Ham Radio Fun

Communications Simplified, Part 16

Peter A. Stark K2OAW
P.O. Box 209
Mt. Kisco NY 10549

So far we've covered a few of the basic antenna types—the dipole, 1/4-wave vertical, and the yagi beam. Let's look at some more general concepts.

Types of feedline

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 twinlead 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 twinlead.

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.

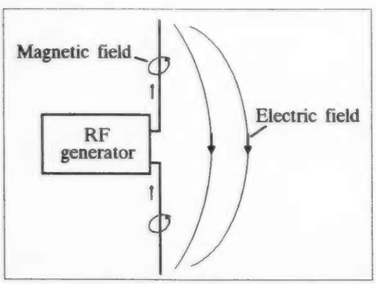

Fig. 1. Fields around a dipole.

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.

"There is one concept, often forgotten, that is crucial to success."

The counterpoise

In the electric and magnetic fields of **Fig. 1**, we specifically refer to a dipole, and we show 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. 2**. In other words, the ground plane (and the coax shield it connects to) is an integral part of the 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 amateur antenna builders, 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.

Loop antennas

Two paragraphs ago, we used the phrase "any antenna that directly generates an electric field." 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 1,000 feet or so, the

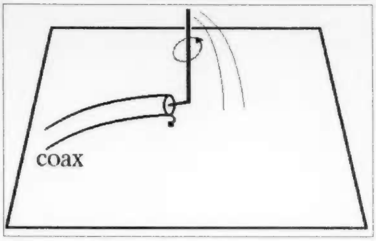

Fig. 2. Fields at a vertical antenna.

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 band would be about 34 feet long; a loop antenna for that band is less than one tenth that size.

Collinear 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 collinear 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 collinear 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, and squeeze it. The normal radiation pattern wastes some signal by sending it down into the ground and up into the clouds; the collinear antenna reduces the radiation in those directions, and sends it out more 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 canceled out. The antennas we've discussed so far in this chapter were resonant also; that is, their length (some multiple of a quarter 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 1/4-wavelength antenna would be about 5% shorter than 1/4 of a wavelength in air.

END OF DETOUR

Many antennas, however, are nonresonant, or perhaps resonant at some frequency other than where we want to use them. 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 of a capacitance or inductance to the circuit to cancel out the reactance of the antenna.

"This trick is often used to shorten an antenna."

This trick is often used to shorten an antenna. For example, a 1/4-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 canceled out with a loading coil (inductance) at the base or near the bottom of the antenna. Likewise, a 1/4-wave whip for a 2-meter amateur handie-talkie would be about 19-1/4 inches long; the antenna can be shortened but then appears capacitive. Many such radios use a rubber ducky" antenna, which winds the antenna in a coil and thus adds inductance to 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 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.

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 2/3 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 collinear antenna. Most mobile cell phone antennas look like **Fig. 3**. If we break down the antenna into its parts, we see a 1/4-wave vertical at the bottom, with a 1/2-wave antenna above it, making a collinear antenna. But the 1/2-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

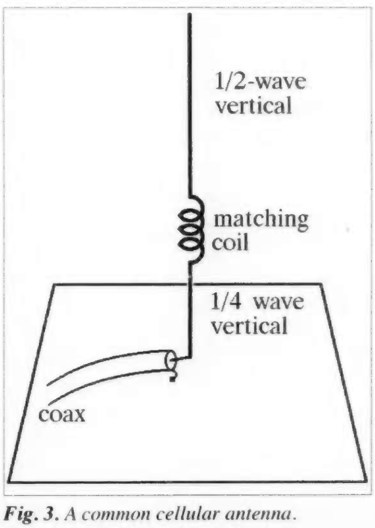

Fig. 3. A common cellular antenna.

stronger signal in its desired direction than a nondirectional 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 1/4-wave vertical transmits nothing off its ends. So to be able to do any meaningful comparisons, we have to invent a nondirectional 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 Radiated Power or ERP coming out of the isotropic antenna is also P watts.

The idea of ERP 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 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 1,000-watt transmitter using an isotropic antenna; the 1-watt transmitter and its directional antenna would then be putting out an ERP of 1,000 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 ERP 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 ERP of P watts, we would write this as

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

"This antenna is impossible to build, but useful to imagine."

Let's try an example. The power density of a 10-watt signal being transmitted by an isotropic antenna (which has an ERP of 10 watts), calculated 1,000 meters away (about 2/3 of a mile), is

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

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 it. A receiver broadside to the dipole will now get more of a 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 ERP 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 1,000 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 backwards 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 ERP is now up to 32.7 watts.)

In our example, the power density would then be

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

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

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

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:

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

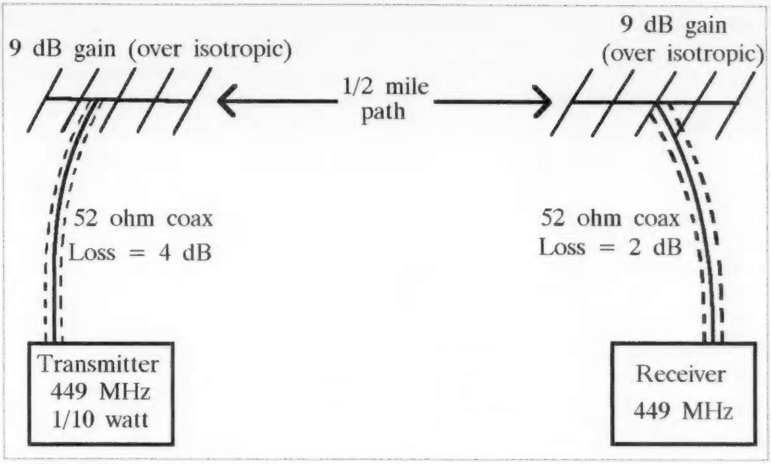

Fig. 4. *A practical example from ham radio.*

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. If that antenna was two square meters in area, then it 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

received power = power density x 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 a useful concept 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 backwards. 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:

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

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.

Practical example

Fig. 4 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, 1/2 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 x 100 = 316 milliwatts.

In other words, the ERP 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

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

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 x 2 x 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

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

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

received power = power density x capture area

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

= 0.011 microwatts.

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:

$V^2 = P \times R$

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

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

$= 5.99 \times 10^{-4}$ = 600 microvolts.

"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.

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

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

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

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

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 a 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.

"Did the antenna manufacturer tell the truth?"

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 2-meter receiver is located 1/5 mile (322 meters) away from someone else's transmitter on a nearby frequency; in other words, 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 might be smaller once these are taken into account.