is negative and large, corresponding to a series capacitance with a low value. An antenna tuning network that increases the low value of R and tunes out the high value of capacitive reactance will need at least one high-value inductor. However, the limited Q of practical inductors will introduce losses that result in very low efficiency. A counterpart to the elemental dipole is the elemental current loop or *elemental magnetic dipole* (as opposed to *elemental electric dipole*). The loop is assumed to have constant current and its radiation pattern is the same as the pattern of the elemental dipole. A very small loop antenna has an extremely low radiation resistance and a high positive (inductive) reactance. It is an inefficient antenna in that its ohmic resistance is usually higher than its radiation resistance.

In practice, most dipoles have an overall length of $\lambda/2$. At this resonant length, the impedance of a thin wire dipole seen at the center gap is $73.1 + j0$ ohms, an entirely practical value. Resonant loops, with a circumference of the order of a wavelength, are also practical antennas. While transmission line arguments can

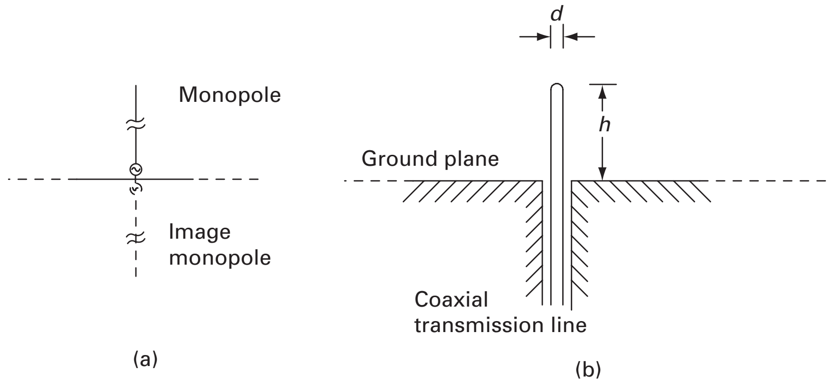


Figure 20.5. Fat monopole geometry.

be used to show that the current distribution on a wire dipole is nearly sinusoidal, going to zero at the open ends, rigorous calculations of current distributions on arbitrary antennas require elaborate self-consistent treatments of the currents and fields. A benchmark problem is the fat dipole antenna shown in Figure 20.5(b). This "monopole" is really a dipole, since the ground plane induces an image monopole as shown in Figure 20.5(a).

This geometry is appealingly simple; the radiating element, as it enters the hole in the ground plane, becomes the center conductor of a coaxial line. You can find theoretical and experimental plots of input impedance of this antenna in most standard antenna textbooks. You can also model this dipole if you have access to a program for three-dimensional electromagnetic field simulation, such as Ansoft's HFSS.

20.4 Antenna directivity and gain

The radiation from any antenna is always stronger in some directions than in others; no antenna can be an isotropic radiator.[2] Let the energy flux (W/m^2) produced in a given direction be denoted as $S(\theta,\phi,r)$, where we are using standard polar coordinates. As before, the integral of S over an enclosing sphere will be the total radiated power transmitted, P_{RAD}:

$$\int S(\theta,\varphi,r)r^2\mathrm{d}\boldsymbol{\Omega} = P_{\mathrm{RAD}}. \tag{20.7}$$

[2] It is topologically impossible to have a constant tangential vector field on the surface of a sphere, a theorem sometimes expressed as "You cannot comb the hair smooth on a billiard ball." Thus, in the far field of an antenna (at a distance greater than the square of the largest physical dimension of the antenna divided by the wavelength), where the E and B fields are transverse (tangent to the surface of a sphere surrounding the antenna), the fields cannot be everywhere constant. The radiation pattern of an antenna, therefore, cannot be isotropic.

For any given direction, the ratio of the flux to the average flux is defined as the *directivity* of the antenna

$$D(\theta, \Phi, r) = \frac{S(\theta, \Phi, r)}{P_{\mathrm{RAD}}/(4\pi r^2)}. \tag{20.8}$$

An equivalent statement is that the directivity is the factor by which the flux in the strongest direction exceeds the flux from a hypothetical isotropic antenna radiating the same total power. We know that the radiation flux eventually falls off as r^{-2}. For large r, then, the far-field directivity, D, is a function only of the polar coordinates θ and ϕ. Combining Equations (20.7) and (20.8), we see that the average directivity of an antenna is unity:

$$\frac{1}{4\pi}\int D(\theta, \Phi)\mathrm{d}\boldsymbol{\Omega} = 1. \tag{20.9}$$

Directive gain, $G(\theta,\phi,\mathrm{r})$, has the same definition as the directivity except that the radiated power is replaced by P_{inc}, the power incident on the antenna terminals:

$$G(\theta, \Phi, r) = \frac{S(\theta, \Phi, r)}{P_{\mathrm{inc}}/(4\pi r^2)}. \tag{20.10}$$

If an antenna has no ohmic losses and its feedpoint impedance matches the transmission line impedance (so that no power is reflected), all the incident power will be radiated and the directive gain will be equal to the directivity. In most antennas used for transmitting, the losses are no more than a few percent and the distinction between directivity and gain is unimportant.[3] The maximum value of an antenna's directive gain is simply called the gain. A transmitter connected to an antenna having a gain of 20 dB will produce a directed signal 100 times more powerful than if it were connected to a hypothetical lossless isotropic radiator.

Since we have already calculated the total power radiated by an elemental dipole (Equation 20.4), we can easily calculate its directivity. The maximum flux density, calculated from Equation (20.3), is $S_{\max} = (60/\pi)[\, I_0 k\,\delta\, l\,/(4\pi r)]^2$, where we have substituted 120π for $(\mu_0/\varepsilon_0)^{1/2}$. Dividing $\mathrm{S}_{\max}$ by the average flux density, S_{avg} = total power/$(4\pi r^2)$, yields a directivity of 1.5. As explained above, this will also be the directivity of a short dipole. The directivity of a half-wave dipole is 1.64 and its radiation pattern differs little from that of the elemental dipole.

[3] Small inefficient antennas are adequate for low-frequency receivers. Even though antenna losses add noise at the receiver input, signal strengths in the AM broadcast band and short-wave bands must already be considerably higher than this added noise in order to exceed the noise power from static (atmospheric electricity). Even with an inefficient antenna, the total power delivered to the receiver is much greater than the thermal noise added by the antenna and by the receiver itself.

https://doi.org/10.1017/CBO9780511626951.021 Published online by Cambridge University Press

20.5 Effective capture area of an antenna

The distance between transmitting and receiving antennas is generally so large that a plane wave can be assumed incident on the receiving antenna. The energy extracted from the incident wave is, of course, proportional to the average energy flux density, $\frac{1}{2}E \times H$ W/m^2. The proportionality constant is called the *effective area* (capture area) of the receiving antenna. It turns out, as shown below, that the effective area is proportional to gain:

$$A_{\text{eff.}} = G\lambda^2/4\pi. \tag{20.11}$$

Since gain and effective area are proportional, there is really no distinction between transmitting antennas and receiving antennas; the best transmitting antenna (most gain) is also the best receiving antenna (most capture area). A standard derivation of Equation (20.11) begins by applying the reciprocity theorem[4] to a system of two arbitrary antennas. The two antennas, as shown in Figure 20.6, need not be identical. We can suppose they are both matched to the same impedance value and that we have both a generator and a receiver that match this impedance. First we connect the generator to Antenna 1 and measure the power from Antenna 2.

If we now interchange the generator and load, the reciprocity theorem states that the power delivered to the load will be unchanged. Expressing this in terms of gain and effective area, we have $G_1 A_{\text{eff }2} = G_2 A_{\text{eff }1}$ or $G_1/A_{\text{eff }1} = G_2/A_{\text{eff }2}$, from which we see that the ratio of gain to effective area has the same constant value for any and all antennas. We can pick any conveniently simple antenna

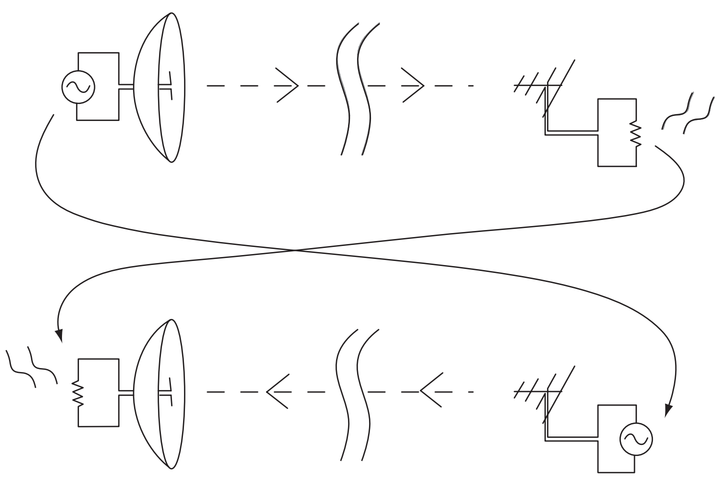


Figure 20.6. Reciprocity: received power is unchanged when source and load are interchanged.

[4] The remarkable reciprocity theorem finds application in mechanics, optics, and acoustics as well as electrical engineering.

and use electromagnetic theory to calculate its gain and effective area. When this is done, it is found that the ratio of gain to effective area is $4\pi/\lambda^2$. A half-wave dipole antenna has a maximum gain of 1.64. Its effective area is therefore $1.64\ \lambda^2/4\pi$. If the dipole is made of thin wire, it has no real physical area, only a length, yet it has a nonzero effective area and can extract energy from an incident electromagnetic wave. Note that even a very short dipole has nearly the same effective area as a half-wave dipole. The effective area of a microwave dish antenna can approach its physical area, as we will see below, but normally it is from 50 to 80 percent of the physical area. This fraction, known as *aperture efficiency*, is usually set not by ohmic losses but rather by the illumination pattern of the primary feed antenna. A perfectly uniform illumination pattern (and no ohmic loss) produces an aperture efficiency of unity. The Arecibo radio telescope dish uses an aperture 700 ft in diameter (area = 35 800 m^2). Its aperture efficiency is about 70% at $\lambda = 12$ cm (the wavelength used there for radar astronomy) so its gain, using Equation (20.11), is $G = 4\pi(0.70 \times 35\,800)/(0.12)^2 = 22 \times 10^6$ or 73 dB.

20.6 Reflector and horn antennas

Radar antennas and satellite TV dishes are familiar examples of reflector antennas. From a transmitting standpoint, a primary or “feed” antenna illuminates the reflector. Currents induced in the metallic surface of the reflector become secondary radiators and their radiation forms the beam of the antenna. The larger the dish, the more directive the antenna, as we will see. These antennas are examples of “aperture antennas;” one can readily identify the aperture (usually circular) from which the radiation emanates, as if from a searchlight.

Let us use Equation (20.3) to calculate the radiation from a large flat rectangular metal plate on which there is a surface current density, J_S, which has the same amplitude and phase at every point.[5] For surface current, $J_S\mathrm{d}x\mathrm{d}y$ replaces $I_0\mathrm{d}l$ in Equation (20.3). To find the antenna pattern we integrate over the plate, summing up the contributions from each element of area. These are phasor contributions and the far field is actually an interference pattern. Once we have calculated the antenna pattern, we will integrate the power over a bounding sphere, as we did for the elemental dipole, to determine the total power radiated. With this, we can then calculate the gain.

Figure 20.7 shows the geometry for this antenna. The size of the plate is $2a$ by $2b$. Radiation will be strongest in the z-direction, perpendicular to the plate. We

[5] Such a current distribution could be established by illuminating the plate with radiation from a dipole far out in front of the plate, though this would not be an efficient feed antenna, since the plate would intercept only a small fraction of the dipole’s radiation.

https://doi.org/10.1017/CBO9780511626951.021 Published online by Cambridge University Press

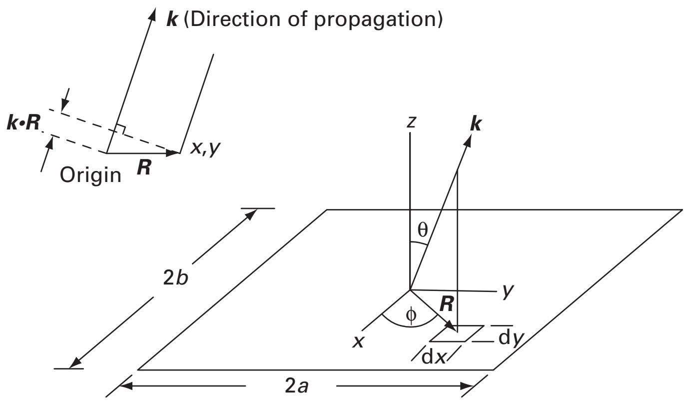


Figure 20.7. Rectangular aperture antenna geometry.

will assume that both a and b are much greater than λ, anticipating that this will form a very concentrated beam.[6]

The vector from the center of the plate to a distant observation point is denoted by $\boldsymbol{r}$, and the vector from the center of the aperture to an arbitrary point x,y on the aperture is denoted by $\boldsymbol{R}$. Again using the first term (the radiation term) of Equation (20.3), we see that the contribution to the distant H field from an element of area at the center of the plate is given by $\delta H = J_S\,\mathrm{d}x\mathrm{d}y\;\mathrm{j}k(4\pi r)^{-1}\cos(\theta)\mathrm{e}^{\mathrm{j}(\omega t-kr)}$ where, as always, $k=2\pi/\lambda$ and $\boldsymbol{k}$ is directed from the plate to the point $\boldsymbol{r}$. Note that the $\sin(\theta)$ factor in Equation (20.3) is here $\cos(\theta)$ because the current is perpendicular, not parallel, to the z-axis. We will replace this $\cos(\theta)$ by unity, anticipating that, when both a and b are much greater than λ, the pattern will be highly concentrated around the $\theta=0$ direction. Away from the origin, radiation to the point $\boldsymbol{r}$ from any other point on the plate will be shifted in phase by $\boldsymbol{k}\cdot\boldsymbol{R}$ radians, as shown by the inset in Figure 20.7. Using standard polar coordinates, this phase shift is given by

$$\boldsymbol{k}\cdot\boldsymbol{R} = xk_2 + yk_y = k\,\sin(\theta)[x\cos(\phi) + y\sin(\phi)]. \tag{20.12}$$

If we assume a uniform field over the aperture, the field at r, θ, ϕ can now be written as

$$\begin{aligned} H(r,\theta,\phi) &= \frac{jkJ_S\mathrm{e}^{\mathrm{j}\omega(t-kr)}}{4\pi r}\int_{-b}^{b}\int_{-a}^{a}\mathrm{e}^{\,\mathrm{j}k\cdot R}\mathrm{d}x\mathrm{d}y \\ &= \frac{jkJ_S\mathrm{e}^{\,j\omega(t-kr)}}{4\pi r}\int_{-b}^{b}\int_{-a}^{a}\mathrm{e}^{\,jk\theta(x\cos\phi+y\sin\phi)}\mathrm{d}x\mathrm{d}y, \end{aligned} \tag{20.13}$$

[6] When used at a wavelength of 12 cm, the 210-m diameter illumination on the dish at Arecibo forms a beam whose width between half-power points is only 12/21 000 radians or 0.03 degree, a "pencil beam."

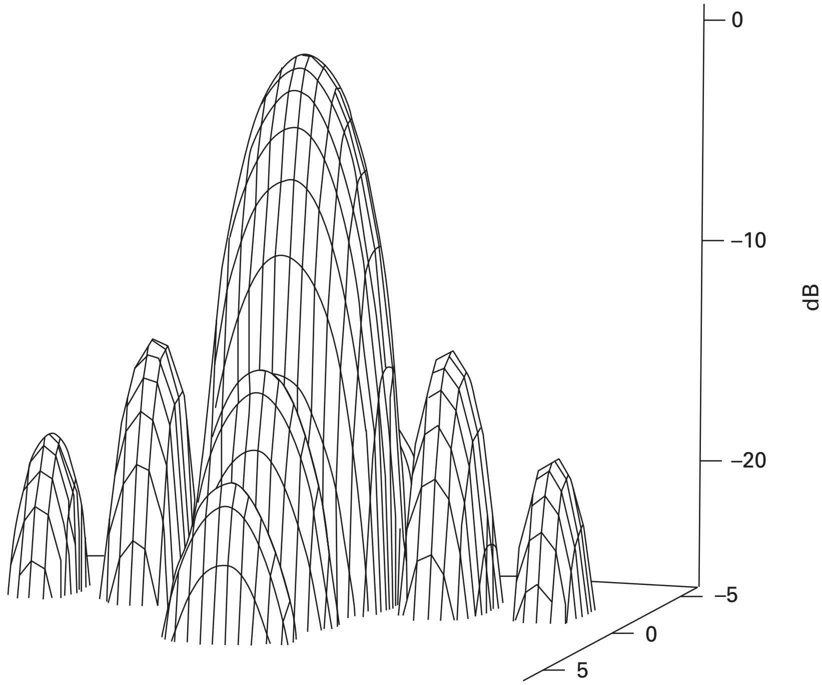


Figure 20.8. Radiation pattern (power in dB) from a rectangular aperture.

where we have replaced $\sin(\theta)$ by θ, again anticipating that the power will be negligible except for the region close to $\theta=0$. Evaluating the integral we find

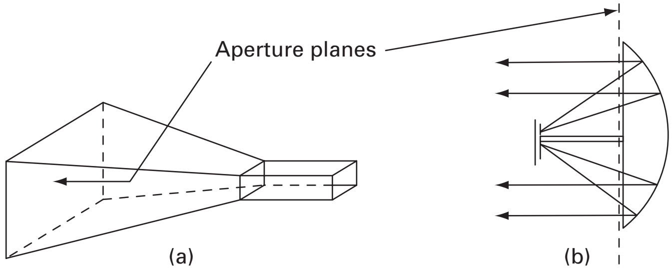
$$H(r,\theta,\phi)=\frac{\mathrm{j}kJ_{\mathrm{S}}ab\mathrm{e}^{\mathrm{j}\omega(t-kr)}}{4\pi r}\left(\frac{\sin(ka\theta\cos(\phi))}{ka\theta\cos(\phi)}\frac{\sin(kb\theta\sin(\phi))}{kb\theta\sin(\phi)}\right). \qquad (20.14)$$

Since the average power density (watts/m^2) is given by $EH = ½(\mu_0/\varepsilon_0)^{1/2}H^2$, the antenna pattern is proportional to the square of the term in brackets. Figure 20.8 shows the pattern, which has a strong central *main lobe* surrounded by *sidelobes*, which for this $(\sin(x)/x)^2$ power pattern are 13.3 dB below the peak of the main beam.

This plot uses rectangular "sky coordinates," u and v, measured in radians, where $u=\theta\sin(\phi)$ and $v=\theta\sin(\phi)$. The term in brackets is, therefore, $\sin(kau)/(kau)\ \sin(kbv)/(kbv)$. This plot was made for a square aperture; a rectangular aperture produces an elongated beam. For $a<b$, the beam is broader in the u-direction than in the v-direction. We can integrate the power density of the pattern to get the total average power:

$$\begin{aligned}\mathrm{Pwr}&=1/2\sqrt{\frac{\mu_0}{\varepsilon_0}}\left|\frac{jJ_{\mathrm{S}}kab\mathrm{e}^{\mathrm{j}\omega(t-kr)}}{4\pi r}\right|^2\iint r^2\left(\frac{\sin(kau)}{kau}\frac{\sin(kbv)}{kbv}\right)^2\mathrm{d}u\mathrm{d}v\\&=1/2\sqrt{\frac{\mu_0}{\varepsilon_0}}\left(\frac{J_{\mathrm{s}}kab}{4\pi}\right)^2\frac{\pi^2}{k^2ab}.\end{aligned} \qquad (20.15)$$

Figure 20.9. Horn and dish antenna aperture planes.



This let us calculate the gain:

$$G = 1/2\sqrt{\frac{\mu_0}{\epsilon_0}}|H(r,0,0)|^2 \div \frac{\text{Pwr}}{4\pi r^2} = \frac{4\pi(4ab)}{\lambda^2} = \frac{4\pi A}{\lambda^2}. \quad (20.16)$$

This example antenna serves to establish the relationship between gain and effective area, $G = 4\pi\ A_{\text{eff}}\ /\lambda^2$, as long as we can argue that the effective area (capture area) is equal to the actual physical area. To argue this case, consider that a wave transmitted by this antenna is essentially a plane wave as it leaves the large aperture (as contrasted with the spherical wave emitted by a small antenna). Therefore, if we invoke time reversal, the wave entering the antenna would be a plane wave, just as if it had come from a distant transmitting antenna. Since the time-reversed wave is completely "absorbed" by the antenna, the effective area is equal to the physical area, $4ab$.

The integral in Equation (20.13) is just a two-dimensional Fourier transform of the aperture illumination. In this example, we used uniform illumination, so we have transformed the shape of the aperture with uniform weighting. The aperture can have any shape. For example, a circular plate of radius $a \gg \lambda$ produces a field pattern proportional to $J_1(ka\theta)/(ka\theta)$, where J_1 is the first-order Bessel function of the first kind. (The function $J_1(x)$ resembles $\sin(x)[x+1]^{-1/2}$). For any given aperture, uniform illumination produces the highest gain. But "tapered" illumination, where the current density is less near the edges of the aperture, is often used because it reduces the amplitude of the sidelobes.

A simple *microwave horn antenna* is essentially a waveguide funnel, fed at the small end, as shown in Figure 20.9. The other end fans out, usually in both transverse directions, to form a large aperture. The beam radiated from the aperture is comparable to the beam radiated from a dish antenna having equal area. The field distribution in the aperture plane is the same as the field at any cross-section in the waveguide, expanded to fill the aperture.

The horn antenna has no secondary radiators in its aperture plane. Of course there are currents on the inside walls of the waveguide and, in principle, Equation (20.3) can be used to calculate the radiation pattern far out in front of the horn. A much simpler way to find the radiation pattern is to use Huygen's principle, where the aperture *field* is regarded as a *source*, a two-dimensional array of wavelets, each of which re-radiates as if it were a current element.[7] If you know the *E*-field (or *H*-field) in the aperture plane, the squared magnitude of its two-dimensional Fourier transform yields the radiation pattern, just as the squared magnitude of the transform of the currents on a metal plate yielded the radiation pattern in the above example. This analysis technique is also applied to dish antennas. Ray-tracing methods (geometric optics) are used to find the phase and amplitude of the field at an aperture plane – usually just in front of the dish. This field is then regarded as a Huygen's source; its Fourier transform gives the far-field pattern without requiring an intermediate calculation of the currents on the reflector surface. This Fourier transform method is also applied to systems having multiple antennas, such as arrays of dish antennas, used together as an *interferometer* to form an extremely narrow beam.

20.7 Polarization

The elementary dipole of Figure 20.2 produces radiation with *linear polarization*; at any observation point the electric field vector has a fixed direction and an amplitude that oscillates sinusoidally. A dipole receiving antenna would have to be placed parallel to the incoming polarization. If placed perpendicular to the incoming *E* vector, it would pick up no signal. In all the preceding discussions, we have implicitly assumed that the polarization of a receiving antenna was matched to that of a transmitting antenna. *Circular polarization* can be produced with an antenna made of crossed dipoles. The dipoles are fed by signals that are equal in amplitude but 90° out of phase. This results in a transmitted *E* vector that rotates in the plane perpendicular to the direction of propagation. At any observation point, the magnitude of the *E* vector remains constant, but the vector rotates one turn for every period of the wave. A 90° hybrid can be used as a power splitter to feed the two dipoles. An identical antenna (a pair of crossed dipoles with a 90° hybrid as a combiner) makes a polarization-matched receiving antenna. Note that a single dipole can serve as a receiving antenna for circular polarization, but it collects only half the available power. If the incoming *E* field, viewed from the receiving position, rotates in the counterclockwise sense, the radiation has, by definition, *right-hand circular polarization (RCP)*. If the cables from the receiving antennas were to be interchanged at the hybrid, the resulting *LCP* antenna would have the wrong handedness and could receive no power from an RCP transmitting antenna. The most general polarization is elliptical, which is an admixture of linear and circular. Linear and circular polarizations are just special cases of elliptical polarization.

[7] See Reference [1] for a theoretical justification of Huygen's principle.

https://doi.org/10.1017/CBO9780511626951.021 Published online by Cambridge University Press

20.8 A spacecraft radio link

Consider the following example of a spacecraft telemetry link for which we wish to find the maximum range. Suppose we have a 1-W telemetry transmitter aboard a spacecraft and that the data rate requires a channel capacity corresponding to a signal-to-noise ratio of at least unity in a 1-Hz bandwidth. This link uses a frequency of 3 GHz (10 cm wavelength). The transmitting antenna on the spacecraft is a 2-m diameter dish. The ground station antenna is a 10-m diameter dish, as shown in Figure 20.10. Assume that both these dish antennas have an effective area equal to 60% of their physical apertures. Assume also that there is no pointing error, i.e., the antennas always point directly at each other and that the system temperature of the ground station receiver is 25 K. (The system temperature is the sum of the equivalent receiver noise temperature, the antenna noise temperature, and the sky noise temperature.)

1. What is the equivalent input noise power of the receiver? Boltzmann's constant, k, is 1.38×10^{-23} W/Hz/K, so the equivalent input noise power, kTB, is $1.38 \times 10^{-23} \times 25 \times 1 = 3.45 \times 10^{-22}$ W.
2. What is the effective area of the receiving antenna? The physical aperture is $\pi \mathrm{R}^2$ so the effective area is $0.60\ \pi R^2 = 0.60\pi 5^2 = 47.1\ \mathrm{m}^2$.
3. What is the gain of the transmitting antenna? The physical aperture is πR^2 and the aperture efficiency is 0.60 so the effective aperture is $0.60\ \pi(1^2) = 1.88\ \mathrm{m}^2$. The gain is $4\pi A_{\mathrm{eff}}/\lambda^2 = 4\pi(1.88)/(0.1^2) = 2369$.
4. What is the maximum range, R, in kilometers for the spacecraft to maintain the required signal-to-noise ratio? Here we simply set $P_{\mathrm{noise}} = P_{\mathrm{received}} = A_{\mathrm{rcvr}}\ P_{\mathrm{trans.}} G_{\mathrm{trans}}/\ (4\pi R^2)$ from which we have $R^2 = (4\pi)^{-1} P_{\mathrm{trans.}}\ G_{\mathrm{trans}}/P_{\mathrm{noise}}$. Using the parameters calculated above, we have $R = \sqrt{(4\pi)^{-1} \times 1 \times 2369/(3.45 \times 10^{-22})} = 7.4 \times 10^{11}\,\mathrm{m} = 740 \times 10^6\,\mathrm{km}$. This is roughly the mean distance to Jupiter.

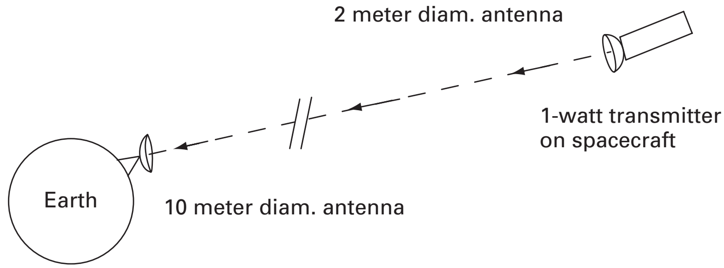


Figure 20.10. A spacecraft telemetry link.

20.9 Terrestrial radio links

VHF and UHF two-way radios used by cellular phones, emergency vehicles, etc. have transmitters with several watts of power but their range is limited by the curvature of the Earth to only a few miles or tens of miles. (Radio waves do not propagate through the highly conductive Earth though they do diffract slightly, so the radio horizon is somewhat beyond the optical horizon.) Mobile radios also use antennas that have gains only of the order of unity so they do not have to be pointed accurately – or even at all. Finally, in ground-to-ground radio links, the signal usually arrives from an angle near the horizon; the receiving antenna will pick up noise from the ground (thermal radiation). In these systems, then, extremely low-noise receivers are of no benefit. Broadcasting stations for FM and television also use VHF and UHF frequencies so their range is also essentially line-of-sight. Long-distance propagation in AM and short-wave broadcasting depends on reflection from the ionosphere.

20.10 The ionosphere

At altitudes above about 60 km the atmosphere is ionized by ultraviolet radiation from the Sun; electrons are stripped from the neutral particles (mostly oxygen atoms and O_2 and N_2 molecules) to produce a mixed electron and ion gas. During the day the density of this ionized gas is highest at around 250 km, the peak of the "F-region." Above the peak the ionization is less because the thinner atmosphere presents fewer particles to be ionized. Below the peak the ionization is less because the denser atmosphere exhausts the supply of ultraviolet photons; the electrons they produce quickly encounter nearby ions and recombine. At night, without sunlight, the ionization rate is zero. Recombination quickly neutralizes the ionization at the lowest altitudes, around 100 km, and depletes the F-region until sunrise.

20.10.1 Wave propagation in the ionosphere

An electromagnetic wave induces electric currents in the electron gas of the ionosphere. (The electrons, by virtue of their low mass, are accelerated by the incident wave to much higher velocities than the ions, so the ion contribution to the current is negligible.) The effective dielectric constant of an electron gas is not difficult to calculate from Maxwell's equations. (We will see that this dielectric constant becomes imaginary below a certain critical frequency which depends on the electron density; below this frequency, then, electromagnetic waves cannot propagate through the plasma.) No longer in a vacuum, we must use the general curl H equation which includes the real electric current, J, in addition to the displacement current:

https://doi.org/10.1017/CBO9780511626951.021 Published online by Cambridge University Press

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}. \tag{20.17}$$

The electrons are in rapid thermal motion, but this motion is random and contributes nothing to the current. Yet all the electrons near any given point are accelerated equally and move together to produce a net current, $\boldsymbol{J} = Nev$, where N is the electron density, e is the electron charge, and v is the component of velocity imparted by the electric field. We will neglect the weak $\boldsymbol{v} \times \boldsymbol{B}$ force from the magnetic field so Newton's second law of motion is just $m\mathrm{d}\boldsymbol{v}/\mathrm{d}t = \boldsymbol{F} = e\mathbf{E}$. For a sinusoidal time dependence, $\mathrm{e}^{\mathrm{j}\omega t}$, we can write this equation of motion as $m(\mathrm{j}\omega \boldsymbol{v}) = e\mathbf{E}$. Substituting $\boldsymbol{J} = Ne\mathbf{v} = Ne^2\boldsymbol{E}/(\mathrm{j}\omega m)$ in Equation (20.17) gives

$$\nabla \times \boldsymbol{H} = \frac{Ne^2\boldsymbol{E}}{\mathrm{j}\omega m} + j\omega\epsilon_0\boldsymbol{E} = \mathrm{j}\omega\epsilon_0\left(1 - \frac{Ne^2}{\epsilon_0 m\omega^2}\right)\boldsymbol{E}. \tag{20.18}$$

Note that the term in brackets is the relative dielectric constant and that it becomes negative for low frequencies, in particular for $\omega^2 < \omega_\mathrm{p}^2$ where $\omega_\mathrm{p}^2 = Ne^2/(\varepsilon_0 m)$. This happens because the conduction current (the electron current) becomes greater than the displacement current. The total current (conduction current plus displacement current) changes sign and has the wrong polarity to source the H field of a traveling wave. This critical frequency, ω_p, is known as the *plasma frequency*. (If the local charge neutrality of an electron–ion gas is disturbed, the densities will oscillate at this frequency the way a spring and mass system oscillates at its resonant frequency.) For a wave to propagate in the plasma, the dielectric constant must be positive; only waves with frequencies lower than the plasma frequency will be reflected. The free-electron gas that gives metals their conductivity is dense enough to reflect visible light but the alkali metals (lithium, sodium, etc.) have relatively lower electron densities and are transparent in the ultraviolet.

20.10.2 Reflection of waves from the ionosphere

The reflection of radio waves is normally a process of refraction because the waves are not vertically incident on the ionosphere. As they travel obliquely upward, the dielectric constant decreases so the phase velocity increases, causing the propagation vector, which is perpendicular to the wavefront, to turn around gradually. This is shown in Figure 20.11.

For signal strength calculations over an ionospheric path the ionosphere can be considered a specular mirror and field strengths can be calculated by taking the inverse square of the total path length of the ray.

20.10.3 Daytime vs. nighttime propagation

Short-wave broadcasts from distant transmitters on the higher frequency bands are not heard at night because the ionospheric electron density is too low for

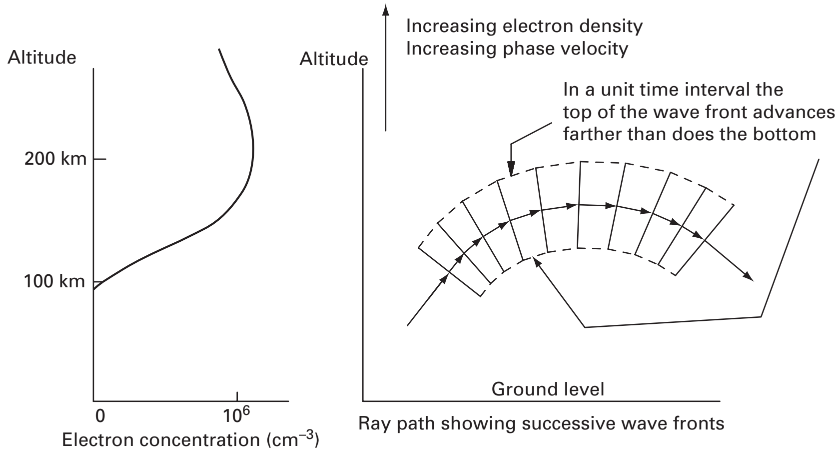


Figure 20.11. Ionospheric refraction.

reflection, i.e., the waves cannot get turned around sufficiently. On the other hand, low-frequency stations, such as those in the AM broadcasting band, are received from great distances *only* at night. During the day their energy is dissipated in the lower ionosphere through collisions between the accelerated electrons and neutral atoms and molecules. Why don't higher frequency waves suffer this daytime attenuation in the lower ionosphere? Consider a low-frequency and a high-frequency wave of equal power, that is, equal field strengths. Both cause the ionospheric electrons to execute synchronized sinusoidal motion as described above. But the low-frequency wave, because its period is long, will produce higher electron velocities; we saw above that $v = Ee/(j\omega m)$. The average kinetic energy of each electron is therefore $m\langle v^2\rangle/2 = m[Ee/(\omega m)]^2/4$ where E is the amplitude of the electric field. All the electrons in the lower ionosphere suffer collisions with the neutral particles (which are present since the ionization is not 100%). The average collision leaves the electron with a random velocity, i.e., its share of the synchronized sinusoidal motion is lost. The frequency of collisions does not depend on the frequency of the electromagnetic wave (or on there being any electromagnetic wave present at all) so the rate of energy loss is inversely proportional to the square of ω, the wave frequency, and long-distance AM listeners have to wait for nighttime.

20.11 Other modes of propagation

Besides reflection by the ionosphere, there are a number of other ways that an electromagnetic wave can get around the curvature of the Earth. These include scattering from the ionized trails of meteors entering the Earth's atmosphere, scattering from irregularities in the ionosphere even when the ionosphere is otherwise not dense enough to turn the waves around by refraction, scattering from density irregularities in the neutral atmosphere (i.e., fluctuations in the index of propagation), and ducting beneath atmospheric temperature inversion layers.

Problems

Problem 20.1. The voltage at the terminals of a receiving antenna is proportional to the E field of an incident electromagnetic wave. The "effective length" (or effective height if the antenna is vertical) is defined as the open-circuit voltage at the terminals divided by the incident E field: volts / (volts/meter) = meters. Show that the effective length is given by

$$\text{effective length} = (4RA_{\text{eff}}/Z_0)^{1/2},$$

where R, the radiation resistance, is the real part of the antenna impedance, A_{eff} is the effective area ($A_{\text{eff}} = G{\cdot}\lambda^2/(4\pi)$), and $Z_0 = 377$ ohms $= (\mu_0/\epsilon_0)^{1/2}$, the impedance of free space. Find the effective length of a half-wave dipole ($G = 1.6$ and $R = 73$ ohms).

Problem 20.2. Suppose we have a 1-W transmitter connected to a dipole antenna which is aligned to provide the maximum signal strength at a distant receiving antenna. Needing more signal strength, we obtain a second, identical dipole and, using a power splitter, feed each dipole with ½W. We space the second far enough from each other so that they do not interact. We make sure that both antennas are aligned toward the receiver and we also make sure that the cables from the power splitter have equal length. At the receiving antenna, each transmitting antenna provides a field amplitude that is less than the original field by $1/\sqrt{2}$. But the two signals are in phase so the total amplitude is increased by $2/\sqrt{2} = \sqrt{2}$. Squaring this we see that the received signal strength is doubled. Have we gotten something for nothing? Could we repeat this process to increase the received power even more?

Problem 20.3. Let the individual antennas of Problem 20.2 be AM broadcast towers with omnidirectional patterns and vertical polarization. Suppose the spacing between these antennas is $\lambda/2$. As before, they are fed symmetrically, that is, with the same power and same the phase. Find the radiation pattern in the horizontal plane: make a polar plot of the relative field strength vs. azimuth angle for a distance far from the antennas. Hint: at any observation point in the horizontal plane at a distance r from the center of the line joining the two antennas, the total voltage is the sum of the contributions from the two antennas, $e^{j\phi_1}$ and $e^{j\phi_2}$. The phases π_1 and π_2 are the phase path lengths corresponding to r_1 and r_2, the distances from the observation point to the respective antennas. These phase paths are just $2\pi r_1/\lambda$ and $2\pi r_2/\lambda$. The field strength is given by $(|e^{j\phi_1} + e^{j\phi_2}|)$.

Problem 20.4. Consider a pair of crossed dipoles. The first dipole points in the z-direction and carries a current $\cos(\omega t)$. The second dipole points in the x-direction and carries a current $\sin(\omega t)$. Find the type of polarization and the relative power density of the radiation in the $+z$, $-z$, $+x$, $-x$, $+y$, and $-y$ directions. If you have a program like MATLAB or Mathcad, make a three-dimensional surface plot, like Figure 20.3, for which the distance from any point on the surface to the origin is proportional to the radiation power in that direction.

References

[1] Collin, R. E., *Antennas and Radiowave Propagation*, New York: McGraw-Hill, 1985.
[2] Davies, K., *Ionospheric Radio Propagation*, New York: Dover Publications, 1966.
[3] Kelley, M. C., *The Earth's Ionosphere: Plasma Physics and Electrodynamics*, New York: Academic Press, 1989.