Chapter 6
Wireless

The last medium we shall cover is the "ether" — which supposedly permits wireless transmission via radio waves.

A century or two ago, when physicists first started intently to study light, they were convinced that, just as sound waves need air or some other medium to go through, so light (and eventually radio waves) needed some medium to pass through. They called this medium *ether*, and spent years and years trying to prove its existence in the laboratory. Not until after 1873, when James Clerk Maxwell developed his famous *Maxwell's Equations*, did physicists finally abandon the search for this "Lost Medium."

We now know that not just radio waves, but light, X-rays, gamma rays, and cosmic rays are all *electromagnetic* waves that can travel through empty space. In this chapter we will confine ourselves to radio waves, but this is somewhat of a misnomer since what we call radio waves are also used for TV, digital data, radar, and other applications which we can lump under the generic name *wireless.*

Since all of these wireless applications require an antenna to launch the wave into space, that is where we will begin. Although we'll look primarily at transmitting antennas, it's important to note that a good transmitting antenna will also generally work well for receiving (though not always the other way around.)

The dipole antenna

In Chapter 4, we learned that a ¼-wavelength section of transmission line that is open at the far end looks like a short at its input. Let's now take such a section of balanced line, and connect it to an RF signal generator, as at Fig. 6-1*a*. Since the generator sees a short circuit, it is not feeding any power into the cable. That is, there is current, but no voltage; since the power is the product of current times voltage, the power leaving the generator must be zero. That makes sense, since there is no place for any power to go.

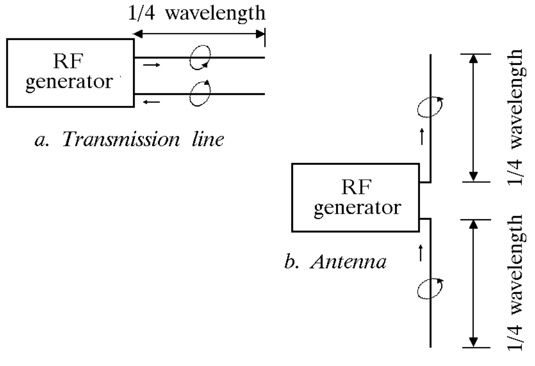

Fig. 6-1. A transmission line vs. an antenna

Now let's separate the two wires of the balanced line, bending them completely out so they point in opposite directions, as in Fig. 6-1*b*. If we now take some measurements, we will see that there is not only current flowing out of the generator, but there is also a voltage across its output. (And, for the purists among our readers, the current and voltage are in phase.) There is thus power leaving the generator and going into the wires. So where is it going?

What is happening is that, by opening up and separating the two wires, we have changed the transmission line into an antenna. The power coming out of the generator is being transmitted out into space. If we actually measured the impedance looking into the antenna of Fig. 6-1*b*, we would measure a resistance of about 73 ohms, rather than a short circuit. If the generator outputs the voltage V, then the power going into the antenna is $P = V^2/R = V^2/73$.

Look back at Fig. 6-1 again. The small straight arrows along the wires show the current direction at some particular instant of time. In each case, the current on the top wire comes out of the generator, while the current in the bottom wire goes back into the generator. In the transmission line, the two currents go in opposite directions, while in the antenna, the two currents both go in the same direction — up.

These two currents both cause magnetic fields to go around the wires; these go in a circle around each wire, and are shown in Fig. 6-1 as small circles, with tiny arrows indicating the direction of the magnetic field.

In the transmission line, Fig. 6-1*a*, the two currents go in opposite directions, and so the two magnetic fields also go in opposite directions. Moreover, the two currents are equal, so the two magnetic fields are also equal. So they cancel. That is, if you were to stand a few feet away from the wires, you couldn't measure any magnetic field because the two fields cancel.

In the antenna, however, the currents in the two wires always go in the same direction. Hence the magnetic fields also go in the same direction. So, if you were to stand back a few feet, you would be able to detect the magnetic fields (if you have sensitive enough equipment) because they add, rather than cancel.

In addition to generating a magnetic field, the antenna wires also generate an electric field.

DETOUR

The concept of a *field* is hard to explain, yet important for us to understand. In simple terms, a field is a system of forces filling some space, which can cause something to move. An analogy is the wind in a storm. At any particular spot, the wind has a certain strength and a direction. If you let a balloon loose at that spot, the wind will move it. How fast it moves depends on the strength of the wind (field) at that spot, and the direction it moves depends on the direction of the wind (field) at that spot. Moreover, as it moves from one place to another, the balloon may change its motion because the field at the new location may have a different strength and direction.

The current through the wires causes a magnetic field which, at any particular spot in space, has a certain strength and a certain direction. Fig. 6-1 doesn't show the strength, but the arrows on the circles show the direction. If you place a compass in the field, the arrow shows which way the compass would point.

We can also generate an electric field, though in a different way. Suppose we have two metal balls, as in Fig. 6-2, and place a lot of negative charges (electrons) on the top ball, and a lot of positive charges (protons) on the bottom ball, as shown. If you now place a single electron at the spot labelled A, that electron will move (assuming there's nothing in its way). The famous rule in electricity is that "unlike charges attract, while like charges repel each other." The protons on the bottom ball will attract the lone electron, while the electrons on the top ball will repel it, so the electron will move down, in the direction of the arrow. So the charges on the two metal balls generate a field which can cause that electron to move.

In the same way, electron B will move down and to the left, also in the direction of the arrow, and electron C will also move in the direction shown by the arrow. The thin curved lines going through electrons B and C show the direction of that field — the path that the electron would take. We could draw an entire series of lines in Fig. 6-2, which would show the path an electron would take if it was dropped anywhere in the space around the two balls. These lines would then specify the direction of the field.

END OF DETOUR

So let's take a look at the antenna shown in Fig. 6-3. Imagine that the small arrows along the wires indicate the flow of electrons along the wires. Electrons flow up on the top wire, eventually generating an excess of electrons at its tip. At the same time, electrons flowing upward on the lower wire leave a lack of electrons — or an extra number of protons — at the end of the bottom wire. So we now have a negative charge at the top, and a positive charge at the bottom. This generates an electric field, just as the two charged balls generated in Fig. 6-2.

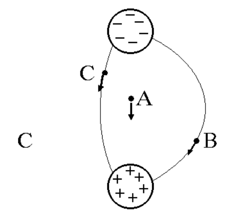

Fig. 6-2. An electric field

At the same time, the currents in the two wires generate the magnetic field, which circles the wires as shown by the arrows on the circles around the wires.

Remember, though, that the voltage and current are constantly changing. The generator is sending out an ac sine wave. So both the magnetic and electric fields are constantly Increasing and decreasing, and even changing direction as the voltage changes from positive to negative and vice versa.

The next step in our explanation requires a leap of faith; it can't be justified without a *lot* of math and physics. Maxwell's Equations, which we mentioned at the beginning of this chapter, show that the magnetic and electric fields interact. What essentially happens is that the buildup and collapse of the electric field pushes the magnetic field away from the antenna; in exactly the same way, the buildup and collapse of the magnetic field pushes the electric field outward. The result is an *electromagnetic* field which radiates outward from the antenna into space, like a radio wave. This electromagnetic wave explains not just radio and TV waves, but even light, which is just another kind of electromagnetic wave, though of much higher frequency (and shorter wavelength) than even microwaves.

In short, the two ¼-wavelength wires in Fig. 6-3 make up the simplest kind of antenna, called a *dipole*. Most dipoles, however, rather than consisting of vertical wires, are horizontal. Note that the currents in the two

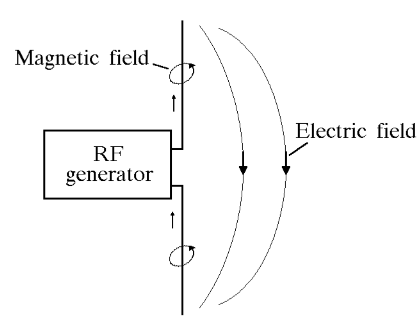

Fig. 6-3. Fields around a dipole

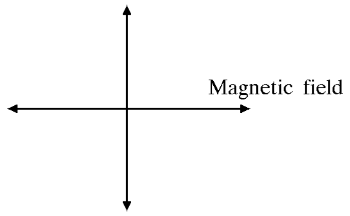

Fig. 6-4. Fields with vertical polarization

halves of the dipole need to be equal but opposites. The dipole therefore needs a balanced transmission line; it cannot be properly fed by a coax cable.

Antenna Polarization

If you stand back from the vertical dipole antenna of Fig. 6-3 and watch the electrical and magnetic fields go through space past you, because the dipole antenna goes up-down, you will note that the electric field is always up-down, while the magnetic field goes left-right, as shown in Fig. 6-4. This is called *vertical polarization*, and the antenna is *vertically polarized*. If the antenna were horizontal, then the electric field would be horizontal, the magnetic field vertical, and we would say that the entire signal is horizontally polarized.

The best way to receive a vertically-polarized signal would be with another vertically-polarized antenna. The electric field will make electrons in the antenna move up and down (causing an electric current), and the magnetic field changing around the receiving antenna will also make electrons move up and down (also causing an electric current). If the receive antenna were horizontally polarized, then the current flow in the wire would be across the diameter of the wire, not along it, and the receiver would receive very little signal. So we need to pay some attention to the polarization of antennas. For example, most AM broadcast stations use tall towers as the antenna; these are vertical, and so a vertical antenna (like a car antenna) works well. TV stations, on the other hand, always use horizontal transmitting antennas, so you need a horizontally-polarized antenna to receive them.

Satellite communications are a problem, since satellites sometimes rotate in space, and signal polarization is also affected by travelling through the earth's atmosphere. So satellites often use *circular polarization*. This involves two antennas, one horizontal, the other vertical, positioned ¼ wavelength behind each other, and both transmitting the same signal. The signal from the closer antenna arrives first with one kind of polarization; the signal from the other arrives ¼ of a cycle later, but is polarized at 90 degrees from the other. So it looks as though it has turned 90 degrees. Depending on how you position the antennas, you get either right-hand or left-hand rotation.

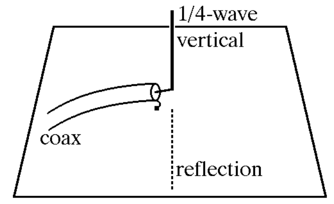

Fig. 6-5. The 1/4-wave vertical

¼-wavelength vertical antenna

A vertical antenna, such as the whip you might see on a police car, is basically half of a dipole, as shown in Fig. 6-5. Suppose you were to take just one ¼-wavelength wire, but position it vertically above a mirror, as shown. If you then looked at the wire's reflection in the mirror, you would think that there are in fact two wires — a dipole. The "mirror" under the antenna is called a *ground plane*, and it makes the vertical antenna look like a dipole.

Because radio waves reflect from a conducting surface, any large sheet of metal will act as a ground plane. So mounting a ¼-wave whip on a metal surface, such as a car roof, works well. If a vertical antenna is mounted on the ground, it's necessary to make the ground under it more conductive so it acts as a better mirror. Commercial AM broadcast stations often float their antenna towers on a raft in the middle of a swamp, since the water in the swamp makes it a good conductor. (Incidentally, the tower doesn't actually touch the ground under it. The base of the tower sits on an insulator.) Amateur radio operators, on the other hand, often bury wires (called *radials*) in the ground around the vertical antenna. They are called radials because they spread outward from the antenna, like the radius of a circle. Radials are also needed when mounting a whip antenna on the body of a car with a fiberglass body; hams often use conductive aluminum tape for that purpose.

In order to "see" a full reflection of the wire, the ground plane has to extend far enough out from the vertical antenna. It should extend at least ¼ wavelength out in each direction, although more is better. Even so, imagine that the ground plane is, say, the length of the ¼-wave whip. If you look at it from above, you see the reflection of the whip in the ground plane, but if you look from a shallow angle, you only see a very small part of the whip in the “mirror”. Hence a vertical tends to transmit slightly upward, rather than horizontally along the ground.

Since there is no bottom wire to the antenna, there is no place to connect the second wire in a balanced transmission line. Hence vertical antennas are usually fed with coax cable. The inner conductor goes to the vertical, while the outer shield connects to the ground plane just under or next to the whip. Since there is only half of a dipole, the impedance of the whip is ½ of the dipole’s impedance, or about 37 ohms.

Carriers and Modulation

The two antennas we have so far discussed are ½ and ¼ wavelength long, respectively. There are other kinds of antennas; some are bigger, and others are smaller. But a general rule of thumb is that an efficient antenna cannot be much smaller than these two. This gives us a rough idea of how large antennas have to be to work well.

So let’s say that we want to transmit an audio signal. Since audio covers the frequency range from 20 to 20,000 Hz, using the formula for the wavelength

$$\lambda = \frac{\text{velocity}}{\text{frequency}}$$

gives us wavelengths from 9300 miles at 20 Hz, down to 9.3 miles at 20,000 Hz. A one quarter wavelength antenna then becomes somewhere between 2325 miles and 2.325 miles long. This is simply not practical, partly because the lengths are just too long, but also partly because any particular length will only work for some frequencies — there is no compromise length which will be good for the entire frequency range.

To make a practical antenna, we have to shorten it to some more reasonable length, and that requires that we shorten the wavelength. Looking at the above equation, we see that there are only two ways to do that — either reduce the speed of light (not exactly feasible), or increase the frequency.

We could all learn to talk like Donald Duck or the Chipmunks, but this would not be enough. To reduce the antenna length a lot, we must increase the frequency a *lot*.

The solution is simple: instead of sending the voice or music by itself, send instead a much higher frequency signal called the *carrier*, and let the voice, music, picture, or whatever, ride on top of that carrier. The process of putting the desired information (voice or whatever) on the carrier is called *modulation*, and will be discussed in the next two chapters.

For example, if you look at the dial of an ordinary AM radio, you will see numbers ranging from 540 up to 1600. These numbers represent the frequencies of the carriers for the AM broadcast stations, which (for this type of radio) range from 540 kHz up to 1600 kHz.

Consider, for example, a radio station at 880 kHz. A wavelength is then

$$\lambda = \frac{\text{186,000 miles per second}}{\text{880 kHz}} = \text{0.21 mile}$$

which is about 1116 feet. A quarter-wavelength antenna would therefore have to be 1116/4 or about 279 feet long — a quite reasonable thing to do.

TABLE 6-1. THE ELECTROMAGNETIC SPECTRUM

Type of radiation	Subdivided into	Frequencies (approx)	Typical uses
Radio frequencies	VLF — Very Low Frequenccies	10 to 30 kHz	Special purposes
"	LF — Low Frequencies	30 to 300 kHz	Foreign broadcasting
"	MF — Medium Frequencies	300 kHz to 3 MHz	AM broadcasting
"	HF — High Frequencies	3 to 30 MHz	International broadcasting
"	VHF — Very High Frequencies	30 to 300 MHz	TV, FM broadcast, mobile
"	UHF — Ultra High Frequencies	300 MHz to 3 GHz	TV, Cell phones
"	SHF — Superhigh Frequencies	3 to 30 GHz	Microwaves, satellites
"	EHF — Extra High Frequencies	30 to 300 GHz	Not yet in wide use
Light	IR — Infrared	3×10^{11} to 4.3×10^{14} Hz	TV remotes, night viewing
"	Visible	4.3×10^{14} to 1×10^{15} Hz	Seeing
"	UV — Ultraviolet	1×10^{15} to 6×10^{16} Hz	
X-Rays		6×10^{16} to 3×10^{19} Hz	Radiology
Gamma Rays		3×10^{19} to 5×10^{20} Hz	Not man made
Cosmic Rays		5×10^{20} to 8×10^{21} Hz	Not man made

TABLE 6-2. EXAMPLES OF CARRIER FREQUENCIES	
Frequency Range	Typical Uses
10 kHz – 100 kHz	Research; submarine communications
100 kHz – 550 kHz	Foreign Broadcast
550 kHz – 1700 kHz	AM Broadcast
1.7 – 30 MHz	Shortwave broadcast and other communications
30 – 54 MHz	Mobile and amateur radio
54 – 88 MHz	TV Channels 2 – 6
88 – 108 MHz	FM Broadcast
108 – 136 MHz	Aircraft
136 – 174 MHz	Mobile and amateur radio
174 – 216 MHz	TV Channels 7 – 13
216 – 450 MHz	Government and amateur
450 – 470 MHz	Mobile applications
470 – 806 MHz	TV Channels 14 – 69
806 – 905 MHz	Cellphones and mobile
905 and up	Experimental, industrial applications, microwaves, satellite communications, cellphones, etc. etc.

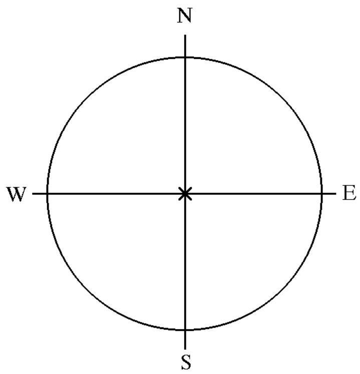

Fig. 6-6. Radiation pattern around a vertical dipole

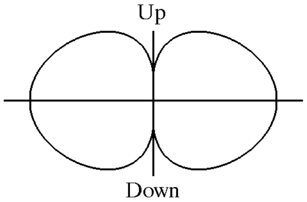

Fig. 6-7. Dipole vertical radiation pattern

Each different radio station (as well as TV stations, radar, navigation transmitters, etc.) uses a different carrier frequency. These are generally assigned by the Federal Communications Commission (or its equivalent in other countries) so as to avoid interference between different stations. There are lot more transmitters out there than the typical radio or TV stations most people are familiar with, and typical carrier frequencies range from a low of 10 kHz up to the gigaHertz range. Table 6-1 shows where radio waves fit into the overall electromagnetic wave spectrum, while Table 6-2 gives a brief summary of the carrier frequencies used for different communications purposes.

Radiation Patterns

The term “radiation pattern” describes the directionality of an antenna. An antenna which transmits (or receives) equally in all directions is called an *isotropic antenna*. But there is no such thing — it is impossible to build one. Instead, every real antenna transmits better in some directions, and worse in others.

Consider, for example, a plain vertical dipole as shown in Fig. 6-3. You might see such an antenna, for instance, hung from a weather balloon, with the transmitter actually hanging in the middle of the antenna. Such an antenna would transmit equally well in all horizontal directions — north, south, east, and west. If you imagine that we’re flying above the antenna, looking down at it, we would draw the horizontal radiation pattern as in Fig. 6-6. The small X in the middle signifies the position of the antenna, and the circle around it shows the directions in which it transmits. Because the radius of the circle is the same in all directions, the signal strength is also the same in all directions. We say that such an antenna is *omnidirectional.* (On the other hand, an ellipse which was longer north-south than east-west would mean that there is more signal going north and south than east and west.)

If, on the other hand, you look at the antenna from the side, you would note that it transmits fairly well horizontally, but not at all up or down. That is because the dipole doesn’t send (or receive) any signal off the ends of the wire. In this case, the vertical radiation pattern (as seen from the side) would look like the figure-8 pattern in Fig. 6-7. Since the pattern is actually three-dimensional, it looks more like a donut; what we see in Fig. 6-7 is just a cut through the donut.

A ¼-wavelength vertical antenna, has the same omnidirectional horizontal radiation pattern as Fig. 6-6 since it transmits equally well in all horizontal directions, but its vertical radiation pattern is more like Fig. 6-8, since it transmits slightly upward rather than straight to the sides. (Actually, the horizontal radiation pattern is a circle only if the ground plane extends equally far out in all directions. Car-mounted vertical

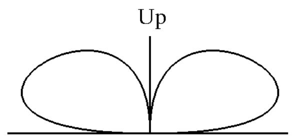

Fig. 6-8. 1/4-wave vertical radiation pattern

antennas usually don't satisfy that requirement, and so they often transmit and receive better in some directions than others.

Directional antennas

Although a single vertical antenna has an omnidirectional horizontal radiation pattern, it is possible to change that by using two or more vertical antennas. You may have noticed that many AM broadcast stations use more than one tower. This allows them to tailor their radiation pattern to their coverage area. If the station is on one end of a city, it may want to direct more of its power toward that city, and less to other, less-populated areas. In addition, some radio stations must reduce their transmitted power in some directions so as not to interfere with stations farther away.

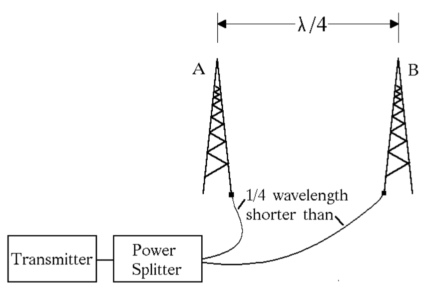

Fig. 6-9. Typical AM broadcast station towers

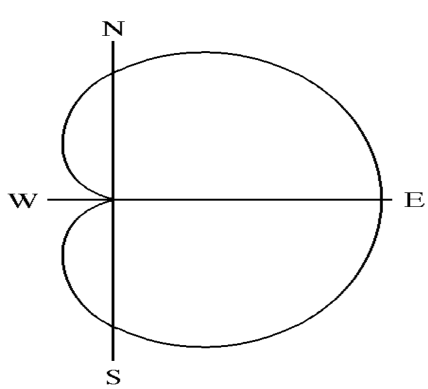

Fig. 6-10. Cardioid pattern from the antennas in Fig. 6-9

Fig. 6-9 shows an example. Suppose a station has two towers A and B, separated by ¼ wavelength from each other, fed by two different lengths of coax cable, so that the cable to tower A is ¼ wavelength shorter than the one to tower B.

Now imagine that you are at the far right of the drawing, east of the antennas. The signal from Tower B gets to you ¼ cycle sooner than the one from tower A (since it is closer to you), but since it had a longer cable, it was transmitted ¼ cycle later. So the transmitted signals from the two antennas arrive at your location at the same time — in phase — and they add.

On the other hand, suppose you are at the far *left* of the drawing, west of the antennas. The signal from tower B gets to you ¼ cycle later (since it is farther away.) It was also transmitted another ¼ cycle later (because of the longer cable), so it arrives at your location ½ cycle later than the signal from tower A. The two signals are therefore out of phase, and they cancel. So the station transmits well to the east, but not to the west. The resulting radiation pattern in Fig. 6-10 is called a *cardioid* because it resembles a heart shape.

DETOUR

The program at the top of the next page, written in plain IBM Basic, lets you plot the radiation patterns of various combinations of vertical antennas. To run it, you must first specify how many vertical towers or whips there are. Then for each one, you must specify where it is in relation to the transmitter (compass direction in degrees, and distance from the transmitter in wavelengths), the length of the feed cable from the transmitter (in wavelengths), and the attenuation in the cable (including any attenuation added on purpose. For example, one of the towers may purposely be getting reduced power to change the pattern.)

For example, Fig. 6-11 shows the radiation pattern for a two-tower setup, where one tower is .25 wavelength north (0 degrees) from the transmitter, and the other is .25 wavelength south (180 degrees). Both are fed with a cable .25 wavelength long, and both get equal

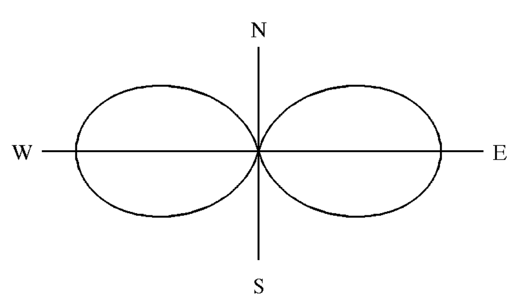

Fig. 6-11. A figure-8 pattern from two towers


```
10 ' PROGRAM TO GRAPH ANTENNA RADIATION PATTERNS
20 INPUT "number of towers = "; NR
30 FOR I=1 TO NR
40    PRINT "For tower"; I; "enter:"
50    INPUT "  Angle (0 degrees = north)"; ANGLE
60    INPUT "  Dist. (in wavelengths)"; DIST
70    INPUT "  Cable length (wavelengths)"; TL(I)
80    INPUT "  DB attenuation"; DB
90    GAIN(I)=10^(DB/20)
100   Y(I)=DIST * COS(ANGLE/57.2957795)
110   X(I)=DIST * SIN(ANGLE/57.2957795)
120 NEXT I
130 ' SET UP SCREEN
140 KEY OFF
150 SCREEN 2
160 CLS
170 'calculate and plot
180 LINE (0,100)-(639,100)
190 LINE (320,0)-(320,199)
200 FOR D=0 TO 360
210    XSUM=0 : YSUM=0
220    X=1000 * COS(D/57.2957795)
230    Y=1000 * SIN(D/57.2957795)
240    FOR R=1 TO NR
250       DIST=SQR((X-X(R))^2 + (Y-Y(R))^2)
260       PHASE=DIST + TL(R)
270       MAG=GAIN(R)*3000/SQR(DIST)/NR
280       XSUM=XSUM+MAG * COS(PHASE*2*3.14159)
290       YSUM=YSUM+MAG * SIN(PHASE*2*3.14159)
300    NEXT R
310    TOTAL=SQR(XSUM^2+YSUM^2)
320    X=320+2.4*(TOTAL*COS(D/57.2957795))
330    Y=100-TOTAL*SIN(D/57.2957795)
340    IF D=0 THEN LINE -(X,Y),,,0
350    IF D<>0 THEN LINE -(X,Y)
360 NEXT D
370 IF INKEY$="" THEN 370 ELSE SCREEN 0:STOP
```

power (assuming 0 db loss.) You see that we get a figure 8 pattern. Listeners to the east and west get the same signal from both towers, so the two signals add. Listeners north and south, however, get the two signals ½ cycle apart (because one tower is ½ wavelength closer than the other) and so the two signals cancel. You can experiment with different combinations in the program.

END OF DETOUR

The above examples were all based on AM broadcast station towers; in that case, the towers are actually fed from the transmitter through some sort of a matching network. It is possible, however, to build a directional antenna without actually supplying power to the other components. Figure 6-12 shows how.

Fig. 6-12 is a *beam* antenna, also often called a *Yagi* antenna. In the center we have a boom, which holds a number of elements. One of these, connected to the transmitter by the feed line, is called the *driven element* because it is actually driven by the transmitter signal. It essentially acts as a dipole.

Mounted parallel to the driven element are one or more *parasitic elements*, which are not directly connected to the transmitter. There is usually one *reflector*, and one or more *directors*. Each one of these parasitic elements acts as a separate little antenna.

A normal dipole consists of two ¼-wavelength pieces, connected to a transmission line. Suppose, however, that the transmission line was missing, and instead the two sections of the dipole were shorted together in the middle. Any signal received by the dipole would go to the middle, hit the short, be totally reflected back into the dipole, and then be retransmitted into the air. This is precisely what the reflector and directors do — since they are so close to the driven element, they pick up a small amount of signal, don't know what to do with it, and so they send it right back out again. The signals transmitted by the driven element, and retransmitted by the parasitic elements, then add or subtract to make the overall antenna directional.

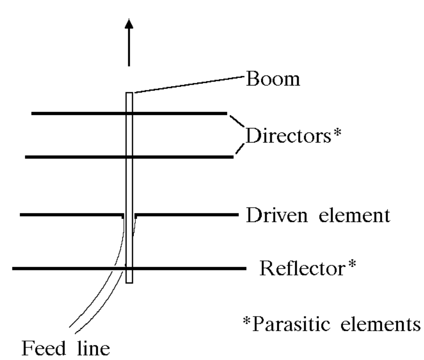

Fig. 6-12. A 4-element Yagi beam

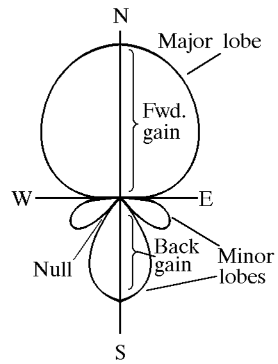

Fig. 6-13. Radiation pattern of a yagi

Looking at Fig. 6-12, you'll note that the reflector is slightly longer than the driven element, while the directors are slightly shorter. The different lengths change the phase angles of the reflected signals to make sure that the signal is sent in the correct direction. The reflector reflects the signal, whereas the directors act as a lens to direct the signal forward. The arrow in Fig. 6-12 is at the front of the antenna — it shows the direction where most of the signal goes (the other end is obviously called the back).

Fig. 6-13 shows the radiation pattern of a typical yagi; we assume that the antenna is aimed north in this case. We see four lobes; one *major lobe* which shows that most of the power goes north, and three *minor lobes* which show other directions where some of the other power goes. Between these lobes are directions which get no power; these are called *nulls*.

The length of a lobe represents the gain of the antenna in that direction. We will define the gain of an antenna shortly; for now, let's just note that the gain of this antenna in the forward or front direction (toward the directors) is higher than the gain in the back direction (toward the reflector). The ratio of these two gains is called the *front-to-back ratio* and it is usually expressed in decibels.

Other things to consider

We've so far discussed only a few of the more basic antenna types — the dipole, ¼-wave vertical, and the Yagi beam. There are many other kinds, and an exhaustive description of them all would take us many more pages. Let's instead look at some more general concepts.

Types of feed line

When an antenna consists of two identical parts, such as the two halves of a dipole or the driven element in a Yagi, it can be fed by a balanced line. The two sides of the antenna get equal, but opposite voltages. For low-power applications, 300-ohm twin-lead could be used, but for higher powers, or if line losses are important, an open-wire line is more common. This consists of two conductors, kept apart by insulated spacers every few inches. These spacers have less loss than the continuous strip of plastic used in the twin-lead.

But when the antenna consists of unlike parts, such as a vertical antenna and its ground plane, you should use an unbalanced line, such as a coax cable. You can mix and match by using a balun to match a balanced load to an unbalanced line, or vice versa. With a transmitting antenna, however, you must be sure that the balun can handle the power. The balun can be a transformer, as discussed in our transmission line chapter, or it can be made from coax cable.

People sometimes use a coax cable to feed a dipole; although this works, it greatly distorts the pattern of the antenna, because the coax shield now becomes part of the antenna, and itself radiates.

The counterpoise

When we drew the electric and magnetic fields in Fig. 6-3, we specifically referred to a dipole, and we showed the electric field extending from one end of the dipole to the other. A similar thing occurs with a vertical antenna, except that this time the electric field extends from the top of the vertical whip down to the ground plane under it, as shown in Fig. 6-14. In other words, the ground plane (and the coax shield it connects to) is an integral part of the antenna.

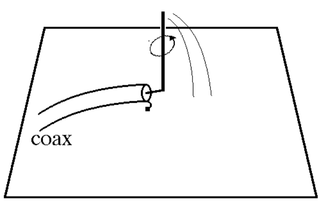

Fig. 6-14. Fields at a vertical antenna

In general, any antenna that directly generates an electric field needs two parts between which the field can extend. If only one part of the antenna is up in the air, then the other part has to be down at the bottom somewhere, so it can act "against" the top part. It is therefore often called the *counterpoise.*

This is a concept often forgotten by antenna experimenters, but it is crucial to success. If an antenna does not supply its own counterpoise (such as the other half of a dipole, for example), then an external counterpoise (usually grounded) must be provided. Sometimes the transmitter itself, such as a handheld cellular phone, itself becomes the counterpoise for the antenna on its top.

Loop antennas

A few paragraphs ago, we used the phrase "any antenna that directly generates an electric field". It turns out that there are antennas that do not.

We mentioned that radio waves consist of an electromagnetic field, which is a combination of an electric field and a magnetic field. There are antennas which generate (or detect) mainly the magnetic field; they let the buildup and collapse of the magnetic field generate the electric field which is ultimately necessary to transmit the signal through the air.

A simple example is the loopstick antenna used in almost all AM broadcast receivers. It is simply a short rod of ferrite (an insulating rod which contains metal powder), with a coil wound around it. As the magnetic component of the electromagnetic field passes through it, the coil generates a voltage. The advantage of such an antenna is that it can be quite small — even though a half-wavelength at the AM broadcast band is on the order of 1000 feet or so, the loopstick antenna is usually just a few inches long.

There are also several models of commercial loop transmitting antennas. They are not as efficient as some other antennas, but they feature small size. For example, a dipole antenna for the 20-meter (14 MHz) amateur antenna would be about 34 feet long; a loop antenna for that band is less than one tenth of that size.

Colinear antennas

In introducing directional antennas, we discussed using multiple radiators whose signals add in some directions, and cancel in others. Our prior examples used radiators which were parallel to each other; these radiators could also be placed end-to-end, in which case the antenna is called a *colinear* antenna, because all the radiators are on the same line.

A common example consists of two or three vertical dipoles, placed one above the other. A receiver at the same height as the colinear transmitting antenna will get the sum of the dipoles' signals, but the signals heading for a receiver at a slightly higher or lower altitude will partially cancel. The effect is to take the dipole's normal vertical radiation pattern, shown in Fig. 6-7, and squeeze it (much like taking the donut that it represents, and sitting on it!) The radiation pattern of Fig. 6-7 wastes some signal by sending it down into the ground and up into the clouds; the colinear antenna reduces the radiation in those directions, and sends out more of the signal horizontally.

Nonresonant antennas

You probably know that in a resonant circuit, the capacitive reactance and the inductive reactance are equal, and they therefore cancel. That is, a resonant circuit appears as a pure resistance because the reactance is cancelled out. The antennas we've discussed so far in this chapter were resonant also; that is, their length (some multiple of a ¼ wavelength) made them appear as a pure resistance load.

DETOUR

When you calculate the length of an antenna in wavelengths, remember to consider the speed of the signal in the antenna wire — the velocity factor. The velocity factor of a plain wire depends slightly on the diameter of the wire, but it is about 0.95, so a ¼-wavelength antenna would be about 5% shorter than one-quarter of a wavelength in air.

END OF DETOUR

Many antennas, however, are nonresonant, or perhaps resonant at some frequency other than what we want to use them at. This adds a capacitive or inductive reactance, which means that there will be some mismatch to the resistive Z_o of the line that feeds them. The common solution is to add just enough a capacitance or inductance to the circuit to cancel out the reactance of the antenna.

This trick is often used to shorten an antenna. For example, a ¼-wave vertical antenna for the 27 MHz CB band would be about 102 inches long, a bit unwieldy for most mobile operators. The antenna can be shortened, but then it has a capacitive reactance. This can be cancelled out with a loading coil (an inductance) at the base or near the bottom of the antenna. Likewise, a ¼-wave whip for a 2-meter amateur walkie-talkie would be about 19¼ inches long; the antenna can be shortened but then appears capacitive. Many such radios thus use a "rubber duckie" antenna, which winds the antenna in a helical coil and thus adds inductance to cancel out some of the capacitance and make it resonant.

The disadvantage is that this greatly reduces the efficiency of the antenna. Shortening an antenna by 50%, for example, reduces its efficiency by much more than 50%. This doesn't matter much in most receive applications, but is important in a transmitter because the extra inductance tends to heat up and absorb power that should be transmitted.

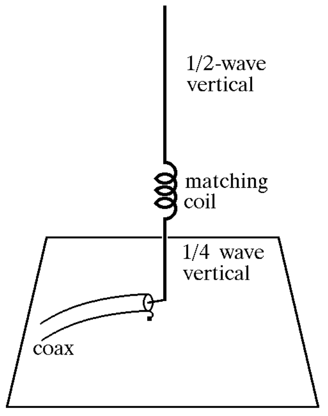

Fig. 6-15. A common cellular antenna

Feed methods

So far, we've seen antennas with the feedline connected in the middle (as in the dipole or the driven element in the beam) and at the end (in the vertical antenna). Antennas can also be fed at other points, such as slightly off the middle, or at the ⅔ point. In general such antennas do not provide a resistive load, and so some extra capacitance or inductance is needed to make them a good load for the transmission line.

Modern cellular phone antennas are an interesting example of a combination of different feed methods to make a colinear antenna. Most mobile cell phone antennas look like Fig. 6-15. If we break down the antenna into its parts, we see a ¼-wave vertical at the bottom, with a ½-wave antenna above it, making a colinear antenna. But the ½-wave antenna at the top is fed at its bottom end rather than in the middle like a dipole. A short inductor between the two antennas takes some of the signal from the bottom antenna and couples it into the top antenna.

Antenna gain

We have shown that directional antennas concentrate the power in a desired direction, and reduce the power going off in undesired directions. This implies that the directional antenna puts out a stronger signal in its desired direction than a non-directional antenna would. This improvement is called an antenna's *gain*. So if one antenna puts out a signal that is 3 db stronger than that of a nondirectional antenna, we say that it has 3 db gain. The catch, of course, is that we have to aim the directional antenna correctly.

Well, there is actually another catch too. Every antenna is directional — there is no such thing as a truly nondirectional antenna, since even a simple dipole or ¼-wave vertical transmits nothing off its ends. So to be able to do any meaningful comparisons, we have to devise a non-directional antenna first.

Enter the *isotropic antenna*. This antenna is impossible to build, but it is useful to imagine it anyway. We assume that the isotropic antenna is (1) perfectly efficient, with no losses, and (2) perfectly nondirectional. All the power it gets from the transmitter is sent out into space equally in all directions.

So let's connect the isotropic antenna to a transmitter with some transmission line. If the power going into the isotropic antenna is P watts, then the *Effective Isotropic Radiated Power* or EIRP coming out of the isotropic antenna is also P watts.

The idea of EIRP becomes important when we consider a directional antenna. Suppose the directional antenna aims its signal so that in some desired direction its signal is a thousand times as strong as the isotropic antenna would be. The word "effective" implies that only the power actually going toward the receiver is useful or effective, so the Effective Isotropic Radiated Power of this directional antenna is then also a thousand times as large. A 1-watt transmitter feeding such an antenna would put out as strong a signal *in this one desired direction* as a 1000-watt transmitter using an isotropic antenna; the 1-watt transmitter and its directional antenna would then be putting out an effective isotropic radiated power or EIRP of 1000 watts. What this points out is that it is not a good idea to stand in front of a very directional, high gain antenna, even if the transmitter power is fairly small, because the EIRP in your direction could still be large.

Back to the isotropic antenna. Suppose we send P watts into it, to be radiated into space in all directions. Let's then build a large sphere around the antenna, and collect all the power it radiates — we should then get our P watts back. (Don't worry about how we're going to do this — this is only a theoretical exercise anyway.)

Since this is an isotropic antenna, every part of the sphere gets an equal amount of power. If the sphere has a radius of R meters (the common unit of measurement for this calculation), its surface area is $4\pi R^2$ square meters. Splitting the P watts into $4\pi R^2$ little pieces, each one square meter in size, tells us that the power hitting each and every square meter of the sphere's surface is

$$\frac{P}{4\pi R^2} \text{ watts per square meter.}$$

This number is called the *power density* at that distance from the antenna. More generally, since an isotropic antenna getting P watts also has an EIRP of P watts, we would write this as

$$\text{power density} = \frac{\text{EIRP}}{4\pi R^2} \text{ watts per meter}^2$$

Let's try an example. The power density of a 10-watt signal being transmitted by an isotropic antenna (whose EIRP is thus 10 watts), calculated 1000 meters away (about ⅔ of a mile) is

$$\text{power density} = \frac{\text{EIRP}}{4\pi R^2} = \frac{10 \text{ watts}}{12{,}566{,}360 \text{ m}^2} = 7.96 \times 10^{-7}$$

which is about 0.796 microwatts per square meter.

Let's now switch to a dipole, still assuming little or no loss in the antenna itself. The same 10 watts of power is now being concentrated broadside to the dipole, with little or no power coming off the ends of the dipole. A receiver broadside to the dipole will now get more signal than it got with the isotropic antenna.

Broadside to the antenna, a dipole transmits 1.64 times more power than the isotropic antenna. The dipole therefore has a gain of 1.64 over an isotropic antenna, and the EIRP is now 16.4 watts. Translated into decibels, we get

$$10 \log \frac{1.64}{1} = 10 \times 0.214 = 2.14 \text{ dB},$$

So the half-wave dipole has a gain of 2.14 dB over an isotropic antenna. To remind us that the comparison is with an isotropic antenna, we write that as 2.14 dBi (i for isotropic).

Obviously, then, an antenna with high gain has to be very directional, since we never get something for nothing — what looks like gain is just the antenna aiming most of the radiated power in some preferred direction, at the expense of other directions.

Let's continue with our example. Suppose our 10 watt signal were radiated with a test antenna having a gain of 3 dB over a *dipole*; we say that its gain is 3 dBd (d for dipole). If the antenna has gain, then it is directional and so we must aim it toward the receiver; hence we must talk about the gain in its major lobe.

So we might then ask — what would be the power density 1000 meters away (in the major lobe, obviously)? We already know the power density for an isotropic antenna, so we need to convert dBd to dBi. If our test antenna has a gain of 3 dBd (3 dB over a dipole), and the dipole itself has a gain of 2.14 dBi (2.14 dB over an isotropic), the test antenna has a gain of 5.14 dBi (you add the two dB ratings.)

Using the standard formula for converting power gain into dB, we work it backward to get a power gain of about 3.27:

$$5.14 \text{ dB} = 10 \log \frac{P_{test}}{P_{isotropic}}$$

$$0.514 = \log \frac{P_{test}}{P_{isotropic}}$$

$$\frac{P_{test}}{P_{isotropic}} = 10^{0.514} = 3.27$$

In other words, the power radiated in the desired direction (the major lobe) of the antenna will be 3.27 times that produced by an isotropic radiator, and so will the power density. (And our EIRP is now up to 32.7 watts.)

In our example, the power density would then be

$$3.27 \times 7.96 \times 10^{-7} = 2.60 \text{ microwatts/meter}^2$$

An easier way to get to this same number is to use the EIRP in the numerator of the power density formula, like this:

$$\text{power density} = \frac{\text{EIRP}}{4\pi \text{R}^2} = \frac{3.27 \times 10 \text{ watts}}{12{,}566{,}360 \text{ m}^2} = 2.60\ \mu\text{w/m}^2$$

Signal Strength

The above calculation gives us the power density a certain distance from the transmitting antenna. However there are commercial signal strength meters, which measure the strength of a signal not as a power density, but in units of volts per meter, and it would be useful to be able to convert from one to the other.

Just as we normally calculate power as

$$\text{Power} = \frac{V^2}{R}$$

so we can calculate the power density as

$$\text{Power density} = \frac{\text{field strength}^2}{R}$$

But what is *R*? *R* is the resistance that the signal goes through in space. Say that again?

This is another concept that requires some more advanced physics; let's just say that free space (really vacuum, but air is similar enough) has a *characteristic wave impedance* which, for all intents and purposes, is like the resistance *R* in an electric circuit; its value is 377 ohms.

In this equation, the power density is measured in watts per square meter, while the field strength is measured in volts per meter. To go from a power density to field strength, we have to rearrange the equation to

$$\text{Field strength} = \sqrt{\text{Power density} \times 377 \text{ ohms}}$$

In our example, for instance, we had a power density of 2.60 microwatts per meter^2. The field strength is therefore

$$\text{Field strength} = \sqrt{.00000260 \text{ watts/m}^2 \times 377 \text{ ohms}}$$

$$= \sqrt{0.00098}$$

$$= 0.031 \text{ volts/meter}$$

Like some other concepts in antenna work, field strength is somewhat theoretical. It is based on the idea that, if you could somehow stick two voltmeter probes into the air, exactly one meter apart, the meter would measure a voltage of (in this case) 0.031 volts. This is not really possible, of course; actual field strength meters measure the field strength by measuring the output from a calibrated antenna.

Field strength calculations can be useful if you ever get your hands on a calibrated field strength meter, but otherwise are not very useful.

Capture Area

As you remember, power density is the amount of power that hits a one square-meter area at some distance from the transmitter antenna. Let's now place an antenna at that point, and make the antenna exactly one square meter in size. If the antenna can capture all the power hitting it, it will receive the same amount of power. For example, if the power density was 2.60 microwatts per square meter, as in our previous example, a one-square-meter antenna would receive 2.60 microwatts of power. A two-square-meter antenna would receive twice as much power, etc.

The catch is that the actual physical area of an antenna doesn't always match exactly the amount of power it captures. Some antennas simply don't capture enough of the signal hitting them, while others capture more signal than their size would indicate — they seem to "reach out" into space around them to capture some signal that would otherwise pass on by. So, rather than talk about their physical area, we consider the effective or working area.

The effective area of the antenna is called its *capture area*. Once we know the capture area, we can compute how much signal the antenna actually receives from the formula

$$\text{received power} = \text{power density} \times \text{capture area}$$

The greater the capture area of a receiving antenna, the greater the amount of power it picks up out of the air and sends to a receiver.

As with so many other antenna concepts, the idea of a capture area is purely theoretical. For instance, if it really did what it sounds like it does, namely capture *all* the power existing in a certain area of space, then a second antenna placed behind the first antenna would pick up no signal at all, and we know that is not true. Similarly, putting a reflector behind a dipole would do nothing because there would be no signal there to reflect, whereas we know that reflectors are commonly used in beam antennas. Still, capture area is useful because it allows us to calculate other antenna parameters. Specifically, it lets us know how much rf signal a given antenna will pick up and deliver to the receiver.

Measuring the capture area, however, is difficult, so we usually work backward. Instead of estimating capture area and using it to calculate the gain, we measure the gain and use it to calculate the capture area. The gain of an antenna can be measured by comparing it with that of an antenna with a known gain (such as a half-wave dipole.) Once we have that, we calculate the capture area from the following equation:

$$\text{capture area} = \frac{\text{Gain} \times \text{wavelength}^2}{4\,\pi}$$

where Gain is the gain compared with an isotropic antenna (expressed as a number, not as dBi), and the wavelength is simply the wavelength of the signal which the antenna is trying to pick up.

Let's justify the equation. It's easy to see why the Gain term is in it — if you double the gain of an antenna, that means it picks up twice the signal, which means that it has twice the capture area.

But why the wavelength^2 term, and why is it squared? Let's consider an example. Let's assume that we have a 3 dBi antenna of, say, 2 by 3 feet. Let's now build an identical type of antenna, but for half the frequency. This new antenna will also have 3 dBi gain, since it is the same type of antenna. Yet every dimension of the new antenna has to be twice as large (because the wavelength is twice as large), and so it has a capture area four times as large. So, although the gain has stayed the same, the wavelength has doubled and the capture area has gone up by a factor of 4. So the capture area is proportional to the square of the wavelength.

Dish Antenna Gain

Dish antennas typically have the greatest gain and capture area. They use a spherical or parabolic reflector, which reflects the signal much as the reflector in a flashlight concentrates the bulb's light in a single direction. To be effective, dish antennas have to be substantially larger than a wavelength; this usually limits their use to the higher frequencies. (Dish antennas can be used at lower frequencies, but their size makes them very difficult to build and aim. For example, a dish at Arecibo in Puerto Rico, used for radio astronomy, is 1000 feet in diameter, and actually rests on hollowed-out ground. That huge size makes it usable down to 50 MHz, at which it is about 50 wavelengths in size.)

Dishes for TV reception from satellites are quite common. Today's more powerful satellites can make do with 16-inch or 18-inch diameter dishes; in the past, 10- and 12-foot antennas were needed.

To get an idea of dish antenna characteristics, look at these equations: the gain of a parabolic dish antenna is

$$Gain_{db} = 20 \log (f\,D) - 52.6$$

while the beamwidth (the width of the beam in degrees) is

$$Beam\ Width = \frac{6.875 \times 10^4}{f\,D}$$

where f is the frequency in MHz, and D is the diameter in feet (since dishes are usually measured in feet, we switch from meters to feet).

Thus a 5-foot diameter antenna for 2 GHz would have a gain of

$$Gain_{db} = 20 \log (2000\ Mhz \times\ 5\ ft) - 52.6 = 27.4 \text{ dB}$$

and a beamwidth of

$$Beam\ Width = \frac{6.875 \times 10^4}{2000\ MHz \times 5 \text{ feet}} = 6.875 \text{ degrees}$$

The 27.4 dB gain means that a 1-watt transmitter would have an effective isotropic radiated power (EIRP) of more than 500 watts! This shows why dish antennas are so popular (when the frequency makes their size practical.)

Practical Example

Fig. 6-16 shows a typical problem from amateur radio. It shows a 0.1 watt transmitter on 449 MHz, feeding a 9 dB gain beam through a coax which has 4 dB loss. At the receiver, ½ mile away, a similar antenna feeds a receiver through a 52 ohm coax having a loss of 2 dB. Under these conditions, how much signal will the receiver get?

Our calculations go like this:

1) Transmitter power is 100 milliwatts into the coax.

2) The antenna has 9 dB gain, but there is 4 db loss in the coax cable feeding it, so the total power gain is only 5 dB (in the desired direction!) A 5 dB power gain is a power ratio of 3.16, so the power actually radiated toward the receiver is the same as an isotropic antenna would radiate if it was fed with

$$3.16 \times 100 = 316 \text{ milliwatts.}$$

In other words, the EIRP is 316 mw or 0.316 watt.

3) A half mile is 1609/2 meters, or 805 meters. The power density at that distance is thus

$$\frac{EIRP}{4\pi r^2} = \frac{0.316\ watt}{4 \times 3.14159 \times (805)^2} = 0.0388 \text{ microwatts/meter}^2.$$

4) 9 dB antenna gain on the receiver is a power ratio of 8. (Here's a shortcut to figure that out: 9 dB is 3 dB + 3 dB + 3 dB. Since each 3 dB power gain doubles the power, the power increase is $2 \times 2 \times 2$, or 8.)

The wavelength at 449 MHz is

$$\frac{3 \times 10^8 \text{ meters/sec}}{449 \times 10^6 \text{ cycles/sec}} = 0.668 \text{ meters/cycle}$$

With a 0.668 meter wavelength and a gain of 8, the receive antenna's capture area is

$$\frac{\text{Gain} \times \text{wavelength}^2}{4\pi} = \frac{8 \times (0.668 \text{ m})^2}{4 \times 3.14159} = 0.284 \text{ m}^2$$

and so the received power at the receiver's antenna is

$$\text{received power} = \text{power density} \times \text{capture area}$$

$$= (0.0388\ \mu\text{w/m}^2) \times 0.284 \text{ m}^2$$

$$= 0.011 \text{ microwatts} \quad .$$

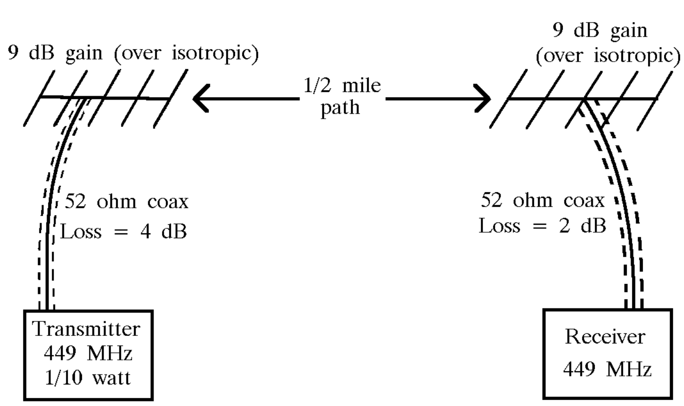

Fig. 6-16. A practical example from ham radio

5) Another 2 dB is lost in the receive coax line; we translate that to a ratio of 1.59 using the equation

$$2 \text{ dB} = 10 \log \frac{P_2}{P_1}$$

so the power arriving at the receiver is only

$$\frac{0.011 \text{ microwatts}}{1.59} = 0.0069 \text{ microwatts.}$$

6) Since $P = V^2/R$, we can find the actual voltage at the 52 ohm receiver input:

$$V^2 = P \times R$$

$$V = \sqrt{P \times R}$$

$$= \sqrt{6.9 \times 10^{-9} \text{ watts} \times 52 \text{ ohms}}$$

$$V = 5.99 \times 10^{-4} = 600 \ \mu\text{volts}$$

"Figures lie, and liars figure"

Time to tell the truth. The above numbers are all nice and exact — but in practice, things never quite work out like that. There are a number of other factors which don't show up in the math, such as

- What is between the transmitter and receiver antennas?
- Do they have a clear line of sight between each other, or are there obstructions? The above math assumes a line of sight.
- What about the curvature of the earth — if the antennas are low enough, the earth may obstruct the path between them.
- Are there any reflections from other objects? Nearby buildings or hills can provide reflections, but so can the earth below! Earth reflections are less likely with vertical polarization, but they can still occur. And reflections can either add to the signal, or cancel part of it; either way, the actual signal strength at the receive antenna can be drastically different.
- How about the coax, antennas, and connections — are they in good shape, or are there additional losses due to old age, moisture, rust, or other factors?
- How well are the antennas aimed?
- Is the polarization of both transmitter and receiver antennas the same?
- And yes ... did the antenna manufacturer tell the truth in specifying 9 db gain?

Since there is so much variability in these factors, it is usually a good idea to assume that the results could be off by a factor of 10 or more. In other words, a real-life system had better provide ten times more power than the calculations indicate is needed. Still, such calculations do give you a rough idea of the *minimum* reasonable power that might do the job.

There are two areas where there is often some honest confusion, and this can also affect your results. One area is the difference between dBd and dBi. Some manufacturers specify antenna gain in dBd, some in dBi, and some in just plain dB, without specifying the reference. It is not always easy to tell which is used. To avoid confusion (and possible deception!), some magazines therefore refuse to accept antenna advertisements which provide any dB gain figures at all.

A second area involves EIRP and ERP. We have been using the abbreviation EIRP for the effective isotropic radiated power — the radiated power of an isotropic antenna. Some users use ERP or just plain effective radiated power for the same quantity; others use ERP to mean the effective radiated power of a dipole, not an isotropic antenna. So if you see the abbreviation ERP, it is sometimes unclear as to which one is meant.

Path Loss

In the above example, we started with a transmitter output of 100 milliwatts and wound up with only 0.0069 microwatts at the receiver. This is a total loss of

$$\text{Loss } in\ dB = 10 \log \frac{0.0069 \text{ microwatts}}{100 \text{ milliwatts}}$$

$$= 10 \log \frac{6.9 \times 10^{-9} \text{ watts}}{1 \times 10^{-1} \text{ watts}}$$

$$= 10 \log (6.9 \times 10^{-8}) = -71.6 \text{ dB}$$

Let's see what the signal had to go through on its way from the transmitter to the receiver: a cable at the transmitter; a transmit antenna; half-mile of air; a receive antenna; and some cable at the receiver. Let's then add up the losses in each of these:

Cable at the transmitter	–4 dB
Transmitter antenna	+9 dB
1/2 mile of air	–X dB
Receive antenna	+9 dB
Cable at the receiver	–2 dB
TOTAL	–4 + 9 – X + 9 – 2 = +12 – X dB

But we already know that the total loss is 71.6 dB, so

$$+12 - X \text{ dB} = 71.6 \text{ dB}$$

$$X = 83.6 \text{ dB}$$

In the above example, the antennas actually contributed an 18 dB *gain* (9 dB for each antenna), while the cable loss added up to 6 dB (4 dB at the transmitter, 2 dB at the receiver). This adds up to a total *gain* of 18 – 6 = 12 dB. In other words, we had an effective gain of 12 dB in the antenna systems, and still lost 71.6 dB in the transmission; this means that the loss in the 1/2-mile path was actually 71.6 + 12 = 83.6 dB. This is called the *path loss.*

The path loss is actually dependent only on the distance and the frequency. It is calculated by assuming that isotropic antennas are used at both the transmitter and receiver, and there are no other losses in the coax cables. We then use the foregoing equations to calculate, step by step, the received power in relation to the transmitted power.

Alternatively, we can combine all of the above equations into one big equation which gives the path loss directly in dB:

$$\text{Path loss } \textit{in dB} = 10 \log \frac{(4\pi)^2 \times \text{distance}^2}{\text{wavelength}^2}$$

where both the distance between the transmitter and the receiver, and the wavelength, must be given in meters.

The path loss is useful not only in cases where we want to get a signal from one place to another, but also in cases where we *don't*. For example, suppose a receiver is located 1/5 mile (322 meters) away from someone else's transmitter on a nearby frequency, and we worry that the nearby transmitter might interfere with our efforts to receive a weak signal. How much interference will the transmitter cause to the receiver? The path loss is a guide to how much the transmitted signal will be attenuated in the 1/5-mile path:

$$\text{Path loss } \textit{in dB} = 10 \log \frac{157.91 \times (322 \text{ m})^2}{(2 \text{ m})^2} = 66.1 \textit{ dB}$$

This means that if both the transmitter and receiver have isotropic antennas and no loss in the coax, the received signal will be 66.1 dB weaker than the transmitted signal. In an actual case, you would have to add in any antenna gains, and then subtract cable or other losses, so the actual signal loss in a real example might actually be smaller once these are taken into account.

Propagation

The above path loss calculations are theoretically satisfying, but in practice we have to also consider propagation; that is, the way in which our radio signals propagate (or travel) from place to place. We can summarize this as follows:

Low frequency signals follow the curvature of the earth (an effect called the *ground wave*), and tend to go around corners and objects. *Very* low frequency signals even travel through water (useful for communications to submarines.)

Very high frequency signals tend to travel in a straight line. In that sense, they behave somewhat like light, which is an extremely-high-frequency electromagnetic wave. They also easily bounce off conductive objects (like light reflects off a mirror). Very high frequency signals are easily absorbed by moist objects, such as leaves on trees. (Think of a microwave oven, where anything moist absorbs the microwaves and rapidly heats up.)

Low frequency signals could bounce off objects too, but these objects must be extremely large and relatively flat, to go with the long wavelength of these signals. There are few such objects around, and so low frequencies seldom bounce, except perhaps from the surface of the earth itself.

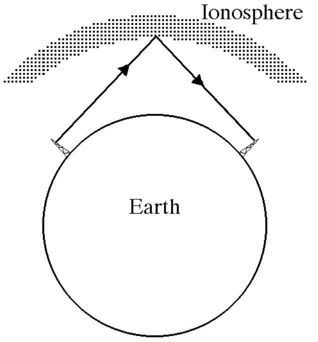

Fig. 6-17. Reflection from the ionosphere

Mid-frequency signals are sort of a compromise. They will follow the curvature of the earth but only for a very short distance. They prefer to travel in straight lines, but will make it around corners and objects. The primary difference here is in their bouncing or reflections, specifically off the ionosphere.

The Ionosphere

The *ionosphere* is a layer of ionized gases ranging from about 30 miles above the earth up to about 400 miles. At this height, the air is extremely thin, and consists mostly of the lighter, and simpler, elements such as hydrogen and helium. Radiation from the sun is also much stronger at this height, and includes strong ultraviolet light, X-rays, etc., which ordinarily do not reach down to lower altitudes. This radiation attacks the thin gas atoms in the air, and tears the electrons away from their nuclei, producing what are called *ions*. These ions are charged (electrons are negative, while the remaining atom nuclei are positive), and they reflect radio signals.

The result, shown in Fig. 6-17, is that mid-frequency signals (called short waves) can travel large distances even though they do not follow the curvature of the earth — they go up to the ionosphere, and are reflected back to the earth. In some cases, they can bounce a second time and provide even longer-distance communications.

The thickness and composition of the ionosphere depends, of course, on the radiation from the sun. It therefore changes between day and night, but it also changes with the rotation of the sun, and also with the number of sunspots — areas on the surface of the sun which generate bursts of energy and radiation. Sunspots have an 11-year cycle, so the reflections also vary over 11-year periods. A sunspot peak in 2001–2002 was expected to provide strong long-distance communications.

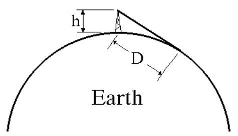

Fig. 6-18. The effect of antenna height on range

Because of ionospheric reflections, it is possible to communicate very long distances — literally around the world — even with relatively low power signals. It is, however, unreliable, since ionospheric propagation varies greatly with time of day, the frequency of the signal, sunspots, and many other factors.

And in some cases, the ionosphere can even reduce the quality of communications. Strong AM broadcast stations can routinely be heard 50 or more miles away, day and night, because their ground-wave signals follow the curvature of the earth. But at night there are additional reflections from the ionosphere, which sometimes add to the ground wave, and other times interfere with it. Hence people who get clear AM reception during the day often find the reception spotty evenings and night time.

Height and HAAT

The height of an antenna is also important, especially at those frequencies which travel mostly in a straight line, and do not appreciably bend around the earth or reflect from the ionosphere. If an antenna is mounted at height *h* above the ground, the signal can only travel as far as the antenna can "see", shown as distance *D* in Fig. 6-18. This distance, called the *radio horizon*, is approximately

$$D = \sqrt{2\,h}$$

where *h* is in feet, and *D* is in miles. For example, the radio horizon for an antenna 20 feet high is a bit over 6 miles.

In practice, even high frequencies, which supposedly travel only in a straight line, do curve just a bit around the surface of the earth. The effect is to lengthen the radio horizon by roughly an extra 15%, which would increase the distance in the above example to roughly 7 miles instead of 6.

The height *h* is sometimes also a bit difficult to measure. A 10-foot tower on a tall mountain can be clearly better than a 100-foot tower in a low valley. Antenna height is therefore often measured as "height above average terrain" or HAAT — the height above the average lay of the land. But it is difficult to define what we mean by "average terrain", since asking five people to estimate the average height of the land around an antenna is likely to give five very different answers.

Some Complications

The ground below an antenna, and the ionosphere above the antenna, can both affect antenna performance, with the result that some of the theoretical factors that we have been discussing may be too simplified.

For example, consider a horizontal dipole shortwave transmitting antenna stretched east-to-west. Theoretically, the dipole has a figure-8 radiation pattern, so that it transmits best at right angles to the antenna (north and south, as well as up), and radiates nothing off its ends (to the east and west.) Hence a receiver off to the east or west should get no signal.

In practice, though, the antenna may transmit nothing exactly horizontally to the east or west, but it does transmit upward toward the eastern and western sky. In fact, at some heights above ground, reflections from the ground may add to the signal to produce a fairly strong signal at an angle of 40 to 60 degrees upward in those two directions. This may be at just the right angle to hit the ionosphere and reflect back to a receiver hundreds or thousands of miles away. The result is to get a signal in an unexpected direction.

Long distances at very high frequencies

As mentioned above, very low frequency signals can curve around the earth for long distances; mid-range frequencies can bounce off the ionosphere and provide long-distance communications as well. But very high frequencies would appear to be limited to line-of-sight distances by the $D = \sqrt{2\,h}$ equation above. How, then, can they provide long-distance communications?

The simple answer is to bounce them off high-altitude objects. For example, it is possible to bounce strong high-frequency signals off the moon. *EME* or *earth-moon-earth* signals can cover a significant portion of the earth, but the long distances involved require high power.

Another common approach is to bounce them off meteor trails — the ionized gas trails left behind by a meteor streaking across the sky — or the dust and water particles in the troposphere, the bottom 10 miles or so of the atmosphere. The resultant *meteor-scatter* or *tropo-scatter* communications have been around for some dozens of years.

A more modern and better approach is to use a satellite.

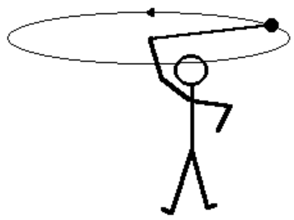

Fig. 6-19. Swinging a rock in a circle

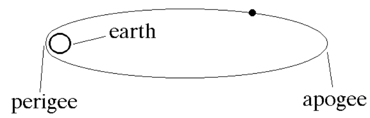

Fig. 6-20. Molnyia orbit

Communication Satellites

Communication satellites have been used since the 1960's to provide long-distance communications at very high frequencies. A signal is usually sent up to the satellite, which receives it, amplifies it, and sends it back to the earth at a different frequency. Sometimes the satellite simply amplifies the signal, other times it may regenerate it or process it in other ways.

Satellites are often classified by the type of orbit: synchronous, low-earth-orbit (LEO) or Molniya. To explain the orbits, imagine that you have a rock attached to a string, and you want to swing it around you in a circle as in Fig. 6-19. You have to keep it moving at a good clip to keep it from falling down. If the string is very short, you have to rotate it very fast; if the string is long, you can rotate it much slower.

The same principle applies to satellites — as the satellite travels around the earth, its own momentum wants to keep it going in a straight line, eventually flying away from the earth. At the same time, however, gravity wants to make it fall down. If the satellite is moving at the right speed, the two forces will exactly counteract each other and prevent it from falling down.

Much like the rock at the end of the string, the farther the satellite is from the earth, the slower it need rotate to keep from falling. For example, the moon is so far away from the earth that it need only make one revolution every 30 days to keep from falling down.

Geo-stationary or Synchronous satellites

Some satellites are placed about 22,300 miles up. At that height, they rotate around the earth in exactly 24 hours, and so are "synchronous" with the earth. Since they are placed directly above the equator and rotate in the same direction as the earth, they are always above the exact same spot on the earth. That is, viewed from the earth, they appear to stand still, hence the name *geo-stationary*. That means that, once you properly aim an antenna at the satellite, it will stay aimed without having to be moved. Moreover, at their height, geo-synchronous satellites "see" almost a third of the entire earth at one time. They can provide signals for a huge chunk of the earth at one time.

But geo-stationary satellites have two main disadvantages. One is their large distance from the earth. Until recently, 10- and 12-foot-diameter dishes were needed to provide enough antenna gain; fortunately, they did not have to be moved to stay pointed at the satellite, but it was still awkward. Only recently have some of the TV satellites become powerful enough to support receive dishes smaller than two feet in diameter.

A second disadvantage is that there is a limited number of "parking spots" where they can be placed. In order to not move, they must be exactly above the equator, and at exactly the right height. Furthermore, to avoid interference between adjacent satellites, they must be separated by 3 or 4 degrees of arc. This places a limit on the number of satellites that can be up there.

Low Earth Orbit or LEO Satellites

LEO satellites are exactly what the name implies — low earth orbit. They orbit the earth, but much lower down than the geo-synchronous satellites, typically just a few hundred miles up. There is a large number of such possible orbits, so there is plenty of room for them.

LEO satellites have one very significant advantage: they are so much closer to the earth that large dish antennas and high power are not needed. Small whip or flat antennas are quite sufficient, so reliable communications can be had with portable equipment. Several companies provide portable equipment that works with LEO satellites.

On the other hand, at their lower height they must rotate much faster than the earth. Viewed from the earth, they seem to constantly be moving. Some are so low that they appear on the horizon, whip overhead in less than a half hour, and then again disappear. To get any kind of continuous coverage, dozens of such satellites are needed.

Since the LEO satellites are constantly moving overhead, their ability to work with small, non-directional antennas is a blessing — having to constantly swing large dishes from horizon to horizon as satellites move overhead would be a tremendous problem.

Molniya Orbits

Fig. 6-20 shows the orbit of a Molniya satellite. The orbit is similar to an ellipse, except that the earth is off at one end of the ellipse, rather than its center.

At the *apogee*, the satellite is very far from the earth and moves very slowly. Then, as it comes closer to the earth, it speeds up and swings behind the earth, through the *perigee*, its closest point to the earth. At this point, it is moving so fast that it swings past the earth in just a few minutes. Then, as it moves back out toward the apogee, it is moving away from the earth and so it slows down.

The Molniya orbit has some of the advantages and disadvantages of the geo-synchronous orbit. At the apogee, the satellite moves so slowly that it almost seems to stand still; on the other hand, it is so far away that the signal is weak and large gain antennas are needed.

At the perigee, on the other hand, it is so close to the earth that very little power is needed to communicate. On the other hand, it is in and out of range so fast that it is almost useless. It is also so low, that it sees only a small part of the earth.

Conclusion

We have described how electromagnetic fields — fields which consist of interacting electric fields and magnetic fields — can be sent through space. We also discussed how antennas are used to launch these fields into space at the transmitter, and receive them at the receiver. This is the basis of many of our modern electronic devices — radios, televisions, cordless and cellular telephones, radar, and more. In the next part, we will discuss some of the Methods that make all this work.

DETOUR

Some student always asks, "Where do the cable TV channels fit? How do they compare with over-the-air TV channels?" So the following table gives the details.

Channel number	Over-the-air frequency	Cable frequency
2 to 4	54 to 72 MHz	same
5 and 6	76 to 88 MHz	same
7 to 13	174 to 216 MHz	same
14 to 22	470 to 524 MHz	120 to 174 MHz
23 to 69	524 to 806 MHz	216 to 498 MHz
70 to 94	none	498 to 648 MHz
98 and 99	none	108 to 120 MHz

Channels 2 through 13 are the VHF channels, while the others are UHF. Over-the-air TV now ends at channel 69, but when the UHF channels were first established, they went up to channel 83 or 890 MHz. Over the years, channels 70 through 83 were taken away from UHF TV, and reallocated to cellular telephones and other services.

The current FCC plans are to change all analog TV over to digital High Definition TV (HDTV) by the year 2006, using only channels 14 through 51. Once the transition is made, channels 52 through 69 (698 through 806 MHz) will be taken away and reallocated to other uses.

END OF DETOUR

