15

Antennas and Propagation

So far, we have made measurements without saying anything about antennas, or about how power gets from the transmitter to the receiver. In our measurements, a 50-Ω load has taken the place of the antenna. However, instead of dissipating the power as heat, an antenna radiates power as electromagnetic waves. One thing that makes antennas interesting is that they necessarily involve both the voltages and currents that we study in circuits and the electric and magnetic fields that make up radio waves. This gives antennas a special place in the history of physics. They were the crucial components that Hertz developed in the 1880s to demonstrate that Maxwell's equations for electricity and magnetism are correct. In the 1960s, a special parabolic antenna allowed Arno Penzias and Robert Wilson at Bell Telephone Laboratories to discover the cosmic background radiation. That measurement earned them a Nobel Prize, and it gave an entirely new interpretation to the history of the universe.

An antenna is characterized by its impedance and pattern, which is a plot of where the power goes for a transmitting antenna. Traditionally, antennas have been analyzed as transmitters, and most antenna engineers think entirely in terms of transmitting antennas. If we know how an antenna transmits, we can use the reciprocity theorem to figure out how the antenna works in reception. The physical description of transmission and reception are actually quite different, and the physics of receiving antennas is in many ways as interesting as for transmitting antennas. In particular, a dipole can be analyzed in a simple way as a receiving antenna, and we will follow this approach. We need a few results from electromagnetic theory to start.

15.1 Radio Waves

Radio waves are predicted by Maxwell's equations. These are complicated, but we will consider the special case of plane waves in free space, where Maxwell's equations are dual to the transmission-line formulas. We write them in curl form with phasors, so that $\partial/\partial t$ becomes $j\omega$:

$$\nabla \times \mathbf{E} = -j\omega\mu_0\mathbf{H}, \tag{15.1}$$

$$\nabla \times \mathbf{H} = j\omega\epsilon_0\mathbf{E}. \tag{15.2}$$

The first equation is Faraday's law, and the second is Ampère's law. In the equations, **E** is the electric field, with units of V/m, and **H** is the magnetic field, with units of A/m. We use bold face to indicate a vector. μ_0 is the *permeability* and ϵ_0 is the *permittivity* (ϵ is the Greek letter *epsilon*). We add the zero subscript to indicate that these are for free space. μ_0 and ϵ_0 are given by

$$\mu_0 = 4\pi \times 10^{-7}\,\text{H/m} \approx 1.26\,\mu\text{H/m} \tag{15.3}$$

and

$$\epsilon_0 \approx 1/(36\pi)\,\text{nF/m} \approx 8.85\,\text{pF/m}. \tag{15.4}$$

In addition to numerical forms, I have given forms with factors of π because they sometimes simplify expressions. The π form for μ_0 is exact.

In a *plane wave*, we assume variation in only one direction, along the z axis, and we let the field vary as $\exp(j\omega t - j\beta z)$. This allows us to rewrite Maxwell's equations in algebraic form as

$$\beta\hat{\mathbf{z}} \times \mathbf{E} = \omega\mu_0\mathbf{H}, \tag{15.5}$$

$$\beta\hat{\mathbf{z}} \times \mathbf{H} = -\omega\epsilon_0\mathbf{E}, \tag{15.6}$$

where $\hat{\mathbf{z}}$ is the unit vector in the z direction, and $\times$ denotes a cross product. A cross product of two vectors is orthogonal to both vectors; thus these equations tell us that **E** and **H** are both orthogonal to $\hat{\mathbf{z}}$. In other words, **E** and **H** have only x and y components. For us it is sufficient to consider *linear polarization*, where the fields lie along one axis. Let us orient the x axis so that it is aligned with the electric field, so that there is an E_x component only. Equation 15.5 can then be written as

$$\beta E_x\hat{\mathbf{y}} = \omega\mu_0\mathbf{H}. \tag{15.7}$$

This means that **H** must be in the y direction (Figure 15.1a). We can rewrite Equations 15.5 and 15.6 as

$$\beta E_x = \omega\mu_0 H_y, \tag{15.8}$$

$$\beta H_y = \omega\epsilon_0 E_x. \tag{15.9}$$

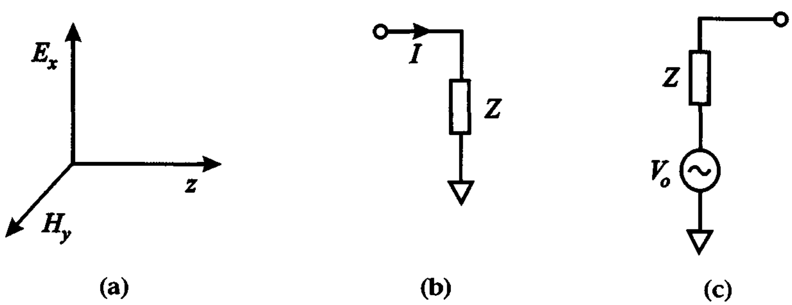


Figure 15.1. (a) Field directions for radio waves. For a linearly polarized wave propagating in the z direction, we can align the x axis with the electric field and the y axis with the magnetic field. (b) Current and impedance for a transmitting antenna. (c) Thevenin equivalent for a receiving antenna.

These are duals to the general transmission-line equations 4.41 and 4.42. The ratio of E_x and H_y is called the *wave impedance*, and it is written as η_0. This corresponds to the characteristic impedance of a transmission line. It is given by

$$\eta_0 = \frac{E_x}{H_y} = \sqrt{\frac{\mu_0}{\epsilon_0}} \approx 120\pi\,\Omega \approx 377\,\Omega. \tag{15.10}$$

We write the phase constant β as

$$\beta = \omega\sqrt{\mu_0\epsilon_0} \tag{15.11}$$

and the velocity as

$$c = 1/\sqrt{\mu_0\epsilon_0} = 299{,}792{,}458\,\text{m/s} \approx 3.00 \times 10^8\,\text{m/s}, \tag{15.12}$$

where c is the traditional symbol for the speed of light. We have exact expressions for both c and μ_0, and therefore we can work backwards to find ϵ_0 to any precision that we want. In transmission lines, the product of voltage and current gives power. The corresponding quantity for radio waves is power density, and we write the average power density in the z direction, S, in W/m^2 as

$$S = \text{Re}\left(\frac{E_x H_y^*}{2}\right) = \frac{|E_x|^2}{2\eta_0}. \tag{15.13}$$

15.2 Impedance

In a circuit, a transmitting antenna is a load, and we can characterize it by an impedance Z (Figure 15.1b). We can write the power delivered to an antenna in terms of the real part R and the current I. We write this power as P_t, with the t standing for "transmitter":

$$P_t = R|I|^2/2. \tag{15.14}$$

We distinguish between power that is absorbed by the antenna materials and power that is actually radiated. To do this we split the antenna resistance into two parts, a *radiation resistance* R_r and a *loss resistance* R_l. We can write the antenna resistance R as the sum

$$R = R_r + R_l, \tag{15.15}$$

and we write the radiation efficiency as

$$\eta = R_r/R. \tag{15.16}$$

In a circuit, a receiving antenna is a source, and we can write a Thevenin equivalent for it (Figure 15.1c). We find the Thevenin source impedance as the look-back impedance with the voltage source V_o turned off. This is just how we find the transmitting impedance. This means that *an antenna's impedance is the same for transmitting and receiving.*

15.3 Directions and Solid Angles

In calculations, we specify directions by a spherical coordinate system with angles θ and ϕ (Figure 15.2a). In this system, θ is the angle from the θ axis, and ϕ is the angle of rotation around the θ axis. This is particularly convenient for antennas with rotational symmetry, where we only need a plot as a function of θ. It is common in measurements to make plots as a function of the *elevation* and *azimuth* angles (Figure 15.2b). This is useful when the antenna has a particular orientation to the earth. The elevation angle is the angle above the horizon, and the azimuth angle is the relative bearing. These are actually quite similar to θ and ϕ, if we point the theta axis straight up. This direction is called the *zenith*. It is the angle from the zenith that actually corresponds to θ, and ϕ becomes the negative of the azimuth.

Antenna formulas are often conveniently written in terms of *solid angles*. We define a solid angle Ω for a surface S and a reference point P(Figure 15.3a). It is the area of the projection of the surface on a sphere with unit radius centered at P. The units are *steradians*. The solid angle for a closed surface that includes P is the

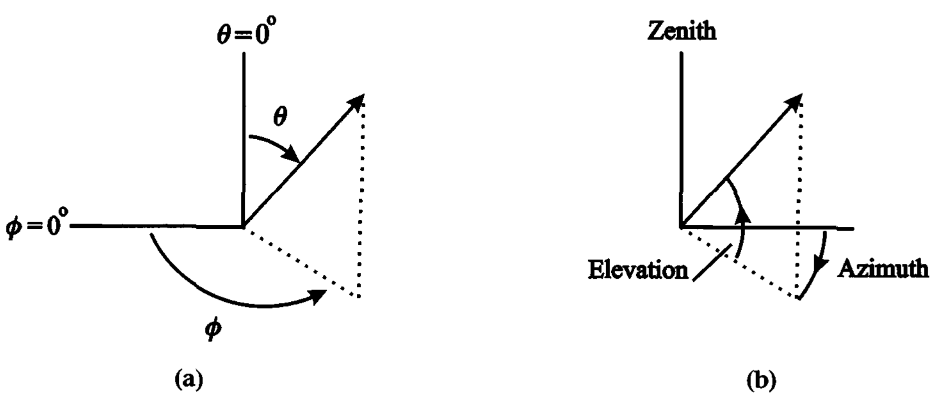


Figure 15.2. Specifying direction with a spherical coordinate system (a) and azimuth–elevation system (b).

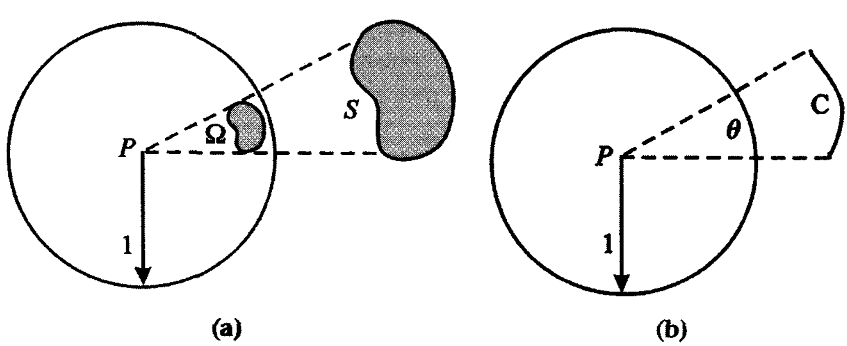


Figure 15.3. Defining a solid angle Ω as the area of the projection of a surface S onto a unit sphere (a). Defining a plane angle as the projection of a curve C onto a unit circle (b).

area of a unit sphere, which is 4π. This definition may seem strange, and it may help to make an analogous definition for ordinary plane angles. We can define an angle θ for a curve C and a reference point P. It is the length of the projection of the curve on a unit circle centered at P (Figure 15.3b). The angle for a closed curve that includes P is the circumference of the unit circle, 2π.

15.4 Transmitting Antennas

We define the *gain* G for a transmitting antenna as the ratio of the power density S to the maximum power density S_r from a reference antenna. We can write

$$G(\theta,\phi) = \frac{S(\theta,\phi)}{S_r}. \tag{15.17}$$

Power density has units of watts per square meter, but gain itself has no units. Gain is used only for transmitting, and we have different measures for receiving. For theoretical work, it is usually most convenient to define a reference *isotropic* antenna that is lossless and radiates uniformly in all directions. We can write the power density for an isotropic antenna S_i in terms of P_t as

$$S_i = \frac{P_t}{4\pi r^2}, \tag{15.18}$$

where P_t is the transmitter power and r is the distance from the antenna. We can then write the gain G as

$$G = \frac{S}{S_i} = \frac{4\pi r^2 S}{P_t}. \tag{15.19}$$

The gain is usually quoted in dB, and the letter "i" is added if it is necessary to specify that the reference antenna is isotropic. We could say that the gain of the isotropic antenna itself is 0 dBi in all directions. Gain is a function of angle, but it is common to say *the* gain for the maximum gain. In measurements, the reference antenna is often a half-wave dipole. In this case, we would write dBd to show that the reference antenna is a dipole. The gain of a dipole in free space is about 2 dBi, but reflections from the ground can raise or lower this as much as 6 dB.

In calculations it is common to assume that an antenna is lossless. Often the losses are small, and this is a reasonable approximation. For a lossless antenna, the integral of the gain over all angles has a particularly simple form:

$$\oint G(\theta,\phi)\, d\Omega = \oint \frac{4\pi r^2 S(\theta,\phi)}{P_t}\, d\Omega = \frac{4\pi}{P_t} \oint r^2 S(\theta,\phi)\, d\Omega. \tag{15.20}$$

A circle is added to the integral sign to indicate that the integral is over all angles. Because an area at a distance r scales by r^2 when we project it onto a unit sphere, we can think of $r^2 S$ as the power density per unit solid angle. We write this as

$$S_\Omega(\theta,\phi) = r^2 S(\theta,\phi), \tag{15.21}$$

where S_Ω is transmitter power density per steradian. We get

$$\int G(\theta, \phi)\, d\Omega = \frac{4\pi}{P_t} \oint S_\Omega(\theta, \phi)\, d\Omega. \tag{15.22}$$

The integral of S_Ω over all angles is just P_t. This gives us

$$\oint G\, d\Omega = 4\pi. \tag{15.23}$$

For a lossless antenna, the integral of the gain over all angles is 4π.

15.5 Receiving Antennas

We characterize a receiving antenna by an *effective length* and an *effective area*. The effective length is defined in terms of the Thevenin equivalent circuit for the receiving antenna (Figure 15.1c). We write the effective length h in terms of the Thevenin voltage V_o as

$$V_o = hE, \tag{15.24}$$

where h is the effective length and E is the incident electric field. The effective length is useful for characterizing short wire antennas, because the peak value is about half the physical length. As an example, consider the effective length of a pair of wires with total length l (Figure 15.4a). This antenna is called a *dipole*. We assume that the effective length is much less than the wavelength. We let θ be the angle from the dipole axis. The voltage at the terminals is just the potential difference between the two wires. This can be determined by symmetry if the potential of one wire is not affected by the presence of the other wire. The voltage of each wire is the potential at the midpoint, so that V_o is given by

$$V_o = (l/2)\, E \sin\theta. \tag{15.25}$$

One important point is that the antenna responds only to the component of the electric field along the wire. The cross-polarized component does not cause

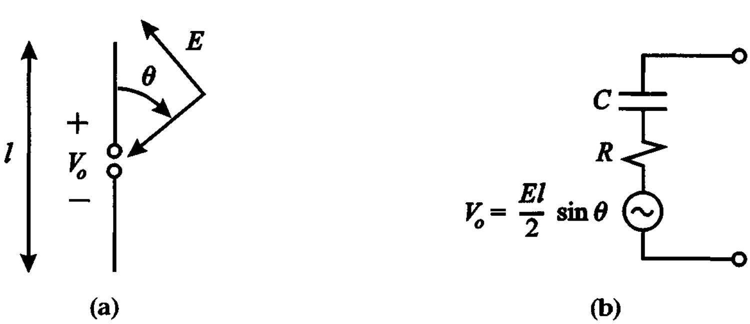


Figure 15.4. Calculating the effective length of a short dipole (a), and the Thevenin equivalent circuit (b).

a voltage. It is true in general that antennas with one pair of output terminals respond only to a single polarization. In our calculations, we will assume that the antenna is oriented to match the polarization of the incident waves. We can write the effective length as

$$h = (l/2)\sin\theta. \tag{15.26}$$

Figure 15.4b shows the Thevenin circuit. In addition to the radiation resistance R, there is also a capacitance C between the wires. We will consider these later.

The effective area A is defined in terms of the available power P_r from the antenna terminals. The r is for "receiver." We write

$$A(\theta,\phi) = \frac{P_r}{S(\theta,\phi)}, \tag{15.27}$$

where S is the incident power density. The effective area is often useful for characterizing large reflector antennas, because the peak effective area turns out to be about half the physical area. In addition, we will see that, by reciprocity, the effective area relates in a simple way to the gain:

$$A(\theta,\phi) = \frac{\lambda^2}{4\pi}G(\theta,\phi), \tag{15.28}$$

where λ is the wavelength. This formula indicates that the angle patterns for receiving and transmitting are the same. It also means that we can use the same antenna for transmitting and receiving in a communications system, and we can measure antenna patterns in either transmission or reception.

We can relate effective area to effective length through the Thevenin equivalent circuit. We write the available power as

$$P_r = \frac{|V_o|^2}{8R} = \frac{|hE|^2}{8R}, \tag{15.29}$$

where R is the antenna resistance. For an incident plane wave we can substitute for E from Equation 15.13 to get

$$P_r = \frac{|h|^2 S\eta_0}{4R}. \tag{15.30}$$

Now, if we compare this formula to the definition of effective area (Equation 15.27), we get

$$A = \frac{|h|^2\eta_0}{4R}. \tag{15.31}$$

15.6 Friis Formula

We can use the gain and effective area to calculate signal levels in a communications system. We consider two antennas separated by a distance r (Figure 15.5). The transmitting antenna has a gain G with transmitter power P_t, and the receiving

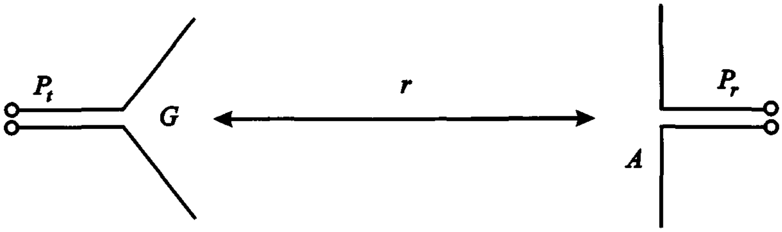


Figure 15.5. Deriving the Friis transmission formula to calculate signal levels transmitted between two antennas.

antenna has an effective area A with available power P_r. We write the far-field power density S as

$$S = \frac{P_t G}{4\pi r^2}. \tag{15.32}$$

The receiver power P_r is given by

$$P_r = SA = \frac{P_t G A}{4\pi r^2}. \tag{15.33}$$

This formula is due to Harald Friis of the Bell Laboratories. In most communications systems, we have additional loss because of ground reflections and absorption and refraction in the atmosphere.

Signals in the HF range from 3 MHz to 30 MHz can propagate long distances, because they reflect off the ionosphere. It is interesting to apply the Friis formula to communication with the NorCal 40A. Antenna gains vary widely, but reasonable values to consider at 7 MHz are $G = 1$ and $A = 150\ \text{m}^2$. For a distance of 2,000 km and a transmitter power of 2 W, we can write

$$P_r = \frac{P_t G A}{4\pi r^2} = 6\,\text{pW}. \tag{15.34}$$

In practice, the loss would likely be somewhat greater than this because of losses in the ionosphere. Figure 15.6 shows typical power levels for a pair of NorCal 40As that are communicating with each other at this distance. The levels span a wide power range, over twelve orders of magnitude. We start in the transmitter, where the VFO output is a microwatt. The signal goes through four successive stages of amplification. We pick up power steadily and transmit two watts from the antenna. At a distance of 2,000 km, we could expect to receive a few picowatts, just above nighttime antenna noise. In the receiver we alternate filters with loss and mixers with gain. The signal level stays low until we reach the audio amplifier, where the power is still below a nanowatt. The audio amplifier gives a huge gain to bring us above a milliwatt, which is a comfortable level for the speaker.

Signals in the VHF range above 30 MHz usually are not reflected by the ionosphere. In this case, the actual range will often not be limited by the Friis formula, but by the curvature of the earth. It helps to have the antenna as high as possible. Figure 15.7a shows the geometry for calculating this line-of-sight limitation, and we can use right-triangle formulas for calculating the relation between the

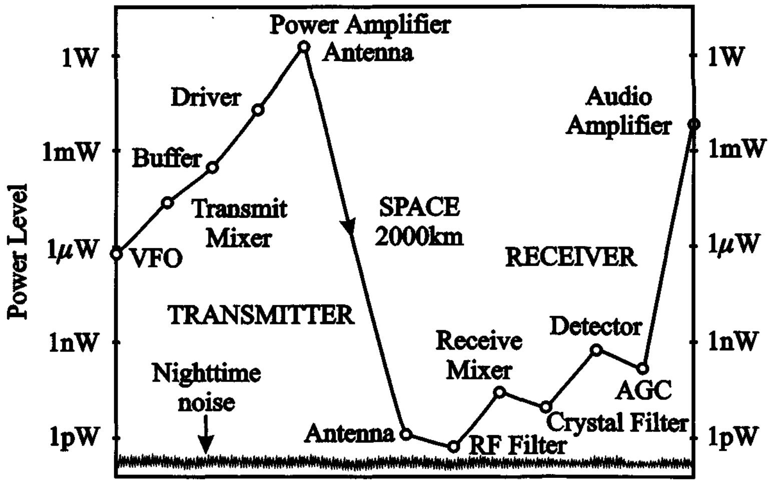


Figure 15.6. Power levels at different stages in a pair of NorCal 40As that are communicating over a distance of 2,000 km.

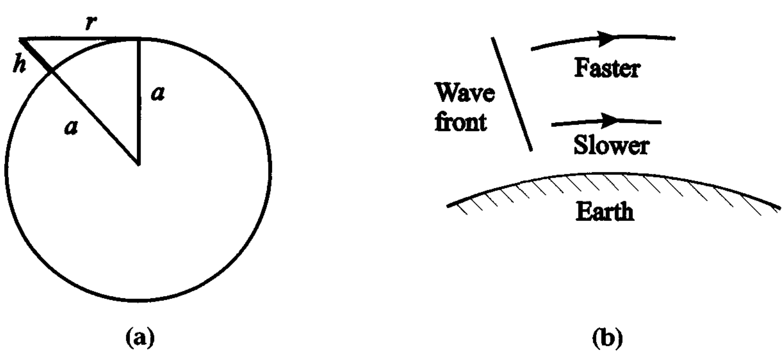


Figure 15.7. Line-of-sight path limit on a curved earth (a). Velocity variation in the atmosphere (b).

antenna height h and the range r to another antenna on the ground. We write

$$(a+h)^2 = a^2 + r^2, \tag{15.35}$$

where a is the radius of the earth, 6,370 km. We can expand and rewrite the formula approximately as

$$r \approx \sqrt{2ah}. \tag{15.36}$$

This formula is easy to apply, but the ranges it predicts are too short. The reason is that the velocity varies in the atmosphere, making the wave follow a curved path (Figure 15.7b). The waves bend toward the denser part of the atmosphere where the velocity is lower, and this gives us longer ranges than we might have expected. The effect varies from time to time and from place to place, but a reasonable approximation is to use an effective earth radius a_e that is 4/3 the actual radius, or 8,500 km. In addition, if the receiving antenna is also elevated, the range increases. We can write the range formula as

$$r = \sqrt{2a_e h_1} + \sqrt{2a_e h_2}, \tag{15.37}$$

where h_1 and h_2 are the antenna heights.

15.7 Antenna Theorem

We found that for a lossless antenna, the integral over all solid angles of the gain was equal to 4π. This formula came directly from the definition of gain. The effective area for lossless antennas satisfies a similar relation, written as

$$\oint A\, d\Omega = \lambda^2. \tag{15.38}$$

This formula is called the *antenna theorem*. It is a beautiful result and not at all obvious. In fact, we might think that we could increase the integral by making the antenna larger. However, this is not so, because the received energy has to be coupled to the output transmission line. Our proof is due to Bernard Oliver of the Hewlett-Packard Laboratories. It requires Nyquist's formula for noise and formulas for blackbody radiation. We consider a lossless antenna in an absorbing cavity at the same temperature T (Figure 15.8). The antenna is connected to a matched load, which is also at temperature T. We consider the power flow between the antenna and the cavity. When the load and the cavity are at the same temperature, they are in thermal equilibrium, which is to say that the temperatures are not changing. This means that the noise power from the load radiated by the antenna must be the same as the power received by the antenna from the cavity, or else the temperatures would change.

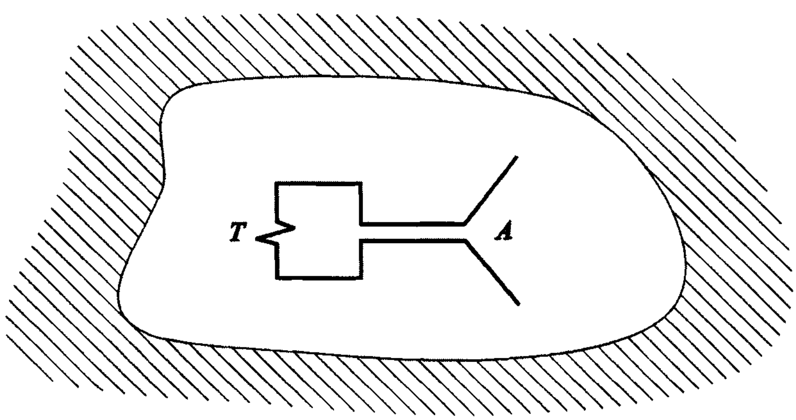


Figure 15.8. Antenna in a radiating cavity. Both the matched load and the cavity are at the same temperature T, and so the system is in thermal equilibrium.

The radiated power from the antenna is Johnson noise from the load, and the power density is given by Nyquist's formula as kT. This power must be the same as the power received by the antenna. To figure out the received power, we need to know the power radiated from the enclosure. This is given by the blackbody radiation law. It is calculated in a manner similar to that for the Nyquist formula we derived in Chapter 14. We use the equipartition theorem to associate the energy kT with each of the resonant modes of a cavity. The quantum-mechanical version is called Planck's formula, and it applies when the photon energy is comparable to the thermal energy. We use the blackbody radiation formula from classical thermodynamics called the Rayleigh–Jeans law, which applies when the photon energy is much smaller than the thermal energy. Counting the resonant modes is difficult, and we will only quote the result here. It is written in terms of E, the energy per unit frequency per unit volume in a single polarization:

$$E = \frac{4\pi kT}{c\lambda^2}. \tag{15.39}$$

This is not a good final form for us, because we can only calculate the solid-angle integral if we know the incident power density per unit solid angle. This quantity is called the brightness B, and its units are W Hz^{-1} m^{-2} steradian^{-1}. We calculate the brightness from the energy density by multiplying by the velocity c, and dividing by the solid angle of a sphere, 4π, to take into account the fact that the power comes from all directions. This gives the brightness as

$$B = kT/\lambda^2. \tag{15.40}$$

We can write

$$kT = \oint BA\,d\Omega = \oint (kT/\lambda^2)A\,d\Omega. \tag{15.41}$$

We can express the received power as an integral such as this because the radiation that comes from different angles effectively comes from different sources and is uncorrelated. This simplifies to

$$\oint A\,d\Omega = \lambda^2, \tag{15.42}$$

which proves the theorem.

15.8 Reciprocity

The reciprocity principle allows us to predict the effect of interchanging the input and the output. For filters, it means that the loss in one direction is the same as the loss in the other. For antennas, reciprocity allows us to relate gain and effective area. However, we cannot just set them equal, because they have different units.

We will see that they are related by the formula

$$G/A = 4\pi/\lambda^2. \tag{15.43}$$

To start, we state, without proof, a form of the reciprocity theorem.

> *The positions of an ideal voltmeter and an ideal current source can be interchanged in a circuit without changing the voltmeter reading.*

Reciprocity is a subtle idea, and there are a long list of conditions that must hold for a circuit to be reciprocal. The network must be linear and time invariant. In addition there can be no bias currents or fields. The conditions for reciprocity usually hold for the antennas themselves. However, the earth's magnetic field sometimes causes the ionosphere to show nonreciprocal behavior. In practice, it is fairly common for the signal-to-noise ratios at the two ends of a radio link to be quite different, even though the transmitter powers are the same. However, this is usually due to different noise and interference levels at the two receivers. This is not itself a violation of reciprocity. In the case we are considering here, where we are just interested in antenna properties, we can assume that the space between the antennas is reciprocal and use the Friis transmission formulas.

Consider the circuit shown in Figure 15.9. The transmitting antenna is driven by a current source I. We let the open-circuit voltage of the receiving antenna be V. We can write the transmitter power P_t as

$$P_t = |I|^2 R_1/2. \tag{15.44}$$

We can write the receiver power P_r using the Friis formula as

$$P_r = \frac{P_t G_1 A_2}{4\pi r^2}. \tag{15.45}$$

Notice that this is available power, rather than delivered power. The delivered power is actually zero, because the output is open-circuited. We can rewrite P_r in terms of the open-circuit voltage V as

$$P_r = \frac{|V|^2}{8R_2}. \tag{15.46}$$

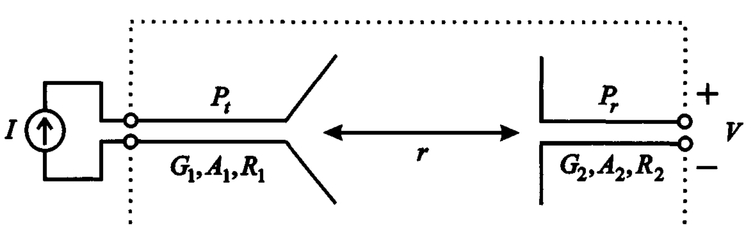


Figure 15.9. Deriving the reciprocal relation between transmitting and receiving antennas. Each antenna is characterized by a gain G, effective area A, and resistance R.

This lets us write the relation

$$\frac{|V|^2}{8R_2} = \frac{|I|^2 R_1 G_1 A_2}{8\pi r^2}. \tag{15.47}$$

Let us regroup to isolate the gain and effective area. We can write

$$G_1 A_2 = \frac{|V|^2 \pi r^2}{|I|^2 R_1 R_2}. \tag{15.48}$$

An interesting thing happens if we move the current generator to the second antenna. By reciprocity, the voltage at the first antenna must be V, and in fact, the entire right side of the equation does not change. However, on the left side, we swap the indices 1 and 2. Thus

$$G_1 A_2 = G_2 A_1 \tag{15.49}$$

or, as quotients

$$G_1/A_1 = G_2/A_2. \tag{15.50}$$

Notice that the left side depends only on antenna 1, and the right side depends only on antenna 2. At this point, we have not specified anything about the antennas or how they are pointed. This means that the ratio of gain to effective area is a universal constant, independent of the antenna orientation or type. It does not depend on whether the antenna is lossless or not.

To find the constant, we consider a particular class of antennas, lossless antennas, and their solid-angle integrals. We saw that from the definition of gain,

$$\oint G\, d\Omega = 4\pi, \tag{15.51}$$

and from the antenna theorem,

$$\oint A\, d\Omega = \lambda^2. \tag{15.52}$$

If the ratio G/A is a constant, independent of angle, then the ratio of the two solid-angle integrals will give us the same constant. We get

$$G/A = 4\pi/\lambda^2. \tag{15.53}$$

15.9 Dipoles

We can use the antenna theorem to find the radiation resistance R_r of a short, lossless dipole. We substitute for the effective area in terms of the effective length from Equation 15.31 and write

$$\lambda^2 = \oint A\, d\Omega = \oint \frac{|h|^2 \eta_0}{4R_r}\, d\Omega. \tag{15.54}$$

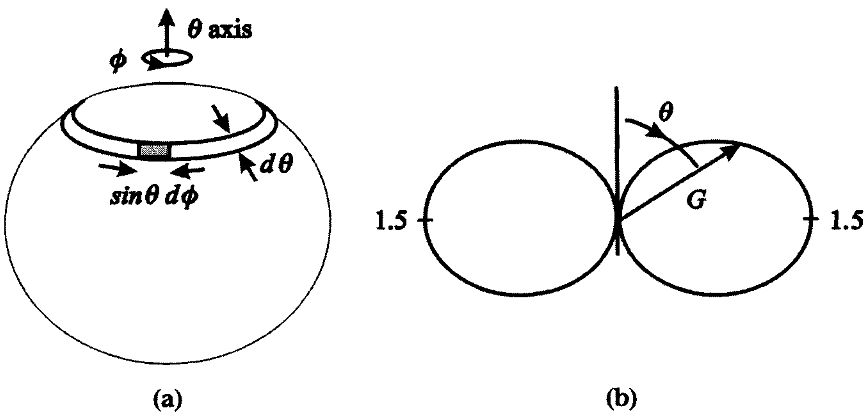


Figure 15.10. (a) Converting a solid-angle integral to the spherical coordinates θ and ϕ. We show a rectangular solid-angle element $d\Omega$ on the surface of a unit sphere. The side lengths can be written as $\sin\theta\, d\phi$ and $d\theta$. (b) Polar plot of the gain pattern for a short dipole antenna.

Substituting for h from Equation 15.26, we write R_r as

$$R_r = \frac{\eta_0 l^2}{16\lambda^2} \oint \sin^2\theta\, d\Omega. \tag{15.55}$$

For the solid-angle integration element $d\Omega$, we can substitute in terms of θ and ϕ (Figure 15.10a):

$$d\Omega = \sin\theta\, d\phi\, d\theta. \tag{15.56}$$

This gives

$$R_r = \frac{\eta_0 l^2}{16\lambda^2} \int_0^{2\pi} d\phi \int_0^{\pi} \sin^2\theta \sin\theta\, d\theta. \tag{15.57}$$

Because there is no ϕ variation, we get a factor of 2π when we integrate over ϕ:

$$R_r = \frac{\eta_0 \pi l^2}{8\lambda^2} \int_0^{\pi} \sin^2\theta\, \sin\theta\, d\theta. \tag{15.58}$$

The θ integral is best attacked by making the substitution $x = -\cos\theta$. This makes the integrand algebraic. The integration element dx is given by

$$dx = \sin\theta\, d\theta. \tag{15.59}$$

We can rewrite the integral as

$$R_r = \frac{\eta_0 \pi l^2}{8\lambda^2} \int_{-1}^{+1} (1 - x^2)\, dx. \tag{15.60}$$

The integral has the value 4/3, so that the radiation resistance becomes

$$R_r = \eta_0(\pi/6)(l/\lambda)^2. \tag{15.61}$$

If we substitute 120π for η_0, we get

$$R_r = 20\pi^2(l/\lambda)^2. \tag{15.62}$$

The radiation resistance is proportional to the length squared. We can substitute back into Equation 15.31 to find the effective area as

$$A = \frac{3\lambda^2}{8\pi}\sin^2\theta. \tag{15.63}$$

We can find the gain by multiplying by $4\pi/\lambda^2$ to get

$$G = 1.5\sin^2\theta. \tag{15.64}$$

This pattern is shown as a polar plot in Figure 15.10b. The maximum gain, 1.5, is at $\theta = 90°$.

Dipoles have a resonance when each arm is a quarter wavelength long, or when the total length is a half wavelength. From a transmission-line point of view, this is what you might expect. We could consider the dipole as an opened-out transmission line, and we would expect a series resonance when the length of the transmission line is a quarter wavelength. Actually the resonant length for practical dipoles is a little shorter than this, about 0.48λ, once the effect of the wire thickness and end supports are taken into account. In practice, the resonant length will also be affected by the presence of the ground; hence dipoles often have to be adjusted for resonance after they are put up. We can estimate the resonant resistance for a dipole by setting $l = \lambda/2$ in Equation 15.62. We get $R_r = 49\ \Omega$. This is a fair estimate of the actual value. A more detailed theory gives $73\ \Omega$, but effects such as the presence of the ground, wire thickness, and insulation conspire to lower it to nearer $50\ \Omega$ in practice. This is a convenient resistance for connecting 50-Ω coaxial cable. Usually, however, we would not connect a coaxial cable directly to a dipole. The problem is that currents run down the outside of the shield and affect both the impedance and the pattern. A transformer called a *balun* is used to isolate the outside of the shield from the antenna.

15.10 Whip Antennas

A common variation on the dipole is a single vertical wire above a ground plane (Figure 15.11a). This is called a *monopole*, or informally, a *whip* or *stub*. The calculations are similar to those for a dipole, except for some factors of two. The effective length doubles because the reflection off the ground plane doubles the component of the electric field along the wire. This effect by itself would raise the radiation resistance by a factor of 4. The pattern above the ground plane is the same as that for the dipole, but no power is received for angles below the ground plane. The θ integral in Equation 15.58 becomes an integral from 0 to $\pi/2$ rather than from 0 to π. This reduces the radiation resistance by a factor of 2. The net

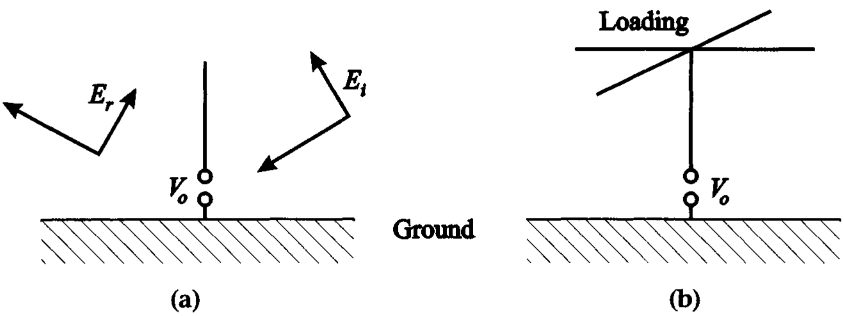


Figure 15.11. (a) Monopole above a ground plane, showing incident and reflected waves. The normal component of the reflected electric field is the same as that of the incident field, and this means that the resulting parallel to the dipole is twice the incident field. (b) Loading a whip antenna.

effect is to double the radiation resistance:

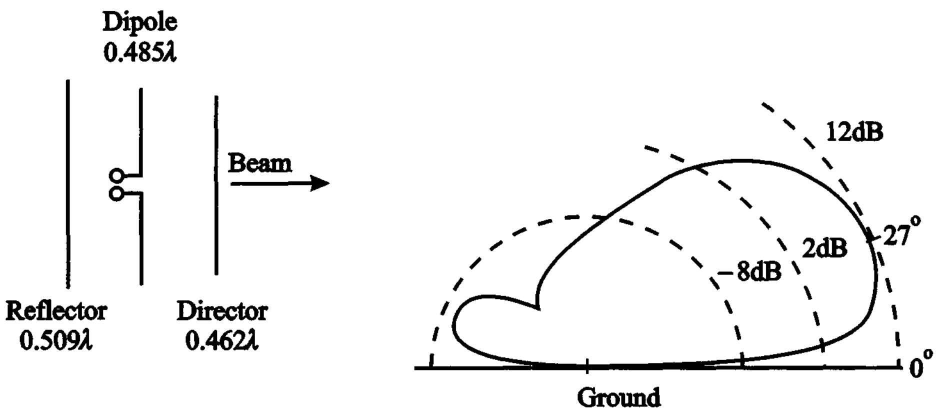
$$R_r = 40\pi^2(l/\lambda)^2. \tag{15.65}$$

A monopole is resonant when the length is a quarter of a wavelength. At low frequencies, it may not be convenient to make the whip this long. For example, the NorCal 40A operates in the 40-meter band. A quarter wavelength is 10 meters, which is too long for a car antenna. However, if we add additional wires at the top, the potential of the antenna can approach the potential at the top rather than half way up (Figure 15.11b). This is called *loading*. Loading doubles the effective length and quadruples the radiation resistance. We can write the radiation resistance for a loaded whip as

$$R_r = 160\pi^2(l/\lambda)^2. \tag{15.66}$$

We can also load the ends of dipoles. In Hertz's original experiments, the dipoles had spheres mounted on each end that had the effect of loading.

The dipoles and monopoles we have discussed so far have low gain. In 1926, Shintaro Uda, a professor at Tohoku University in Sendai, Japan, discovered an ingenious way to increase the gain. He added additional wires in front and behind the dipole (Figure 15.12a). These wires are called *parasitic elements*. They are not directly connected. However, currents are excited on the wires, and these currents reradiate energy with a phase that is determined by the length of the element. One parasitic wire is made shorter than the resonant length, and it acts as a *director*, strengthening the radiated beam in the forward direction. The other element is made somewhat longer than the resonant length, and it acts as a *reflector*, reducing the radiation in the back direction. With the correct dimensions, the gain can be increased substantially. These antennas are usually called Yagi antennas, after Hidetsugu Yagi, who was Uda's laboratory director at Tohoku, and who reported Uda's work in the United States. You may also hear them called Yagi–Uda antennas. The mechanical construction of Yagi antennas is quite simple, and this has made them popular around the world for communications and television.



(a) (b)

Figure 15.12. (a) Yagi antenna. (b) Calculated gain as a function of the elevation angle at 7 MHz for this antenna when it is 20 m above the ground. The array elements are 6 meters apart and have a diameter of 16 mm. This calculation was done with the *EZNEC* program.

The behavior of a Yagi antenna is complicated enough that you need a computer program to predict the patterns. Figure 15.12b shows the elevation pattern for the Yagi antenna in Figure 15.12a. The peak gain is 12 dB. The radiation behind the antenna is suppressed by more than 20 dB. The peak in the pattern is at an elevation of 27° for a 7-MHz antenna. The peak angle is determined by the height above the ground. It is the angle at which the reflected wave from the ground is in phase with the direct wave. This pattern is suitable for long-distance communication by reflection of radio waves off the ionosphere.

15.11 Ionosphere

Early investigators believed that the range of radio transmission would be completely limited by the curvature of the earth. However, radio waves can reflect off ionized electrons in the upper atmosphere. In fact, radio waves can bounce repeatedly, making worldwide radio communications possible. In the upper atmosphere, ultraviolet radiation and X rays from the sun strip electrons from atoms and molecules. This part of the atmosphere is called the *ionosphere*. Figure 15.13 shows a typical daytime electron-density plot. The peak electron density occurs at an altitude of 300 km. The electron density falls off at both higher and lower altitudes. At higher altitudes there are not many atoms and molecules to ionize. At lower altitudes, ultraviolet radiation and X rays become less intense, because they are absorbed in propagating through the atmosphere. The peaks on the plot are given the letter names D, E, F1, and F2. The more energetic ionizing radiation penetrates deeper into the ionosphere, so that X rays are primarily responsible for the D and E layers, and ultraviolet radiation is primarily responsible for the F layers.

The D layer is at an altitude of 70 km. The maximum ionization level occurs at noon each day and is about 10^{10} electrons per cubic meter. The D layer disappears

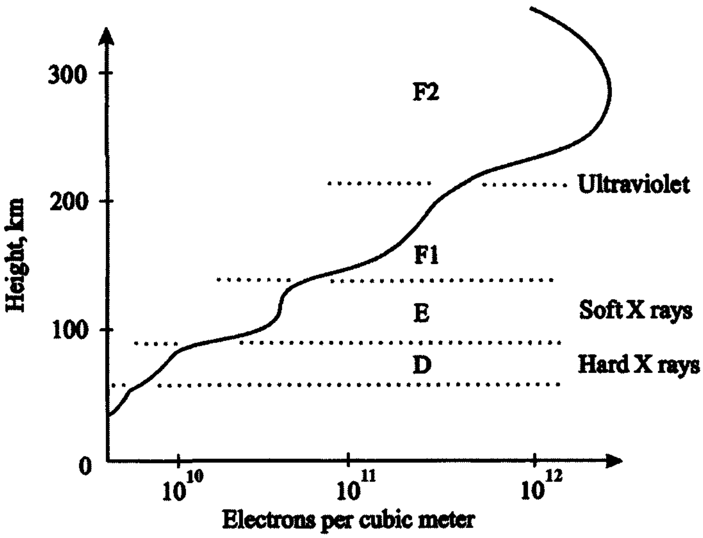


Figure 15.13. Electron density in the ionosphere on a summer day. The layers were named by Sir Edward Appleton in the 1920s. Historically, the first layer that was discovered was called the *electric* layer, or E layer for short. Then new layers were found below and above the E layer, and so these became the D and F layers. Adapted with permission from Figure 1.1 in *Ionospheric Radio*, by Kenneth Davies, published by Peter Peregrinus.

at sundown and reappears at sunrise. The D layer is responsible for most of the absorption in the ionosphere, and it prevents frequencies below 10 MHz from being used for long-range communications in the daytime.

The E layer is at 120 km. The maximum ionization, also at midday, is about 10^{11} electrons per cubic meter. Frequencies up to 15 MHz can be reflected 1,500 km in a single hop. Like the D layer, the E layer disappears at night. In addition, there are patches of ionization at the altitude of the E layer that are called *sporadic E*. The ionization in these patches can be so intense that frequencies up to 100 MHz can be reflected. The patches are about 100 km across, and they do not appear to be caused by solar radiation. One cause of sporadic E is friction at shear layers between fast moving air masses. This is the atmospheric equivalent of making sparks by scraping your shoes on the carpet. It occurs irregularly in summer months. Meteors also cause sporadic E, and this has been used for data transmission from remote weather stations.

The F layers have the highest ionization levels, and they are the main layers used for worldwide broadcasting and communications. Frequencies up to 50 MHz may be reflected 3,000 km in a single hop. The F1 layer disappears in the winter and at night, and the single layer that remains is just called the F layer. The typical peak ionization is about 10^{12} in the day at an altitude of about 300 km. At night the layer may drop 50 km in altitude and a factor of ten in ionization, and frequencies up to 15 MHz may be used. Unlike the E and D layers, whose ionization is determined by the position of the sun, many factors affect the F layer. Electron recombination is quite slow at these high altitudes, and the electron densities are strongly affected by diffusion from lower layers and long-distance drift. Because of this, there is still an F layer in the polar regions in the winter, even though there is no sun for months.

Solar activity has a strong effect on the F layer, and measurements of solar activity are as important for long-distance broadcasting and communications as weather reports are for farmers. The National Institute of Standards and Technology (NIST) broadcasts microwave solar power densities every hour from the standard-time stations WWV and WWVH. This microwave flux is measured every day at noon at 2.8 GHz at the Algonquin Radio Observatory in Algonquin Park, Ontario, Canada. It varies from a minimum of 65×10^{-22} $Wm^{-2}Hz^{-1}$ to a maximum of around 200×10^{-22} $Wm^{-2}Hz^{-1}$. In the broadcast, the exponent and units are dropped, and the announcer will just say 65 for the minimum value and 200 for the maximum. The 2.8-GHz radiation does not ionize the upper atmosphere. However, it is easy to measure, and surprisingly, it correlates well with the ultraviolet and X-ray fluxes, which are much more difficult to measure because they are absorbed in the ionosphere. Solar flux levels tend to repeat each solar rotation period (27 earth days), indicating that the areas on the sun that are responsible for the radiation last for quite a while. Solar flux also correlates with the number of sunspots (Figure 15.14). Sunspots are cool (at least by solar standards) regions with temperatures of 3,000 K or so, compared with 6,000 K, which is typical of the rest of the sun's surface. It might seem surprising that the number of sunspots correlates strongly with solar activity. However, sunspots have enormous magnetic fields, in the range of 0.4 teslas, which is similar to the magnetic field of a large laboratory magnet, and these magnetic fields have a strong effect on the solar plasma. Sunspots have a dramatic 11-year cycle (Figure 15.15). Near the peak of the cycle, frequencies as high as 50 MHz can be used for long-distance

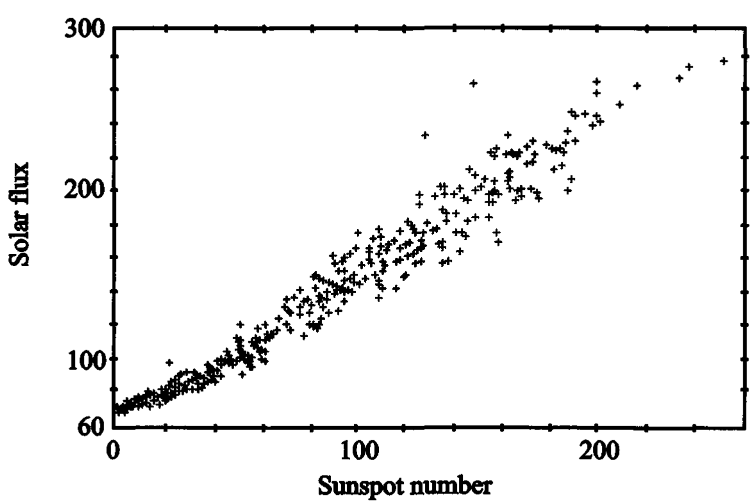


Figure 15.14. Monthly solar flux averages, compared with monthly averages of the sunspot number. Reprinted with permission from Chapter 23 of the *ARRL Antenna Book*, published by the American Radio Relay League.

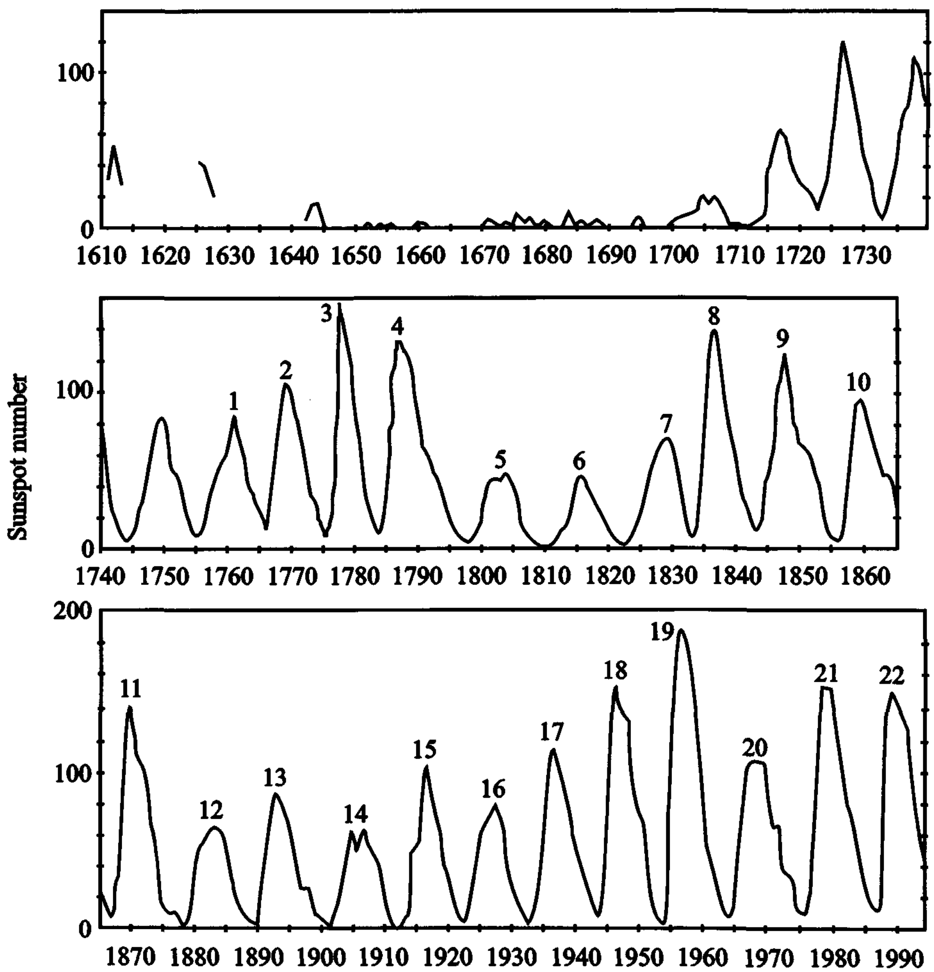


Figure 15.15. Annual sunspot averages going back to 1610. The sunspot number is found by counting the number of individual sunspots and the number of groups of sunspots. The sunspot number is traditionally calculated as the sum of the number of individual sunspots and ten times the number of sunspot groups. Near the minimum, the sunspot number is close to zero, and near the maximum, it can be above 200. The sunspot number is a crude way to measure solar activity, but measurements are available for almost 400 years. The extended quiet period during the 1600s is called the *Maunder minimum*. The cycles are numbered consecutively, starting with the cycle that began in 1755. Reprinted with permission from Figure 2.3 in *Ionospheric Radio*, by Kenneth Davies, published by Peter Peregrinus.

ionospheric communications. Near the sunspot minimum, frequencies higher than 15 MHz can be used only sporadically.

15.12 Radio Waves in the Ionosphere

To investigate how radio waves propagate in the ionosphere, we use Newton's second law to find an expression for the conductivity of the ionized electrons.

Newton's second law says that the applied force is equal to the rate of change of momentum. The force on a single electron due to an electric field is $q\mathbf{E}$, where q is the electronic charge, 1.6×10^{-19} C. Neglect the effects of magnetic fields and collisions, and we can write

$$q\mathbf{E} = m\frac{d\mathbf{v}}{dt}, \tag{15.67}$$

where m is the electron mass, 9.1×10^{-31} kg, and $\mathbf{v}$ is the velocity. We can rewrite this formula in phasor notation, replacing the time derivative by $j\omega$:

$$q\mathbf{E} = j\omega m\mathbf{v}. \tag{15.68}$$

We can solve for the velocity as

$$\mathbf{v} = \frac{q}{j\omega m}\mathbf{E}. \tag{15.69}$$

We relate the velocity to the current density $\mathbf{J}$ by multiplying by the charge density Nq, where N is the number of electrons per cubic meter. This gives us

$$\mathbf{J} = Nq\mathbf{v} = \frac{Nq^2}{j\omega m}\mathbf{E}. \tag{15.70}$$

The current lags the electric field because of the inertia of the electrons. Now we can see how this affects radio-wave propagation by including the current density $\mathbf{J}$ in Ampère's law. We write

$$\nabla \times \mathbf{H} = \mathbf{J} + j\omega\epsilon_0\mathbf{E} \tag{15.71}$$

and substitute for $\mathbf{J}$ to get

$$\nabla \times \mathbf{H} = \frac{Nq^2}{j\omega m}\mathbf{E} + j\omega\epsilon_0\mathbf{E}. \tag{15.72}$$

We can rewrite this equation in the form

$$\nabla \times \mathbf{H} = j\omega\epsilon\mathbf{E}, \tag{15.73}$$

where the effective permittivity ϵ is given by

$$\epsilon = \epsilon_0 - \frac{Nq^2}{\omega^2 m}. \tag{15.74}$$

We can then rewrite the phase constant β using Equation 15.11 as

$$\beta = \beta_0\sqrt{1 - \frac{Nq^2}{\epsilon_0\omega^2 m}}, \tag{15.75}$$

where β_0 is the phase constant for free space. The *refractive index* n is given by

$$n = \frac{\beta}{\beta_0} = \sqrt{1 - \frac{Nq^2}{\epsilon_0\omega^2 m}}. \tag{15.76}$$

The refractive index in the ionosphere is less than 1, which seems strange, because it means that the wavelength in the ionosphere is longer than in free space. Notice that the expression contains the electron mass in the denominator, indicating that the effect comes from the inertia of the ionized electrons.

15.13 Critical Frequency

To simply Equation 15.76, we define a *critical frequency* f_c given by

$$f_c = \frac{1}{2\pi}\sqrt{\frac{Nq^2}{\epsilon_0 m}}. \tag{15.77}$$

If we substitute for the constants, we find

$$f_c \approx 9.0\sqrt{N}. \tag{15.78}$$

We can rewrite Equation 15.76 as

$$n = \sqrt{1 - \left(\frac{f_c}{f}\right)^2}. \tag{15.79}$$

At frequencies below the critical frequency, the propagation constant becomes imaginary, and the wave attenuates. The wave is said to be *evanescent*. Physically, the power is reflected.

The critical frequency can be measured by transmitting pulses from a radar *sounder* straight up, and listening for reflections. Typically the pulses are 30 μs long and have a peak power in the range of 1 kW to 10 kW. The delay in the reflection tells us the height of the layer. Plots of the reflection delay versus frequency are called *ionograms*. Figure 15.16 shows a summer ionogram, with the E, F1, and F2 layers are clearly visible. The E layer is quite sharp, at a height of 100 km, with a critical frequency of 3.8 MHz. The F1 layer is at 200 km, with a critical frequency of 4.9 MHz. The F2 layer is at 370 km, with a critical frequency of 6.7 MHz.

Figure 15.17 shows how solar activity affects f_c. The E-layer critical frequency increases from 3 MHz to 4 MHz as the sunspot number moves from zero to 200. For the F1 layer, f_c increases from 4 MHz to 5 MHz. At higher sunspot numbers, the F1 layer is not always distinct. The most dramatic change is in the F2 layer, where f_c doubles from 5.5 MHz to 11 MHz.

Figure 15.18 shows how the critical frequencies depend on the time of day. For the F2 layer, f_c is relatively constant in the summer, but it peaks sharply in the winter in the early afternoon. Surprisingly, the highest critical frequencies are in the winter, even though the sun is at a lower angle than in the summer. This is a complicated effect involving collisions with molecules in the upper ionosphere. The upper ionosphere is primarily atomic oxygen. However, the density of molecular oxygen increases in the summer, and this reduces the electron recombination time, causing the ionization level to drop.

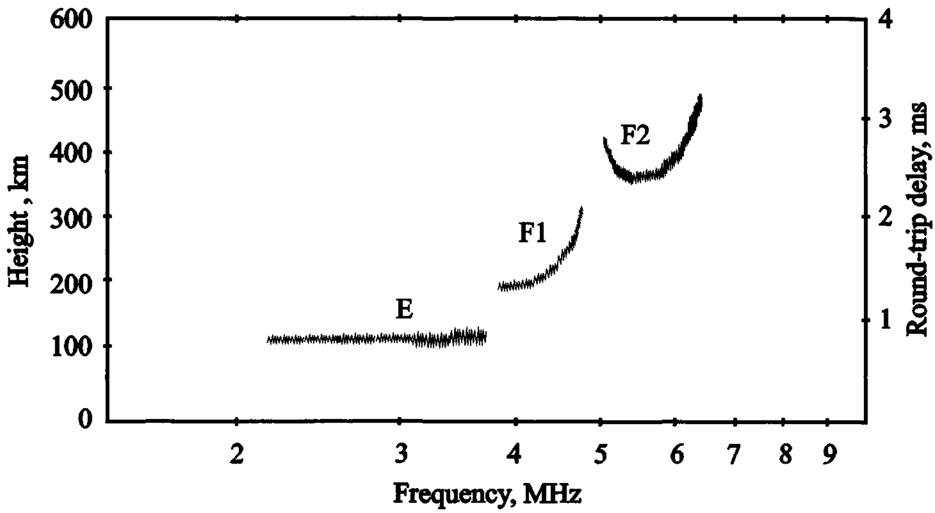


Figure 15.16. Summer daytime ionogram for Boulder, Colorado. The local time is 1:30 pm. The right axis gives the round-trip delay, and the left axis gives the apparent height. Reflection from the ionosphere causes additional delay. This causes the apparent height to be larger than the actual height, particularly near a layer boundary. Redrawn with permission from Figure 4.4a in *Ionospheric Radio*, by Kenneth Davies, published by Peter Peregrinus.

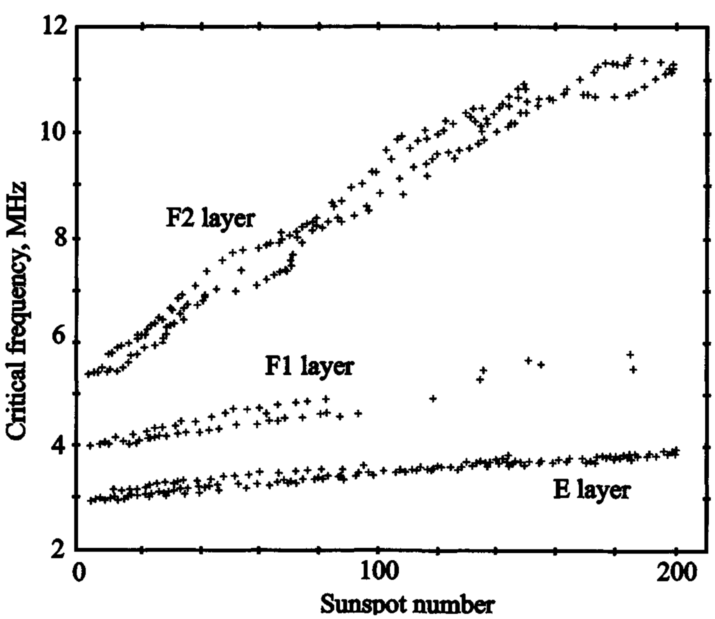


Figure 15.17. Critical frequencies for Washington, DC, for the summer noon F2 layer (upper data points), F1 layer (middle data points), and E layer (lower data points), plotted against the sunspot number over two sunspot cycles. Reprinted with permission from Figure 5.13 in *Ionospheric Radio*, by Kenneth Davies, published by Peter Peregrinus.

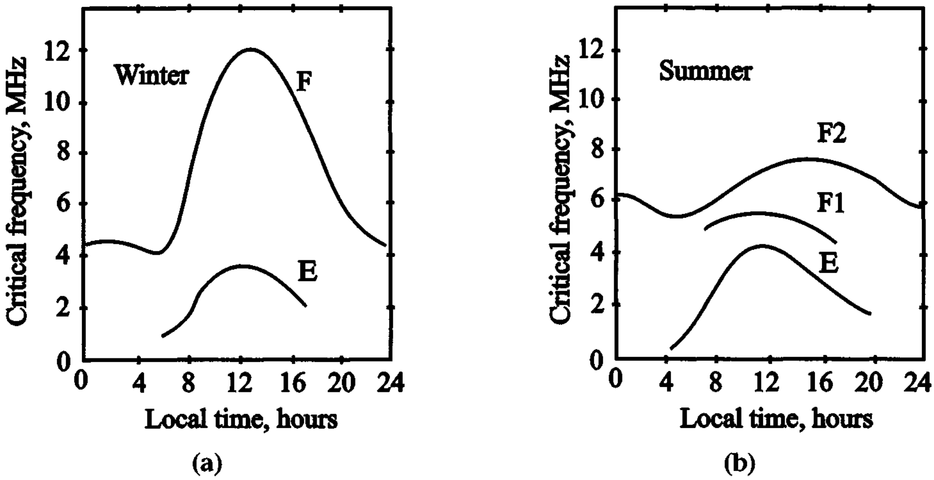


Figure 15.18. Daily variation of critical frequencies for winter (a) and summer (b). From Figure 9.2 in *Radio Wave Propagation*, by Lucien Boithias, published by McGraw-Hill.

15.14 Maximum Usable Frequency

You may have studied Snell's law in optics, which relates the propagation angles for a wave incident on a surface between layers with two refractive indexes (Figure 15.19). It is written as

$$n_i \sin\theta_i = n_t \sin\theta_t, \tag{15.80}$$

where n is the refractive index and θ is the angle from the normal. The subscript i is for the *incident* wave and the subscript t is for the *transmitted* wave. This bending of waves is called *refraction*. We can state Snell's law in a more general way:

$$n \sin\theta \textit{ is a constant.} \tag{15.81}$$

This form also applies if there are more layers or if the refractive index varies continuously. There is also an interesting possibility that occurs if n_i is larger than n_t. We can have

$$n_i \sin\theta_i > n_t. \tag{15.82}$$

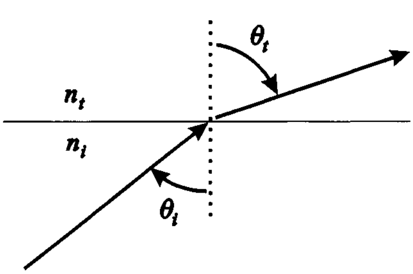


Figure 15.19. Snell's law in optics for transmission through an interface. The arrows show the propagation directions for the incident wave and the transmitted, or refracted, wave.

This makes it impossible to satisfy the Snell's law formula. We define a *critical* incident angle θ_c given by

$$\sin\theta_c = n_t/n_i, \tag{15.83}$$

where $\theta_t = 90°$. At larger incident angles, the wave is completely reflected. This allows us to make optical fiber waveguides by enclosing one dielectric by another with a lower refractive index. The waves bounce back and forth along the guide without loss as long as the incident angle is larger than the critical angle.

Because the refractive index of the ionosphere is less than 1, we can have a critical angle when radio waves are incident from the lower atmosphere. If we take the refractive index of the lower atmosphere to be 1, we can write the critical angle θ_c as

$$\sin\theta_c = n, \tag{15.84}$$

where n is the refractive index of the ionosphere. We can relate the critical angle to the critical frequency by substituting from Equation 15.79:

$$\sin\theta_c = \sqrt{1 - \left(\frac{f_c}{f}\right)^2}, \tag{15.85}$$

or

$$\cos\theta_c = f_c/f. \tag{15.86}$$

We can interpret this formula in two different ways. For a fixed frequency, we can think of it as determining the minimum incident angle that is completely reflected. For a fixed incident angle, we can calculate the maximum frequency that is completely reflected. This is called the *maximum usable frequency* f_m, and we write it as

$$f_m = f_c/\cos\theta_i. \tag{15.87}$$

We should realize that because of the curvature of the earth, the angle of incidence on the ionosphere is not the same as the launch angle on earth. The geometry is shown in Figure 15.20. We can use the sine law to relate θ_i and the launch angle θ:

$$\sin\theta = (1 + h/a)\sin\theta_i, \tag{15.88}$$

where a is the radius of the earth, 6,370 km, and h is height of the ionosphere. We can write the range r in terms of these angles as

$$r = 2a\phi = 2a(\theta - \theta_i). \tag{15.89}$$

We can use these formulas to relate the maximum usable frequency f_m and the range r to the launch angle θ when the F layer is at a height of 300 km; see Figure 15.21. The corresponding elevation angle is shown at the bottom. This plot shows that waves can be reflected at much higher frequencies than the critical frequency. The largest value of f_m, $3.4f_c$, and the greatest range, 3,840 km, are for low

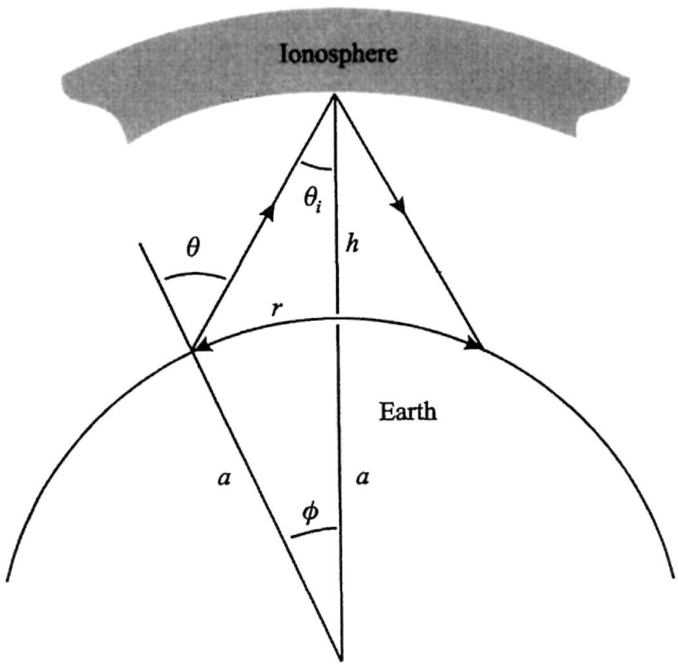


Figure 15.20. Relating the angle of incidence on the ionosphere θ_i to the launch angle θ on earth.

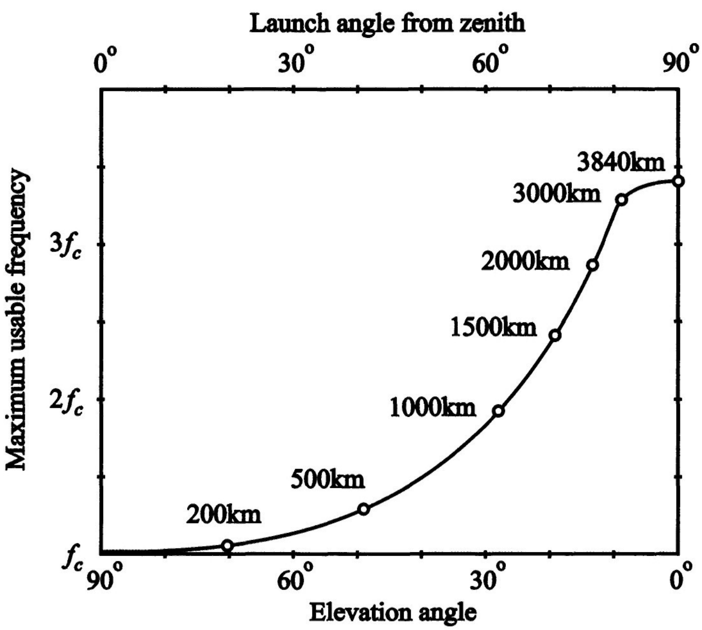


Figure 15.21. Maximum usable frequency f_m versus elevation angle for an F-layer height of 300 km. This curve is calculated from Equations 15.87 through 15.89. The range for a single reflection is marked on the curve. In practice, radio waves bounce repeatedly, and much greater ranges are possible.

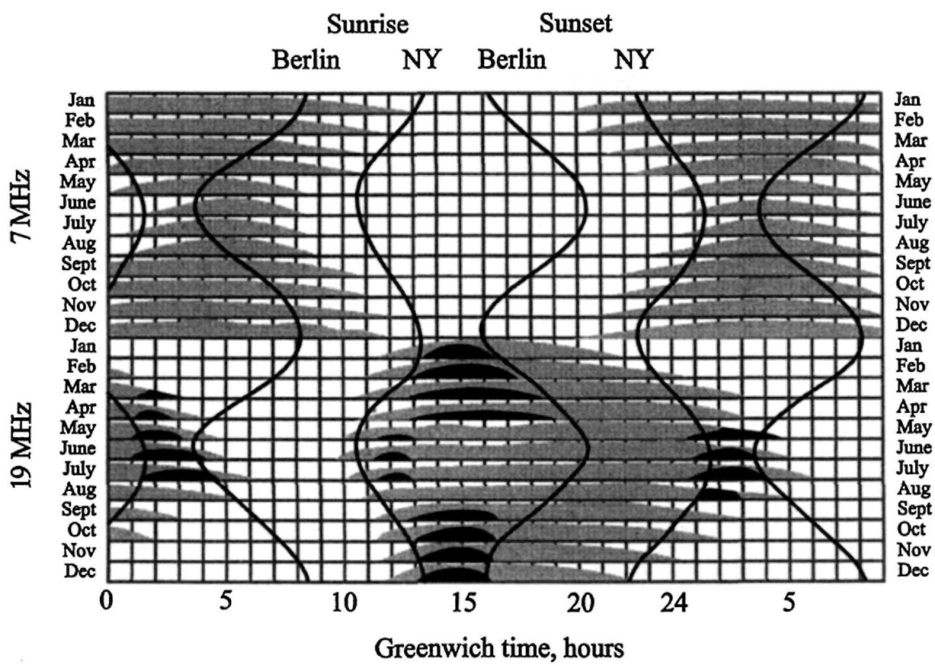


Figure 15.22. Received signal levels between Berlin and New York, measured over an entire year. The signal strength is indicated by the height of the vertical lines. This is Figure 45.4 in *Radio-Wave Propagation and the Ionosphere,* Volume 2, by Y. L. Al'pert, published by the Consultants Bureau, a division of the Plenum Publishing Company.

elevation angles. A peculiar feature of ionospheric radio communications is that when the frequency is greater than f_c, there is a minimum range. As the range decreases, the incident angle on the ionosphere becomes less than the critical angle, and the wave passes through the ionosphere without reflection. This minimum range is called the *skip* distance. The skip distance can also be determined from the figure. For example, if we operate at $2f_c$, the maximum elevation angle is 25°, and the skip distance is 1,100 km. To communicate at closer distances, we would need to reduce the operating frequency.

Figure 15.22 shows a plot of signal reception reports between Berlin and New York over an entire year for 7 MHz and 19 MHz. The distance is 6,000 km, so that two or three reflections from the F layer are needed. The two frequencies show many different effects. The 7-MHz signal is greatly absorbed by the D layer. This means that it has poor propagation during the day. However, the frequency is low enough that it is still reflected throughout the night, even as the electron concentration of the F layer drops. In contrast, 19 MHz is high enough that the signals are not significantly absorbed by the D layer. However, at night, the critical frequency drops and 19-MHz signals fade away. These patterns hold well when there is daylight all along the path, or darkness all along the path. The patterns follow the changes in the hours of sunrise and sunset. Notice also that 19-MHz propagation continues for several hours after sunset in Berlin. This is because the electron concentration in the F layer decays slowly after the sun sets.

FURTHER READING

The book *Antennas and Radio Wave Propagation,* by Robert Collin, published by McGraw-Hill, gives excellent coverage of both antennas and propagation. The *ARRL Antenna Book,* published by the American Radio Relay League, is good for information on building antennas. For studying wire antennas, simulation software is essential. My favorite program is called *EZNEC.* It is available from Roy Lewallen, P.O. Box 6658, Beaverton, OR 97007, w7el@teleport.com. A good reference for the detailed derivation of the blackbody radiation formula is given in *Thermal Physics,* by Charles Kittel, published by W. H. Freeman. Hertz's experiments were extraordinary, and his own descriptions of his experiments make good reading even today. See *Electric Waves,* by Heinrich Hertz, published by Dover. In addition, John Bryant's short book, *Heinrich Hertz, the Beginning of Microwaves,* published by the Institute of Electrical and Electronics Engineers, gives an excellent description of Hertz's experiments. For a proof of the reciprocity theorem, see Desoer and Kuh's *Basic Circuit Theory,* published by McGraw-Hill. The story of the discovery of the cosmic background radiation is told by Steven Weinberg in the *The First Three Minutes,* published by Bantam Books. Shintaro Uda has written an interesting book about the development of the Yagi–Uda antenna, *Short Wave Projector, Historical Records of My Studies in Early Days,* which has been privately published in English.

The standard reference on propagation in the ionosphere is *Ionospheric Radio,* by Kenneth Davies, published by Peter Peregrinus. You can learn more about early radio experiments in *200 Meters and Down* by Clinton DeSoto, published by the American Radio Relay League. Computer predictions of radio-wave propagation can be extremely helpful. I use *Miniprop Plus,* by Sheldon Shallon, published by W6EL Software, 11058 Queensland Street, Los Angeles, CA 90034–3029. The book, *Morse Code, the Essential Language,* by Peter Carron, published by the American Radio Relay League, has more information about the history and use of the code. For a discussion of Snell's law from the perspective of the principle of least time, see *The Feynman Lectures on Physics,* Volume 1, Chapter 26.

PROBLEM 37 - ANTENNAS

A. Use the relation between gain and effective area to rewrite the Friis transmission formula in terms of gain only. Consider a UHF voice radio system for communicating between airplanes. Assume that whatever frequency is finally chosen, the quarter-wave stub antennas will have a gain of 2. Find the maximum possible line-of-sight range between two airplanes at an altitude of 10 km. The required receiver power P_r is –90 dBm. Find the minimum transmitter power P_t required for successful transmission at this range for 100 MHz, 300 MHz, and 1 GHz.

Let us consider a whip antenna for using the NorCal 40A in a car (Figure 15.23). Tuning and matching are done by a tapped inductor, which is a coil with a third wire connected at one of the intermediate turns. This is a transformer where the primary is a part of the secondary. People call this an *autotransformer,* and it acts like an ordinary transformer,

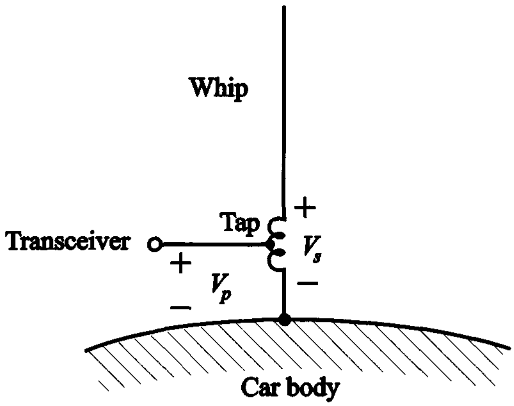


Figure 15.23. Whip antenna with a tapped inductor for using the NorCal 40A in a car.

except that there is no DC isolation. The whip capacitance C is given approximately by

$$C \approx l\epsilon_0, \tag{15.90}$$

where l is the length.

B. Find the inductance required to resonate a 3-m whip. Assuming that the Q of the coil is 200, find the turns ratio required to give the transceiver a 50-Ω load. What is the radiation efficiency?

C. Repeat these calculations for a whip with capacitive end loading, assuming that the capacitance doubles.

PROBLEM 38 – PROPAGATION

The HF band from 3 MHz to 30 MHz is used by thousands of international broadcasters, ships, and amateurs. Many different types of transmissions can be heard, including AM, FM, single sideband (SSB), frequency-shift-keying (FSK), and the oldest of all – Morse Code.

Ionospheric reflection and absorption depend strongly on the frequency, latitude, solar activity level, the time of day, and the season of the year. Signal levels can change quickly, particularly at sunrise and sunset. In addition, on rare occasions, ionospheric storms shut down the HF bands for days. For this reason, some users have switched to microwave satellite links, which are not usually affected by ionospheric disturbances. Others, like amateurs and listeners to international broadcasters, find that the ionospheric "weather" makes the HF bands interesting.

One way to investigate propagation is to listen to *beacon* stations. These are stations that transmit throughout the day and night on a standard frequency. We can tune to these frequencies to find out how radio waves are propagating. Figure15.24 gives reception reports for several beacon stations, taken at hourly intervals over a 24-hour period. The first four stations all transmit at the same frequency, 14.1 MHz, but from different locations. These stations transmit Morse code call signs at regular intervals in a fixed sequence. In addition, the stations transmit a sequence of four tones at power levels of 100 W, 10 W,

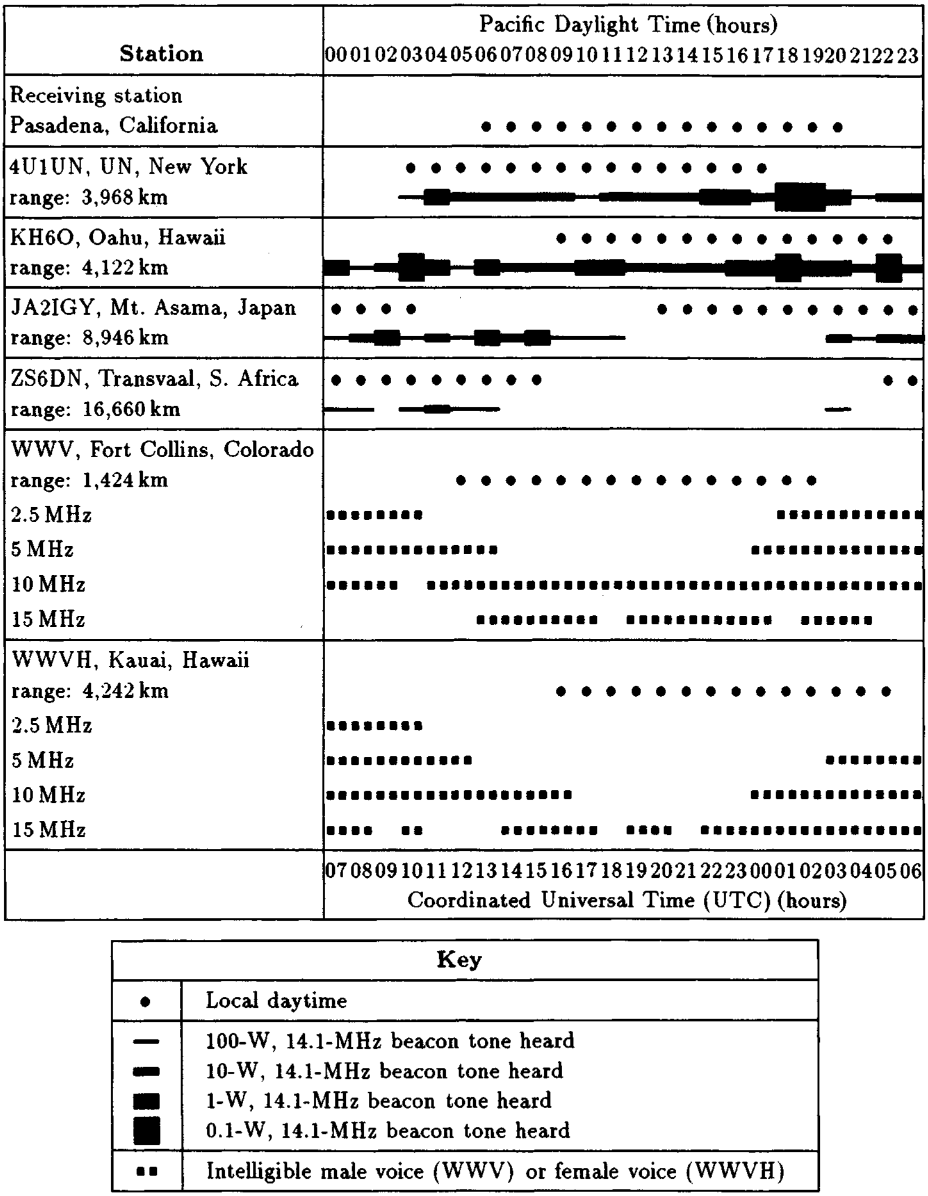


Figure 15.24. Hourly reception reports for June 26, 1993, for Pasadena, California. Courtesy of Kate Rutledge.

1 W, and 0.1 W. By listening to the tones, one can get a feeling for the propagation conditions.

In addition to the beacon stations, there are *time* stations that broadcast the time on particular frequencies. In the United States, the time stations are WWV in Fort Collins, Colorado, and WWVH in Kauai, Hawaii. They use AM. WWV transmits continuously on 2.5 MHz (2.5 kW), 5 MHz (10 kW), 10 MHz (10 kW), 15 MHz (10 kW), and 20 MHz (2.5 kW). WWVH transmits continuously on 2.5 MHz (5 kW), 5 MHz (10 kW), 10 MHz

(10 kW), and 15 MHz (10 kW). Every minute, the Coordinated Universal Time (UTC) is given. This is the same as Greenwich Mean Time, or you may hear it called "Zulu" in voice communications, or "Z" in Morse code. About 15 seconds before the minute, WWVH transmits a woman's voice beginning "At the tone, ..." About 7 seconds before the minute, WWV transmits a man's voice with the same message. This lets us know which stations are being received. The time stations also give ionospheric forecasts and navigation-system status reports. Figure 15.24 also gives reception reports for WWV and WWVH.

Discuss the signal reports in Figure 15.24. You should consider the range and whether the path was in daylight or darkness. Compare sunrise and sunset, and high and low frequencies.

PROBLEM 39 – LISTENING

Your transceivers are for communication by Morse Code. Samuel Morse was the inventor of the telegraph. Morse put a great deal of thought into developing a code for telegraph messages. The first code was based on a code dictionary, with numbers representing words. This is quite slow to encode and decode. Then Morse and his assistant, Alfred Vail, hit upon the idea of using different length pulses and spaces to represent letters. Initially, inking machines with electromagnets were constructed to show these patterns on paper, but operators found that with practice they could interpret the clicking sounds of the instruments by ear, and that they could copy down the message directly. After this, devices called *sounders* were made that were designed for reception by ear. Sometimes you hear sounders in Western movies. They give a distinctive *tick* and *tock* sound at the beginning and end of the pulses. A large network of telegraph lines was built around the world to send messages rapidly, and this was the dominant form of long-distance electrical communication for 100 years.

Morse and Vail incorporated several important ideas in their code (Figure 15.25). To make the sending as efficient as possible, they used shorter patterns for the common letters, like E, T, A, O, I, and N. To find out which letters were the most common, a visit was made to a newspaper printer and the letters in the type cases were counted. In this way, they made the code about 25% faster than it would be if the letters were assigned patterns randomly. Several lengths of pulses are used. You can see these if you compare the pulse lengths for E, T, L, and 0. In addition, they used two different lengths of spaces between the pulses that you can see by comparing I and O. Yet longer spaces separated letters and words.

In Europe, an Austrian, Frederick Gerke, developed a variation of the Morse code that was adopted there (Figure 15.26). Many of the letters are the same in both codes. However, Gerke simplified the code by using only one space length and only two pulse lengths ("dits" and "dahs"). Gerke's code was easier to learn than Morse's, but it was somewhat slower. One interesting character is the letter "O," which is three dahs in Gerke's code. This is much longer than Morse's O, which is a pair of dits. Morse made O short because O is the fourth most common letter in English. However, in German, O is an uncommon letter, ranking only 16th, and for this reason it was made long.

Both the American code and the European code were used when most messages were sent by land telegraph lines. However, when people began sending messages by radio,

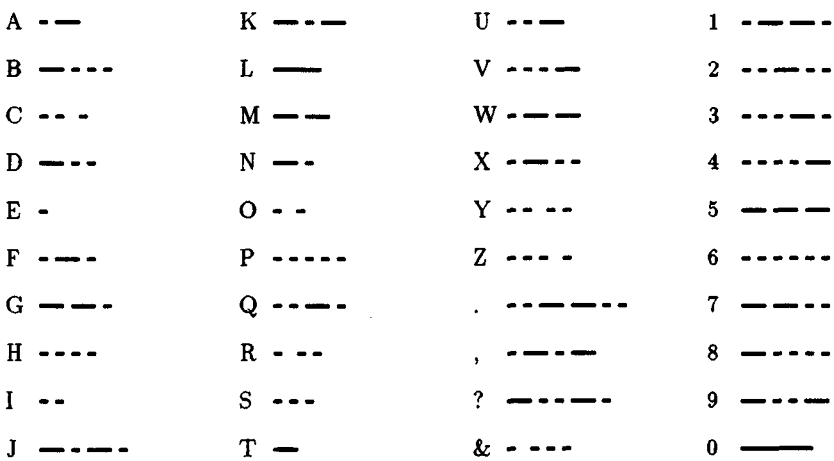


Figure 15.25. The American Morse Code developed by Samuel Morse and Alfred Vail. The length of the lines represents the length of the pulses, and the gap between the lines represents the spacing between the pulses. This code is not in common use today, having been replaced by the International Morse Code.

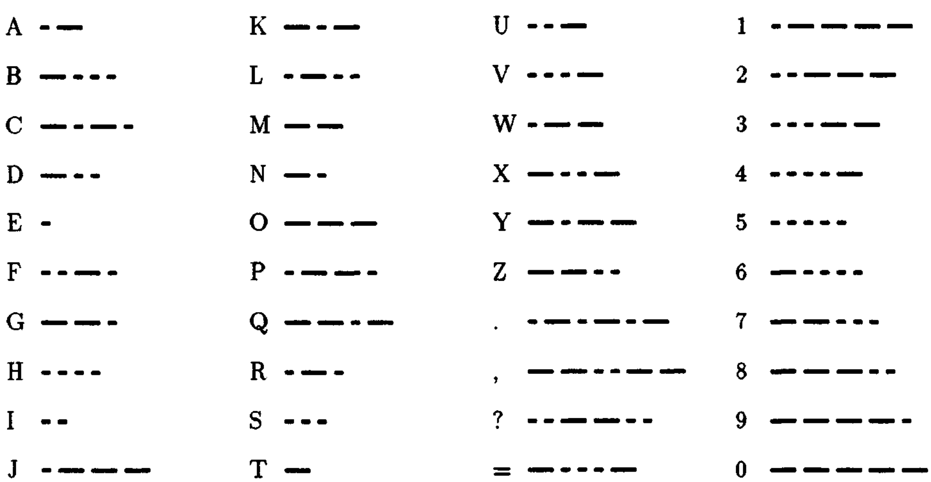


Figure 15.26. The International Morse Code developed by Gerke.

it became important to have a single code, and Gerke's simpler code won out. Known today as the International Morse code, it is universal for languages that use Roman letters. Similar codes have been developed for other writing systems, and the Japanese code is often heard on the air.

Today, Morse code is rarely used in commercial or military communications, where codes intended for automatic decoding are used instead. However, Morse's idea of representing common characters by shorter patterns is used in data compression algorithms, and the idea of conveying information by varying pulse and space lengths is very much alive in radio control systems. In some ways the code is an obsolete form of communications, but in other ways it is no more obsolete than sailing or horseback riding are for transportation. It requires lower power levels, less bandwidth, and simpler equipment than voice communications, and there are a large number of standard abbreviations that allow people who do not share a common language to communicate.

Use your radio to take down a Morse-code two-way conversation off the air. Amateurs call this a QSO. Include an interpretation. For each station, you should try to get the call sign, name, location, signal report, output power, and antenna type. You can do this with a decoding machine tuned to 600 Hz. Plug the audio output from your transceiver into the jack of the decoder. You will probably want to connect your speaker in parallel, because the speakers in decoders are often quite poor. Adjusting the radio takes some care. If the RF gain is set too high, atmospheric noise will cause false letters on a decoder. Usually these will appear as Es. Set the RF gain pot as high as you can without triggering Es. Now look for a signal. Tune carefully through the band with the VFO Tune pot, looking for strong signals. When you find a station, set the frequency precisely by adjusting the RIT pot to make the tune light on the decoder as bright as you can. You should start to see the copy appear. It will be difficult to receive if the signal is weak or if the sending is sloppy. Machines are not as good as people are at copying Morse code! Sometimes you will find a station that is using a radio teletype code called the Baudot code. The Baudot code uses FSK, with two tones 170-Hz apart, and you will hear a continuous warbling. Most decoders also have a setting for the Baudot code. You should try to tune on each of the tones, and see if one gives good copy.

Each station is identified by a call sign that is assigned by the communications authority in that country. An example would be KE6AGH. The first letter "K" identifies the station as American. The number "6" identifies the call district within a country, California in this case. Figure 15.27 shows the call districts in the United States, and Table 15.1 gives the call-sign prefixes for many countries.

Watch carefully to see when one operator quits sending and the other begins. You will usually need to adjust the RIT slightly, unless the two operators are on exactly the

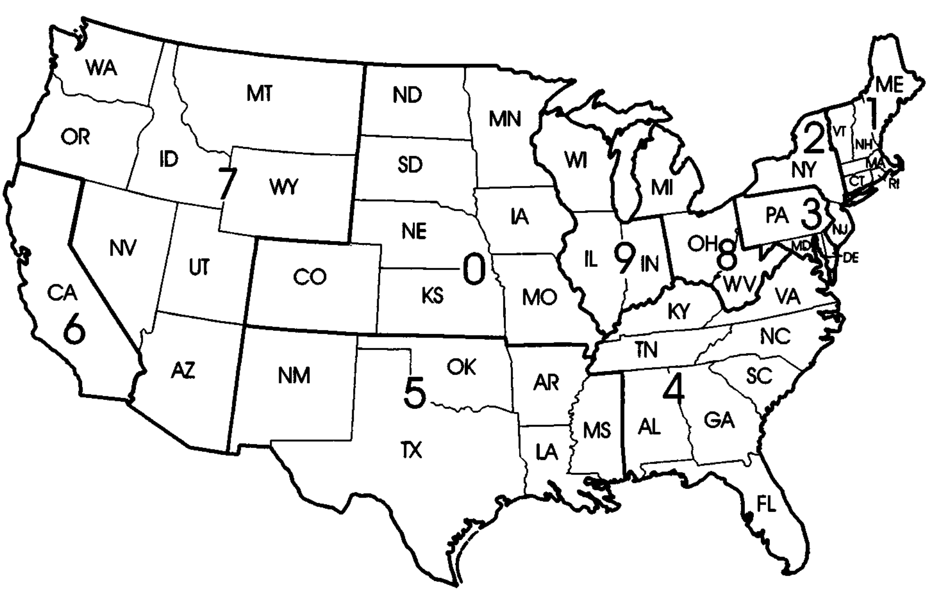


Figure 15.27. United States call districts.

Table 15.1. Some Call-Sign Prefixes and Their Countries.

A, K, N, W United States	HL South Korea	SU Egypt
BV Taiwan	HP Panama	SV Greece
BY China	HR Honduras	TA Turkey
CE Chile	HS Thailand	TG Guatemala
CM–CO Cuba	HT Nicaragua	TI Costa Rica
CP Bolivia	I Italy	U Russia
CT Portugal	JA–JS Japan	VE Canada
CX Uruguay	LA Norway	VK Australia
DA–DL Germany	LU Argentina	VU India
DU Philippines	OA Peru	XE, XF Mexico
EA Spain	OE Austria	YB–YD Indonesia
EI Ireland	OH Finland	YS El Salvador
F France	OK Czech Republic	YV Venezuela
G United Kingdom	OM Slovak Republic	ZL New Zealand
HA Hungary	ON Belgium	ZP Paraguay
HB Switzerland	OZ Denmark	ZS South Africa
HC Ecuador	PA Netherlands	4X Israel
HH Haiti	PP–PY Brazil	7L, 7M Japan
HI Dominican Republic	SM Sweden	9M Malaysia
HK Colombia	SP Poland	9V Singapore

same frequency. There are several ways for an operator to turn over to the other. If an operator has a question, often the question will be sent, followed by BK. This is equivalent to "over" in a radio voice conversation. The other operator will start with BK. It is also common to first indicate the end of a message by sending an A and R run together ($\overline{\text{AR}}$). This sign is a run-together $\overline{\text{FN}}$ (for finished) in American Morse. Then the two call signs would be sent, followed by K, or $\overline{\text{KN}}$. For the last message in a QSO, an operator will send $\overline{\text{SK}}$. This is $\overline{30}$ in American Morse, and it was traditionally sent by telegraphers before a 30-minute lunch break.

Interpreting your copy takes some work, because it is quite different from spoken English. Table 15.2 gives a glossary, and Figure 15.28 gives a typical QSO. These forms are used world wide, and much of it dates back to American land-line telegraphy. For example, the abbreviation for "and" is like the & character in American Morse, and the abbreviation for a laugh is from the American Morse "ho." Many are just abbreviations to save time. Often they help the operator avoid sending an "O," which is quite long. "73" for "best regards" is ubiquitous. It derives from nineteenth-century land-line number codes. "73" has a wonderful sound in International Morse Code, and this may be why it is so common. The articles "a," "an," and "the" appear only rarely, and "is" may be omitted. "=" is typically used instead of a period or comma. You may see a series of them as a pause. It is a good idea to listen for a CQ, which is a call to any station. The CQ is sent before the QSO begins, and if you hear it, you can copy the entire QSO. The first exchange gives a signal report, name, and location. These are typically repeated to help if reception is poor. The first number in the signal report is the *readability*. The readability varies from 1 to 5, and indicates how easy it is to copy the signal. "5" indicates a signal that is easy to copy, and 2 indicates a signal that is quite difficult to copy. The second

Table 15.2. Morse-Code Vocabulary.

ABT about	QRL This frequency is busy. (Asks others to find another frequency)
AGN again	QRL? Is this frequency busy? (A check to avoid interfering with others)
ANT antenna	QRM troubling interference
BK over ("go ahead")	QRN troubling atmospheric noise
BN been	QRP low power (5 watts or less)
BURO bureau (an agency for exchanging QSL cards between countries)	QRS slow down
C degrees Celsius	QRT quit
CONDX ionospheric conditions	QRU I have nothing more to say.
CQ call to any station	QSB troubling fading
CUAGN see you again	QSL confirm the contact by a card
CUD could	QSO a contact
CW Morse code	QST a broadcast to amateurs
DE KD6PFK This is station KD6PFK.	QTH PASADENA, CA gives location
DP dipole	R are, or received, OK ("roger")
DR dear (before a name, as "DR Kate")	RIG transceiver
DX foreign contacts	RPT repeat, or report
E E E error	RST 599 a signal report
EL, ELE antenna element	SIG signal
ES and	SKED a "scheduled" appointment
FB good ("fine business")	SUM some
FER for	SRI sorry
FM from	T zero (much shorter than "0")
GA good afternoon	TEMP temperature
GB good bye	TKS, TNX thanks
GE good evening	TT that
GL good luck	TU thank you
GM good morning	U you
GN good night	UR your
GUD good	VEE a common antenna
HI, HEE telegraphic laugh	VFB very FB
HPE hope	VY very
HR here	W watt or watts
HV have	WID with
HW? How do you copy?	WL well, or will
INFO information	WUD would
K over ("go ahead")	WX weather
M meters	XYL wife ("ex YL")
/M from a vehicle ("mobile")	YL woman ("young lady," often sent after a name, to avoid being called an OM)
N nine (much shorter than "9")	YRS years
NW now	73 best regards
OM form of address for men ("old man")	88 best regards (to a woman)
OP name ("operator")	
PSE please	
PWR transmitter power	

```
CQ CQ CQ DE KN6EK KN6EK KN6EK K

KN6EK DE JA0FCC +

JA0FCC DE KN6EK GE OM ES PLEASED TO MEET U -
UR RST 599 5NN - QTH PASADENA, CA PASADENA, CA -
NAME DAVE DAVE - HW? + JA0FCC DE KN6EK !

RR KN6EK DE JA0FCC GM DAVE SAN -
TNX FB RPT FM PASADENA, CA - UR RST 449 44N -
OP GAKU GAKU - QTH NAGANO NAGANO - HW? + KN6EK DE JA0FCC !

RR JA0FCC DE KN6EK ES TNX RPT FM NAGANO -
RIG HR IS HOMEMADE - RUNNING 2W - ANT 2 EL YAGI AT 25M -
WX CLOUDY ES TEMP 15C - HW? + JA0FCC DE KN6EK !

RR KN6EK DE JA0FCC ALL COPY DAVE SAN -
RIG HR KENWOOD 850 - PWR 100W - ANT HORIZONTAL DP AT 10M -
WX HR RAIN ES TEMP 10C - HW? + KN6EK DE JA0FCC !

RR JA0FCC DE KN6EK TNX INFO GAKU OM -
AGE 46 YRS - HV 3 CHILDREN - JOB TEACHER -
WHAT IS UR JOB GAKU SAN? BK

BK BUDDHIST PRIEST - I LIVE IN A TEMPLE - QSL VIA BURO ? BK

BK MY QSL VIA BURO SURE - TNX FB QSO GAKU SAN -
HPE CUAGN 73 # JA0FCC DE KN6EK !

RR KN6EK DE JA0FCC - TNX FER 1ST QSO DAVE SAN -
73 73 # KN6EK DE JA0FCC E S E

TU E E
```

Figure 15.28. Typical Morse code on the air. Work through the meaning of this QSO before you try to copy one. In the text, Microcraft decoder conventions are used: – for =, + for $\overline{\text{AR}}$, ! for $\overline{\text{KN}}$, and # for $\overline{\text{SK}}$.

number is the *strength*. The strength varies from 1 to 9, with 9 indicating a very strong signal, and 1 indicating a very weak signal. The third number indicates the tone quality of the signal. Nowadays, this is almost always a nine. The endings "E S E" and "E E" are traditional.