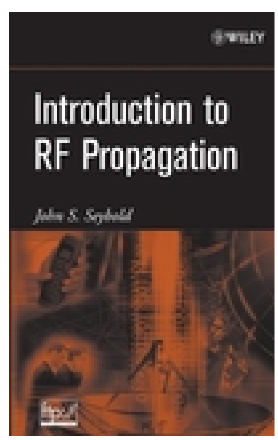
Introduction to Antennas

By

John S. Seybold, Ph.D., P.E.



Introduction to Antennas

INTRODUCTION

Every wireless system must employ an antenna of one kind or another to radiate and receive electromagnetic energy. The antenna is the transducer between the system and free space and is sometimes referred to as the air interface. While this course cannot provide a comprehensive treatment of the subject, it will enable the reader to develop an intuitive understanding about how they operate and provide a few examples of different antenna architectures. Some useful rules of thumb for relating antenna parameters to antenna size and shape are also presented. For many types of antennas it is possible to estimate the gain from the physical dimensions or from knowledge of the antenna beamwidths. A fundamental principle of antennas, called *reciprocity*, states that antenna performance is the same whether radiation (transmission) or reception is considered [1]. The implication of this principle is that antenna parameters may be measured by either transmitting or receiving. The principle of reciprocity means that estimates of antenna gain, beamwidth, and polarization are the same for both transmit and receive. This principle is used freely in most works on antennas.

An important concept in antennas is the wavelength of the electromagnetic wave that is being considered. Wavelength is the distance that the wave propagates during one cycle (period) and it depends upon it frequency and the propagation medium. Wavelength is defined as

$$\lambda = c/f$$

Where *c* is the velocity of propagation

And *f* is the frequency of the wave in Hz (recall that 1 MHz = 10^6 Hz and 1 GHz = 10^9 Hz)

The velocity of propagation is approximately $3.0*10^8$ m/s in free-space.

ANTENNA PARAMETERS

A useful abstraction in the study of antennas is the *isotropic radiator*, which is an ideal antenna that radiates (or receives) equally in all directions, with a spherical pattern. The isotropic radiator is also sometimes called an omnidirectional antenna, but this term is usually reserved for an antenna that radiates equally in all directions in one plane, such as a whip antenna, which radiates equally over azimuth angles but varies with elevation.

The power density, *S*, due to an isotropic radiator is a function only of the distance, *d*, from the antenna and can be expressed as the total transmitted power, *P*, divided by the area of a sphere with radius *d*.

$$S = \frac{P}{4\pi d^2} \tag{1}$$

That is, the power is uniformly distributed over the sphere. Thus for an isotropic radiator, the power density at a given distance from the antenna is constant over all angles and is equal to the average power density at that range.

Gain

For a realizable antenna, there will be certain angles of radiation which provide greater power density than others (when measured at the same distance). The *directivity* of an antenna is defined as the ratio of the radiated power density at distance, *d*, in the direction of maximum intensity to the average power density over all angles at distance, *d*. This is equivalent to the ratio of the peak power density at distance *d*, to the average power density at *d*: Thus an isotropic antenna has a directivity of $D = 1$. When the antenna losses are included in the directivity, this becomes the antenna gain

$$G = \eta \cdot \frac{\text{Power Density at } d \text{ in max direction}}{\mathrm{P_T} / 4\pi d^2}$$

where

P_T is the power applied to the antenna terminals
$4\pi d^2$ is the surface area of a sphere with radius *d*
η is the total antenna efficiency, which accounts for all losses in the antenna, including resistive and taper[1] losses ($\eta = \eta_T \eta_R$)

Antenna gain can be described as the power output, in a particular direction, compared to that produced in any direction by an isotropic radiator. The gain of an antenna is usually expressed in dBi, decibels relative to an ideal isotropic radiator. Thus the gain is expressed as $10 \cdot \log_{10}(g)$, where *g* is the power gain of the antenna.

[1] Taper loss is the inefficiency that is introduced when non-uniform illumination of the aperture is used. Non-uniform aperture illumination is used to control radiation pattern sidelobes, similar to windowing in spectral analysis and digital filtering. [2,3]

Effective Area

An *aperture antenna* is one that uses a two-dimensional aperture such as a horn or a parabolic reflector (as opposed to a wire antenna). The gain of an aperture antenna, such as a parabolic reflector, can be computed using an *effective area*, or *capture area*, which is defined as

$$A_e = \eta A_p \qquad \text{m}^2 \tag{2}$$

where A_p is the physical area of the antenna and η is the overall efficiency of the antenna (generally ranging from 50% to 80%).The expression for the gain of an aperture antenna is

$$G = \frac{4\pi A_e}{\lambda^2} \tag{3}$$

If only the physical dimensions of the antenna are known, one may assume an efficiency of about 0.6 and estimate the gain reasonably well for most antennas using the above expression.

> **Example 1:** Given a circular aperture antenna with a 30-cm diameter, what is the antenna gain if the frequency is 39 GHz?
>
> A 30-cm-diameter circular aperture has a physical area of 0.0707m^2. Assuming an aperture efficiency of 60% yields an effective area of
>
> $$A_e = 0.0424 \text{ m}^2$$
>
> Using the expression for antenna gain, (3) and the wavelength of 7.69 mm yields
>
> $$G = 9005$$
>
> which, when expressed in decibels relative to an isotropic radiator, is 10log(9005), or 39.5 dBi of gain. _

If the physical dimensions of an aperture antenna are not known, but the azimuth and elevation 3-dB beamwidths, θ_{AZ} and θ_{EL}, are known, the gain can be estimated using the following rule of thumb:

$$G \cong \frac{26{,}000}{\theta_{AZ}\theta_{EL}} \tag{4}$$

where the 3-dB beamwidths are expressed in degrees [4].

The effective or capture area of a linear antenna, such as a vertical whip, a dipole, or a beam, does not have the same physical significance as it does for an aperture antenna. For example, a center-fed, half-wave dipole has an effective capture area of $0.119\lambda^2$ [5]. Instead of effective area, the concept of *effective height* is employed for linear antennas [6]. The effective height of an antenna is the height that, when multiplied by the incident electric field, results in the voltage that would actually be measured from the antenna.

$$V = h_e E \qquad \text{volts}$$

Or

$$h_e = V/E \qquad \text{m}$$

where V is the voltage measured at the antenna terminals and E is the magnitude of the incident electric field in volts-per-meter.

Another way to view the effective height is by considering the average of the normalized current distribution over the length of the antenna and multiplying that value by the physical length of the antenna.

$$h_e = \frac{l}{I_0}\int_0^1 I(x)dx \tag{5}$$

Using this information, the effective area of a linear antenna can be found. The current distribution over a half-wave dipole is nearly sinusoidal [7]

$$I(x) = I_0 \sin(\pi x)$$

and for a very short dipole ($l < 0.1\lambda$) it is approximately triangular, with the peak current occurring at the center and zero current at each end. The relationship between the effective height and the effective area is given by

$$A_e = \frac{h_e^2 Z_o}{4R_r} \tag{6}$$

where

A_e is the effective area in square meters

h_e is the effective height in meters
Z_o is the intrinsic impedance of free space, 377ohms
R_r is the radiation resistance of the antenna in ohms

Example 2: Determine the effective height and the effective area of a halfwave dipole at 500MHz.

A half-wave dipole for 500 MHz has a physical length of $\lambda/2$,

$$l = \frac{1}{2}\frac{c}{500\text{ MHz}} = 0.3\text{ m}$$

For a half-wave dipole at 500 MHz, the current distribution is

$$I(x) = I_0 \sin(\pi x)$$

which yields a normalized average current value of $(2/\pi)$ or 0.637. The effective height is then found by multiplying by l = 0.3 m to get,

$$h_e = 0.191\text{ m}$$

Using the radiation resistance for a half-wave dipole, 73 ohms [8], and the impedance of free space, one can readily determine the effective area using (6):

$$A_e = 0.047\text{ m}^2$$

Note that this effective area has no direct relationship to the physical cross section of the dipole antenna. That is to say, it does not depend upon the thickness of the dipole, only its length._

The Radiation Pattern

The radiation pattern of an antenna is a graphical depiction of the gain of the antenna (usually expressed in dB) versus angle. Precisely speaking, this will be a two-dimensional pattern, a function of both the azimuth and elevation angles. This is illustrated in Figure 1 for a circular aperture antenna. In most cases, however, looking at the principal plane cuts—that is, an azimuth pattern while bore-sighted in elevation and vice versa—is sufficient. Figure 2 shows a typical principal plane pattern. Antenna patterns always describe the farfield pattern, where the gain or directivity is a function of the azimuth and elevation angles only and is independent of distance, *d*.

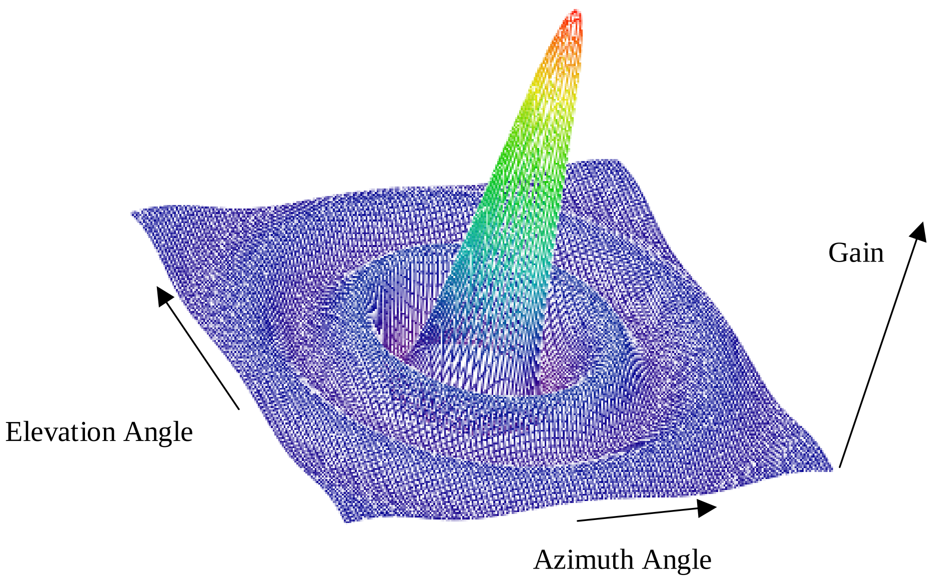


Figure 1 Three-dimensional plot of a two-dimensional antenna radiation pattern.

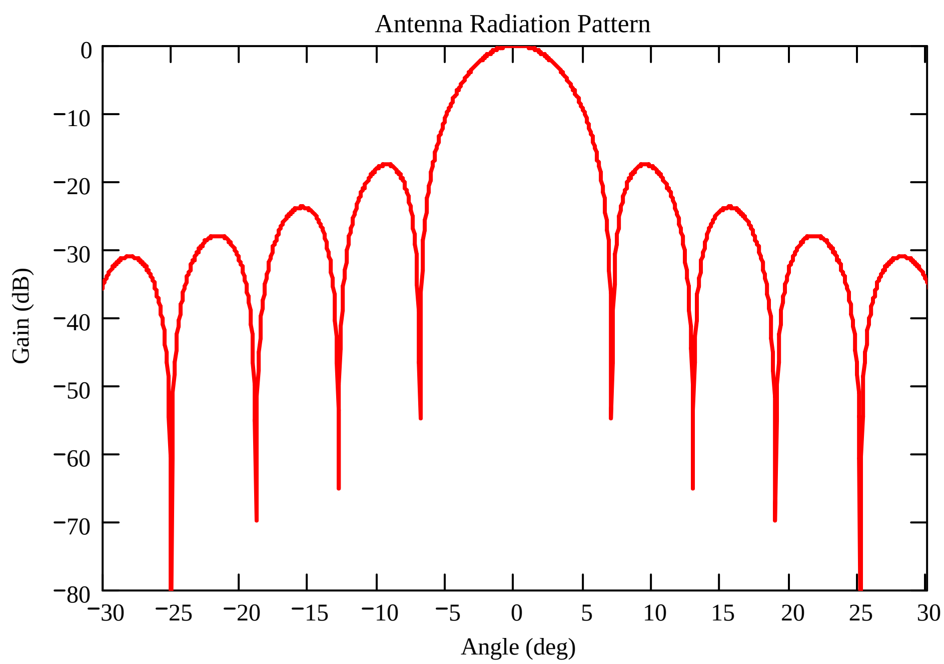


Figure 2 Typical antenna radiation pattern (normalized gain) in one dimension.

The mainlobe of the antenna is the lobe where the peak gain occurs. The *beamwidth* of an antenna is defined as the angular distance between the two points on the antenna pattern mainlobe that are 3 dB below the maximum gain point. From Figure 2, the 3-dB beamwidth of that antenna may be estimated to be about 5 degrees ($\pm$2.5). Another parameter that is often of considerable interest is the maximum or peak sidelobe level. The pattern in Figure 2 shows a maximum sidelobe level of about −17 dB relative to the beam peak, which has been normalized to 0-dB gain. Low sidelobes reduce the risk of undesired signal radiation or reception, which can be valuable in high traffic or multipath environments. Sidelobe reduction is obtained, however, at the expense of widening the mainlobe and reducing the antenna gain (due to increased taper loss), or by increasing the antenna size.

Another often-quoted specification of directional antennas is the front-to-back ratio. This is the ratio of the antenna gain at 0 and 180 degrees azimuth and provides an indication of how well the antenna will reject interfering signals that arrive from the rear of the antenna. The front-to-back ratio is a very important parameter when planning frequency reuse and interference reduction.

The antenna pattern of an aperture antenna is a function of the illumination taper across the aperture. The relationship between the spatial energy distribution across the aperture and the gain pattern versus angle is an inverse Fourier transform [9]. Thus a uniformly illuminated aperture will produce the narrowest main beam and highest possible gain at the expense of producing sidelobes that are only 13 dB below the peak. By tapering the illumination using a window function, the gain is slightly reduced (taper loss) and the mainlobe is broadened, while the sidelobes are reduced in amplitude. Illumination taper functions such as a raised cosine are often used to produce antennas with acceptable sidelobes. Whenever an illumination taper is used, the resulting antenna efficiency (and therefore gain) is reduced. Table 1 shows key parameters for several common window functions.

Table 1 Common Window (taper) functions with their relatives gains beamwidths and first sidelobe levels

Type of distribution, $\|z\| < 1$	Relative gain	Half-power beamwidth in degrees	Intensity of first sidelobe dB below max. intensity
Uniform; $A(z) = 1$	1	51λ/d	13.2
Cosine; $A(z) = \cos^n (\pi z/2)$			
$n = 0$	1	51λ/d	13.2
$n = 1$	0.810	69λ/d	23
$n = 2$	0.667	83λ/d	32
$n = 3$	0.575	95λ/d	40
$n=4$	0.515	111λ/d	48
Parabolic; $A(z) = 1 -(1 - \Delta)z^2$			
$\Delta = 1.0$	1	51λ/d	13.2
$\Delta = 0.8$	0.994	53λ/d	15.8
$\Delta = 0.5$	0.970	56λ/d	17.1
$\Delta = 0$	0.833	66λ/d	20.6
Triangular; $A(z) = 1 - \|z\|$	0.75	73λ/d	26.4
Circular; $A(z) = (1-z^2)^{0.5}$	0.865	58.5λ/d	17.6
Cosine-squared plus pedestal;			
$0.33 + 0.66 \cos^2(\pi z/2)$	0.88	63λ/d	25.7
$0.08 + 0.92 \cos^2(\pi z/2)$, Hamming	0.74	76.5λ/d	42.8

d = aperture width (diameter)
λ = wavelength
Source: *Introduction to Radar Systems*, Merrill I. Skolnik, McGraw-Hill 1980

Polarization

Polarization is defined as the orientation of the plane that contains the electric field component of the radiated waveform. In many cases, the polarization of an antenna can be determined by inspection. For instance, a vertical whip antenna generates and receives vertical polarization. Similarly, if the antenna element is horizontal, the wave polarization will be horizontal. Vertical and horizontal polarizations are both considered linear polarizations. Another type of polarization is circular or elliptical polarization. Circular polarization is similar to linear polarization, except that the polarization vector rotates either clockwise or counterclockwise, producing right-hand circular or left-hand circular polarization. Circular polarization is a special case of elliptical polarization, where the vertical and horizontal components of the polarization vector are of equal magnitude. In general, aperture antennas can support vertical, horizontal, or elliptical polarization, depending upon the type of feed that is used.

Impedance and VSWR

An antenna presents a load impedance or driving-point impedance to whatever system is connected to its terminals. The driving point impedance is ideally equal to the radiation resistance of the antenna. In practical antennas, the driving point impedance will also include resistive losses within the antenna and other complex impedance contributors such as cabling and connectors within the antenna. The driving point impedance of an antenna is important in that a good impedance match between the circuit (such as a transceiver) and the antenna is required for maximum power transfer. Maximum power transfer occurs when the circuit and antenna impedances are matched.[2] Maximum power transfer is desirable for both transmitting and receiving. When the antenna and circuit impedances are not matched, the result is reduced antenna efficiency because part of the signal is reflected back to the source. The square root of ratio of the reflected power to the incident power is called the *reflection coefficient*.

$$|\rho| = \sqrt{\frac{P_r}{P_i}}$$

Clearly, the reflection coefficient must be between zero and one (inclusive). The reflection coefficient can be determined from the circuit and antenna impedances,

$$\rho = \frac{Z_1 - Z_0}{Z_1 + Z_0} \quad (7)$$

The amount of signal passing between the transceiver and the antenna is

$$P_t = (1-\rho^2)P_i$$

Thus the impedance mismatch loss is [10]

[2] Actually, maximum power transfer required a complex conjugate match. Most antenna systems have real or nearly real driving point impedance, in which case a conjugate match and a match are identical. An antenna connected to a cable with a characteristic impedance that is different from its driving point impedance will present a complex impedance at the other end of the cable. In such cases, a transmatch can be used to perform the conjugate matching.

$$L_m = 1\text{-}\rho^2 \tag{8}$$

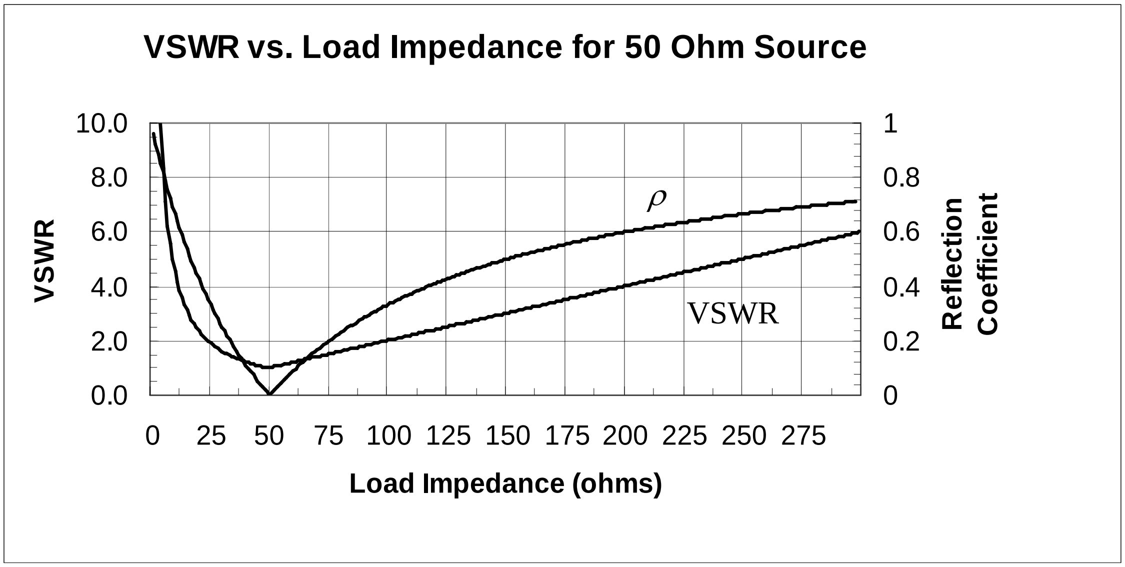
If there is a cable between the antenna and the transceiver, the mismatch creates a voltage standing wave ratio (VSWR, or often just SWR) on the cable. The effect of a VSWR on a cable is to increase the effect of the cable loss [11].
One way to compute the VSWR is

$$VSWR = \frac{1+|\rho|}{1-|\rho|} \tag{9}$$

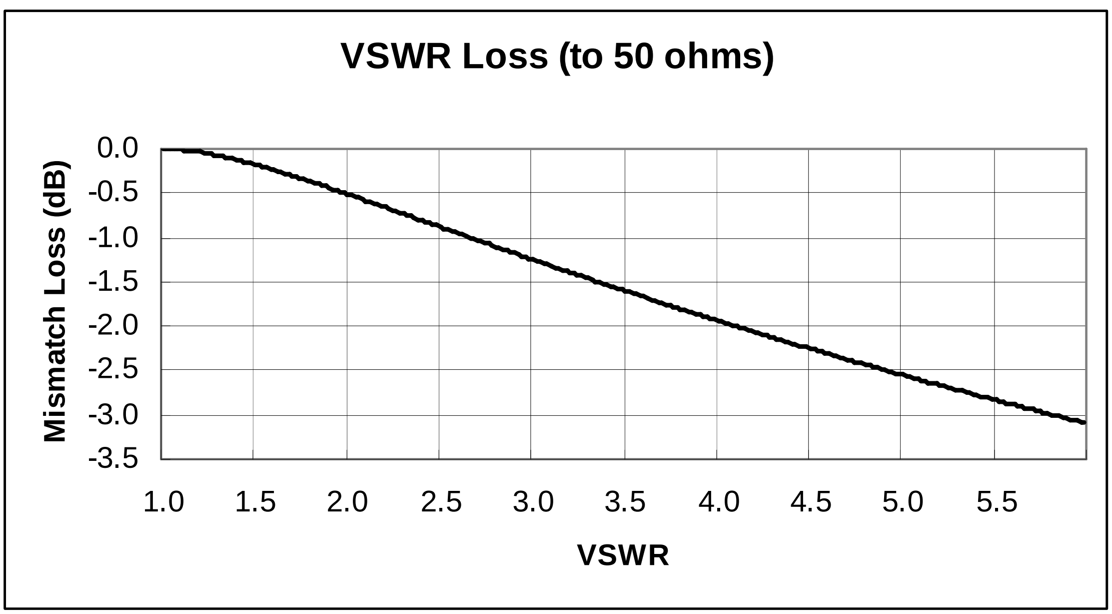
This can also be expressed in terms of the of antenna and transceiver characteristic impedances.

Figure 3a shows a plot of the VSWR and reflection coefficient for a 50 ohm source versus the antenna impedance. Figure 3b shows the corresponding impedance loss.

> **Example 3:** What is the reflection coefficient, VSWR and matching loss for a 50-ohm transceiver driving a half-wave dipole?
>
> The half-wave dipole has a characteristic impedance of 73 ohms, so the reflection coefficient from (7) is $\pm$0.187, depending upon the assignment of Z_1 and Z_0. The corresponding VSWR from (9) is 1.46. Applying (8), the matching loss is found to be 0.965, or -0.15 dB, which is negligible in most applications. _



a.) VSWR and reflection coefficient versus antenna impedance



b.) Impedance mismatch loss as a function of VSWR

Figure 3 VSWR, reflection coefficient and VSWR loss for a 50 ohm source

ANTENNA RADIATION REGIONS

The antenna radiation field is divided into three distinct regions, where the characteristics of the radiated wave are different. While these are not firm boundaries, they represent convention usage and provide some insight to the actual radiated field as a function of distance from the antenna. The *far-field* (Fraunhoffer) region is defined for distances such that

$$d > \frac{2D^2}{\lambda} \tag{10}$$

where *D* is the largest linear dimension of the antenna, and λ is the wavelength. This is the region where the wavefront becomes approximately planar (less than 22.5 degrees of phase curvature over a similar-sized aperture) [12]. In the far-field region, the gain of the antenna is a function only of angle (i.e., the antenna pattern is completely formed and does not vary with distance). In the far field, the electric and magnetic field vectors are orthogonal to each other. For electrically small antennas ($D << \lambda$) the far-field boundary defined by the preceding equation may actually fall within the near-field region. In such cases, the far-field boundary should be taken as equal to the near-field boundary rather than within it. The *radiating near-field* is the region between the far-field region and the reactive near-field region, also called the *transition region*.

$$\frac{\lambda}{2\pi} < r < \frac{2D^2}{\lambda} \tag{11}$$

In this region, the antenna pattern is taking shape but is not fully formed. Thus the antenna gain will vary with distance even at a fixed angle. The radiated wave front is still clearly curved (nonplanar) in this region and the electrical and magnetic field vectors are not orthogonal. The *reactive near-field* is defined as

$$r < \frac{\lambda}{2\pi} \tag{12}$$

This region is measured from the phase center or center of radiation of the antenna and will be very close to the surface of the antenna. In general, objects within this region will result in coupling with the antenna and distortion of the ultimate far-field antenna pattern. Figure 4 is an illustration of the radiation regions for a reflector antenna.

Analysis of near-field coupling can be deceiving. It is unlikely that anything will be close enough to a 40-GHz antenna to cause near-field coupling since the wavelength is less than a centimeter, making the reactive near-field boundary on the order of a couple of millimeters. On the other hand, it is very difficult to avoid coupling with an HF antenna for amateur radio. Antennas for the 75-m amateur radio band (4 MHz) have a reactive near-field boundary of 12 meters. Thus any large conductor (relative to the wavelength) within about 39 ft of the antenna will couple with the antenna and "detune" it. The result can be an altered resonant frequency, radiation resistance (defined shortly), and/or radiation pattern. This can be a significant concern for radio amateurs operating from small lots, where overhead utility lines, house gutters, fascia trim, and other such items are difficult to avoid. Tying these conductors to a good earth ground eliminates their resonant effect and mitigates their impact.

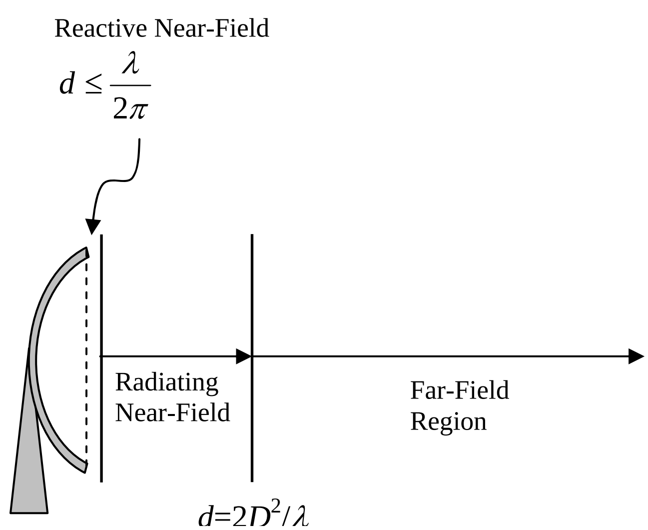


Figure 4 Typical Boundaries for Antenna Radiation Regions

Example 4: How much separation is required between a 140-MHz quarterwave monopole antenna and a 450-MHz quarter-wave monopole antenna if both are mounted on the roof of an automobile and near-field coupling is to be avoided?

To avoid mutual, near-field coupling, each antenna must be outside of the reactive near-field of the other. The reactive near-field of the 140-MHz antenna is larger than that of the 440-MHz antenna, so it determines the minimum required separation. Using (12),

$$d_{\min} = \frac{\lambda}{2\pi}$$

where

$$\lambda = 2.14 \text{ m}$$

The result is that the minimum required separation is

$$d_{min} = 0.341\text{m}$$

SOME COMMON ANTENNAS

The possible configurations of an antenna are limited only by the imagination. In the field of electromagnetic compatibility, any conductor is a potential radiator and may be treated as an antenna. In this section we examine several broad categories of widely used antennas. The reader is directed to the references for more detailed information on these antennas or for insight into other antenna configurations.

The Dipole

The half-wave dipole is configured as shown in Figure 5. It usually consists of two segments, each one-quarter of a wavelength long, with the feed in the middle, although offset-feed and end-fed dipoles are also used. The rings in Figure 4 illustrate the directionality of the radiation pattern of the dipole. It can be seen that there are areas of reduced gain off the ends of the elements. The gain of a half-wave dipole is theoretically 2.14dB. In general, antenna gain may be specified in decibels relative to an isotropic radiator, dBi, or in decibels relative to an ideal half-wave dipole, dBd. Thus a gain in dBd is always 2.14 dB less than the same value in dBi.

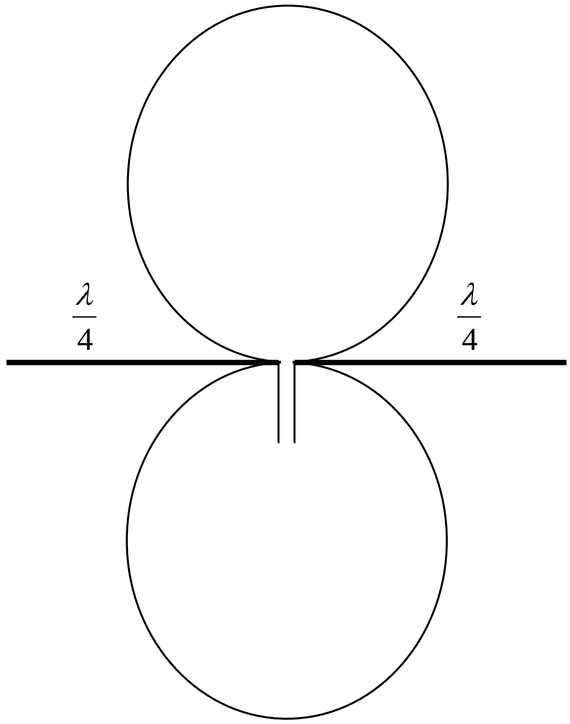


Figure 5 The half-wave dipole antenna and its radiation pattern.

The quarter-wave vertical or monopole antenna is one-quarter of a wavelength long and is often (or at least ideally) mounted on an infinite, perfectly reflective (conductive) ground plane. The ground plane provides a reflection of the antenna, causing it to behave as a dipole antenna. Oftentimes in practical use, the “ground plane” turns out to be more of a lossy counterpoise (such as the transceiver case and the hand and arm holding it), rather than a ground plane that can provide an electrical image of the antenna. Even then, the quarter-wave vertical antenna is still relatively effective. The counterpoise can be thought of as a nonideal implementation of the other half of a half-wave dipole.

The quarter-wave antenna may also be physically shortened as long as the electrical length of $\lambda/4$ is maintained. An example of such an antenna is the flexible rubber duck or helix antenna [13] used on many handheld VHF and UHF radios. While convenient, these antennas can have considerably less gain than a regular quarter-wave antenna, on the order of −3dB or more. All quarter-wave antennas perform best when they have a large conductive ground plane. Using the human hand and arm as a lossy counterpoise degrades both the radiation pattern and the gain. If the antenna is held near the body such as next to the head or on the waist, the net gain may be further reduced by 10dB or more.

An important parameter in antenna performance is the radiation resistance. The radiation resistance of a resonant (half-wave) dipole is approximately 73 ohms, if it is center-fed. For an ideal antenna (no resistive losses), the driving-point impedance seen by the transmitter or receiver is equal to the radiation resistance. The same dipole can be end-fed, but the driving-point impedance then becomes extremely large, on the order of 3000 ohms. A matching network of some type is then required to match the radiation resistance to the feedline and receiver/transmitter impedance to avoid large reflections and SWR (standing-wave ratio, also called voltage standing wave ratio or VSWR). Such matching networks (transmatch) can be somewhat lossy and reduce the efficiency of the antenna.

It is not unusual to see a “loading coil” at the base of an antenna, particularly for mobile applications. This loading coil can be either an impedance matching device if the radiator is an end-fed dipole, or it may be an inductor that is used to increase the electrical length of the antenna. An example is the “Hamstick” antenna used for HF work in automobiles. The antenna is a 4-foot piece of fiberglass with windings running the entire length, forming what is called a normal mode helix. At the top is a 3 to 4-ft whip that is adjusted in length to tune the resonant frequency of the antenna. The body of the car serves as a counterpoise, and the 8-ft antenna is then resonant at an HF frequency such as 3.9 MHz, giving it a resonant length of 75 m. This is shown conceptually in Figure 6. Though the antenna is resonant, it does not have the same gain as a full-length dipole and in fact its gain considerably less than unity.

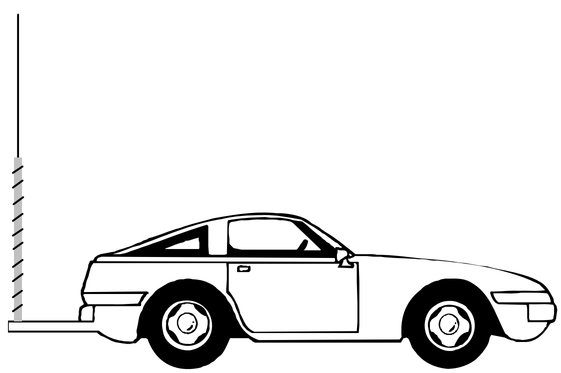


Figure 6 Base-Loaded Mobile HF Antenna

It is also possible to locate “traps” (tank or resonant circuits) along the length of a radiator, so that the same vertical antenna can be resonant at several spot frequencies. Such traps are resonant lumped elements. A trap alters the electrical length of the antenna in one of two ways. It may be used to cut off frequencies above its resonant frequency, so that the remainder of the antenna is invisible above the resonant frequency. The other approach is to have the trap resonance set in between two frequencies of interest, so that at the higher frequency it becomes a capacitive load, effectively shortening the electrical length, and at the lower frequency it becomes an inductive load, thereby increasing the electrical length. Another architecture for the multiband vertical antenna is a parallel array of closely spaced vertical radiators, each tuned to a different band. The close proximity of the radiators means that they have strong interaction, which alters the electrical length of the radiators, so they do not physically measure a quarter or a half-wavelength. The parallel radiators are functioning as distributed elements. Both of these antenna architectures are illustrated in Figure 7.

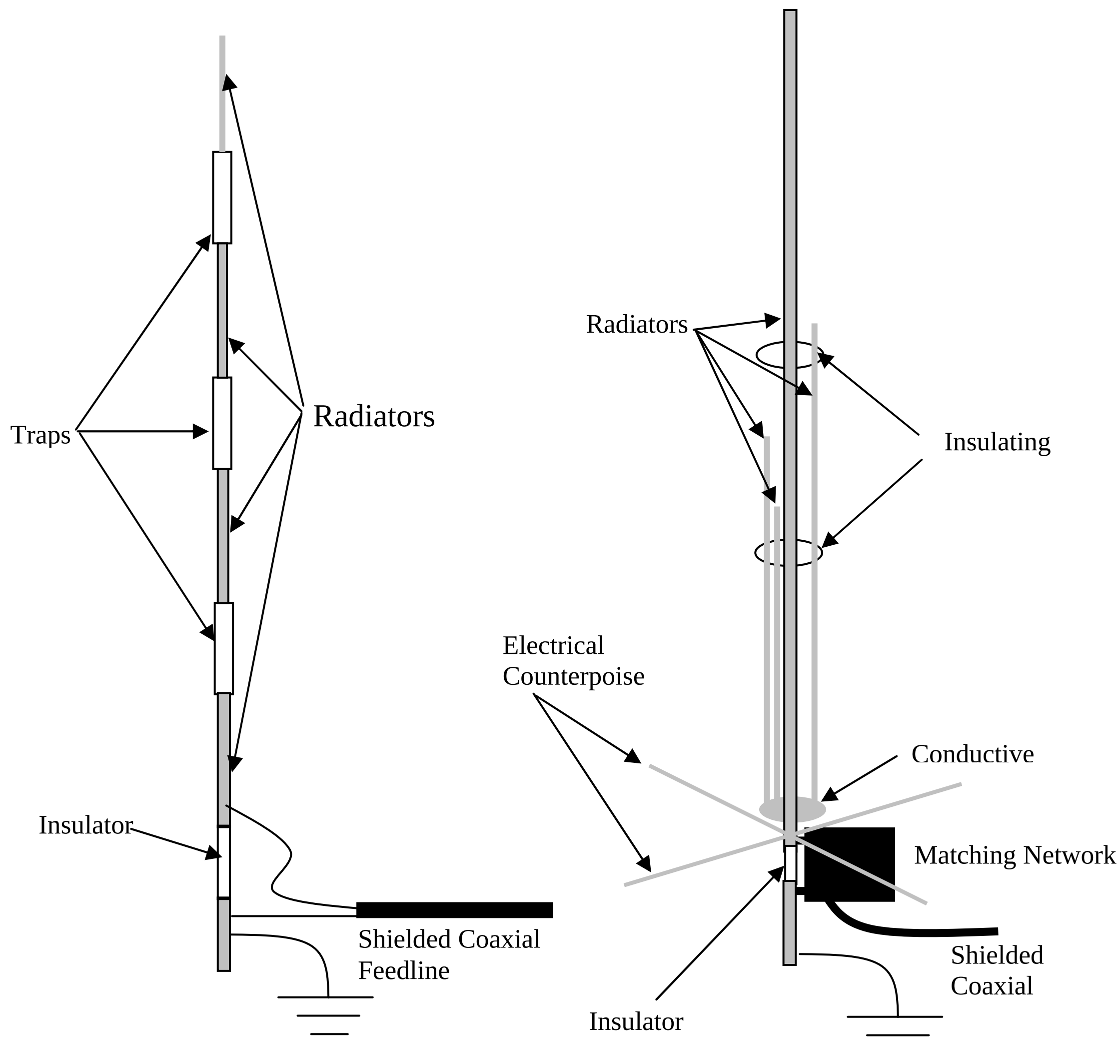


Figure 7 Four-Band Quarter-Wave Trap Vertical and a Four Band Half-Wave Parallel Radiator Vertical Antenna

Beam Antennas

The half-wave dipole antenna also serves as a component of another popular antenna, the Yagi-Uda (Yagi) beam antenna. The Yagi antenna is comprised of a driven element, which is a center-fed half-wave dipole, a reflector element, which is slightly longer than the driven element, and (optionally) several director elements that are progressively smaller than the driven element. This is shown in Figure 8.The boom may be conductive or nonconductive, with the element lengths and antenna properties being slightly different in each case. The Yagi beam is a directional antenna that generates a linearly polarized wave. The gain and front-to-back ratio can be substantial if a long boom and a sufficient number of elements are used in the design. Section 5.4 in Ref. 14 provides an excellent treatment of Yagi antenna design and analysis.

When mounted with the elements oriented vertically, the polarization will be vertical and vice versa. Conventional television broadcast signals are horizontally polarized, which is why television beam antennas are mounted with the elements oriented horizontally. Horizontal polarization was chosen for television broadcast because man-made interference tends to be vertically polarized. By using horizontal polarization, television receivers have less interference to contend with. Mobile radio uses vertical polarization primarily for convenience because it is difficult to mount a horizontally polarized antenna on a automobile or a handset, but also because of its reflective properties.

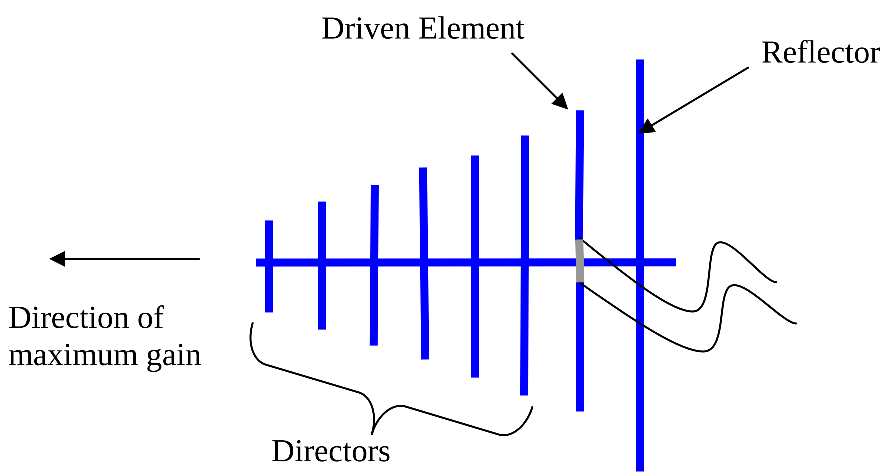


Figure 8 A typical Yagi-Uda antenna (the taper in the element lengths has been exaggerated in this picture for clarity).

Horn Antennas

A horn antenna can be thought of as a flared end of a waveguide. A horn antenna is one example of an aperture antenna. The polarization of the emitted signal is dependent upon the polarization of the waveguide transducer. There are various shapes of horn antennas, and the taper of the flare and the aperture size control the resulting radiation pattern. Horn antennas can provide very high gain and narrow beams, depending upon their design. They are frequently used as gain or calibration standards since their gains are very predictable and repeatable.

Reflector Antennas

Reflector antennas include the parabolic reflector or dish antenna and the Cassegrain antenna. The parabolic reflector has a receiver/transmitter mounted in a small unit that is usually located at the focal point of the parabolic dish. In addition to the efficiency loss due to illumination taper, the supports for the antenna feed produce blockage in the antenna field of view. This causes shadowing and produces diffraction lobes in the antenna pattern. An offset feed can be used to reduce the shadowing effect of the feed supports. An example of a parabolic dish antenna with an offset feed is the direct broadcast TV antenna as shown in Figure 9. The antenna feed (actually it is only a receive antenna) contains a low-noise amplifier and a broadband (block) down-converter that shifts all of the signals to L-band. The feed unit is therefore referred to as a low-noise block (LNB).The block of L-band signals then travels through the coaxial cable to the receiver, which selects the desired frequency for demodulation. It is interesting that Direct TV broadcasts different programming simultaneously on right and left circular polarizations. The Direct TV receiver sends a DC voltage over the receive signal coax to power the electronics in the LNB. The level of the DC also tells the block converter which polarization to receive based on the operator's channel selection at the receiver. This is why Direct TV receivers cannot directly share an antenna connection. For two receivers, a dual LNB is employed, with a separate cable running to each receiver. For more than two receivers, a dual LNB and a receiver control unit (multiswitch) are used to provide the appropriately polarized signal to each receiver, depending upon the channel being viewed. This is an example of frequency sharing by using polarization diversity.

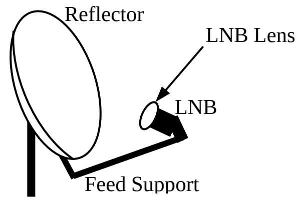


Figure 9 Offset feed reflector antenna for direct broadcast satellite TV.

For some applications, the receiver/transmitter unit is mounted behind the antenna with waveguide or coax cable connecting it to the center-mounted feed. Of course, this configuration entails cable or waveguide loss as the signal is fed to and from the feed point, so it is not suitable for most higher-frequency applications. One way to eliminate this additional loss is to employ a Cassegrain antenna, which is shown diagrammatically in Figure 10. The Cassegrain antenna also employs a parabolic reflector, but it has the antenna feed mounted in the center of the dish, where the signal radiates out to a subreflector that is mounted at the focal point of the parabolic reflector. The signal from the reflector then illuminates the dish and produces the radiation pattern. One advantage of the Cassegrain antenna over the basic parabolic reflector is that the receiver/transmitter unit can be mounted on the antenna without using cables or waveguides and are not required to be as small as those for a standard reflector antenna. Another advantage for very sensitive systems used on earth stations for satellite communication is that the sidelobes from the feed see sky rather than earth (since the feed is pointed upward), which reduces the antenna temperature and thereby the receiver noise floor. The center reflector still requires support, so the Cassegrain antenna also experiences blockage and diffraction lobes both from the subreflector supports and the subreflector itself. There are many different ways that the supports can be mounted, Figure 9 shows one possibility.

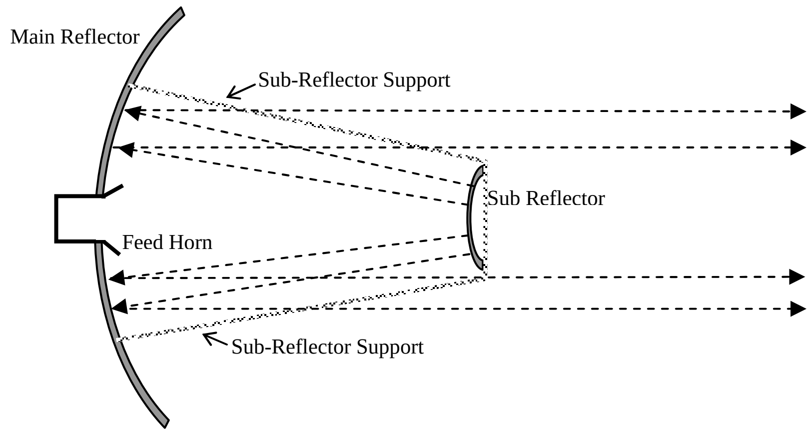


Figure 10 Cassegrain antenna configuration.

Phased Arrays

Phased array antennas were initially developed for large ground-based tracking radars. They provide the ability to focus a beam on each target in the field of view in rapid succession without requiring rapid repositioning of the aperture. The phased array is an array of radiating elements each with a controllable phase shifter. The antenna controller adjusts the phase (and sometimes also the gain) of each element to provide the desired beam position and beamwidth while controlling the sidelobes. Some systems even produce multiple beams simultaneously. Phased arrays suffer from a phenomenon called grating lobes when the look angle approaches 90 degrees off-boresight, which is why phased arrays sometimes include physical pointing. The beamforming process uses digital signal processing techniques in the spatial domain to form beams and control sidelobes. Squinted beams can be formed to permit monopulse angle measurement. As electronics have become increasingly powerful, efficient, and affordable, phased arrays have found application in many other areas. So-called smart antennas [15] have been used for GPS receivers and mobile telephone base stations for interference control. The advantage of the phased array is that it can steer a null in the direction of any detected interference, improving receiver operation. This is often accomplished using an adaptive algorithm to optimize signal-to-interference ratio.

Other Antennas

There are a multitude of antenna types, all of which cannot be even listed here. Some of the more noteworthy configurations are discussed. One interesting antenna is the magnetic field antenna. Most antennas are electric field antennas; they work by intercepting the electrical field component of the desired wave and converting it to a voltage that is sent to a receiver. It is equally possible to detect the magnetic field rather than the electric field, and it is sometimes advantageous to do so. One good example of a magnetic field antenna is the ferrite loop antenna in a small AM radio. The wavelength of AM broadcast is so long that electric field antennas are difficult to fabricate with any efficiency. Early AM radios required very long wire antennas. Later, as the receiver became more sensitive, moderately sized loop antennas where employed. Modern AM receivers use a ferrite rod with windings to sense the magnetic field of the AM wave and provide a signal to the receiver.

Lens antennas operate just as the name suggests. They are treated as an aperture antenna, but a lens covers the aperture and focuses the beam. The lenses are often made of plastic, but other materials can be used as long as they have the appropriate refractive properties at the frequency of interest.

Patch antennas are conformal elements, which are useful for applications where a low profile is required. An array of patch radiators can be used to form a beam. Such antennas are useful for mobile and particularly airborne applications, where low profile equates to less wind resistance and lower fuel costs.

ANTENNA POLARIZATION

Antenna polarization is defined as the polarization of the electromagnetic wave that it radiates. Polarization is defined in the far-field of the antenna and is the orientation of the plane that contains the electric field portion of the wave. In the far-field, the electrical field, magnetic field, and propagation direction vectors are all mutually orthogonal. This mutual orthogonality permits complete characterization of a wave by describing the electric field vector and the direction of propagation. Most antennas and electromagnetic waves have either linear polarization (vertical or horizontal) or circular polarization, which can be right-hand or left-hand. In a general sense, all polarizations can be considered as special cases of elliptical polarization. This provides a convenient mathematical framework for looking at the effects of polarization. An elliptically polarized wave can be expressed as [16]

$$\boldsymbol{E} = \boldsymbol{x}E_x + \boldsymbol{y}E_y$$

where

$$E_x = E_1 \sin(\omega t - \beta z)$$

And

$$E_y = E_2 \sin(\omega t - \beta z + \delta)$$

and

are the **x** and **y** components of the **E** field vector. **E** is the vector that describes the electric field as a function of time and position along the **z** axis (direction of propagation). E_1 and E_2 are the amplitude of the linearly polarized wave in the **x** and **y** directions. The δ term describes the relative phase between the **x** and **y** components of the electric field vector. If δ is equal to zero, ***E*** will be linearly polarized, with the orientation of its electrical field determined by $\mathbf{x}E_1 + \mathbf{y}E_2$. If E_1 and E_2 are of equal magnitude and $\delta = \pm 90$ degrees, then the wave is circularly polarized. Circular polarizations, and elliptical polarizations in general, are defined as follows (Figure 11):

Right-Hand Circular Polarization. Clockwise rotation when viewed in the direction of propagation—that is, from behind (**z** into page).

Left-Hand Circular Polarization. Counterclockwise rotation when viewed from behind (**z** into page).

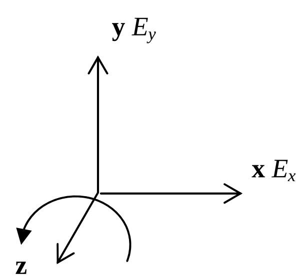


Figure 11 Relationship between orthogonal components of the **E** field and the direction of propagation for right circular polarization ($\delta = -90$ degrees).

These definitions correspond to the IEEE definitions [17, 18]. The labels are frequently shortened to right-circular and left-circular polarization. Using the definitions and the preceding equation, it is straightforward to verify that Figure 12 shows the polarization ellipse as presented by Ref. 16. Note that E_1 and E_2 are the peak amplitudes in the **x** and **y** direction and are therefore fixed, whereas E_x and E_y are the instantaneous amplitudes of the **x** and **y** components, which are functions of time, *t*, position, *z*, and the relative phase, δ. This diagram makes it easy to visualize the following: If E_2 is zero (or if δ is zero), the polarization is linear horizontal; and if E_1 is zero, it is vertical polarization. The tilt angle, τ, is the orientation of the major axis of the polarization ellipse relative to the **x** axis.

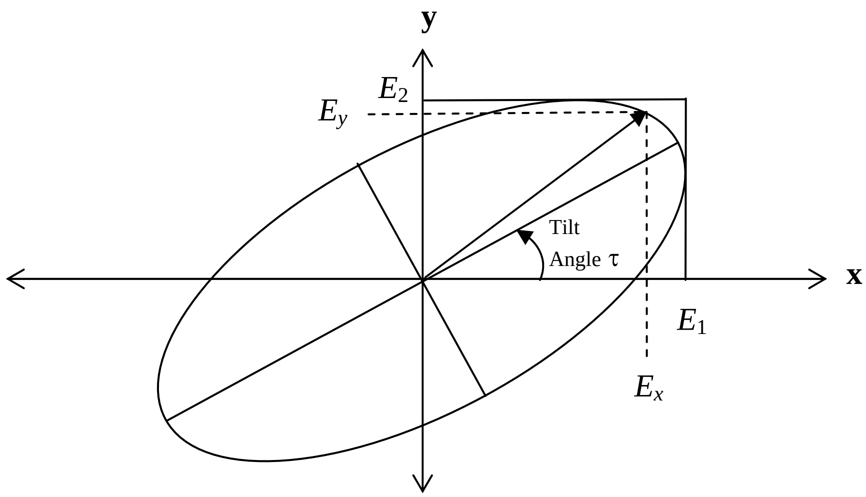


Figure 12 The polarization ellipse (Figure 2.22 from Ref. 16, courtesy of McGraw-Hill.)

The *axial ratio* is an important antenna parameter that describes the shape of the polarization ellipse. The axial ratio is defined as the ratio of the major axis to minor axis of the polarization ellipse when the phase angle between the linear polarization components, δ, is ± 90 degrees. The axial ratio is always greater than or equal to one. Since it is a ratio of amplitudes, it can be expressed in dB using 20 log(*AR*). For precise applications such as satellite communication, the axial ratio is given for circularly polarized antennas, as a metric of the antenna's deviation from ideal circular polarization. This is valuable for determining the potential cross-coupling between an incident wave and the orthogonal polarization channel of the receive antenna and the coupling loss between an incident wave and the receive antenna at the same polarization. Strictly speaking, if the axial ratio is anything other than 0 dB, the antenna is elliptically polarized, but it is common to refer to real-world antennas as circularly polarized with an axial ratio slightly greater than 0 dB.

When an antenna polarization is orthogonal to that of an incident wave, it will theoretically not receive any power from the incident wave. Examples of this are a vertically polarized wave incident on a horizontally polarized antenna or the Direct TV example of RCP and LCP waves being simultaneously incident on the antenna and only one or the other is actually sent to the receiver. Of course, real-world antennas and waves are not perfect, so it is valuable to be able to determine how much energy from an incident wave is coupled to the antenna. In the cross-polarization case, this is characterized by the cross-polarization discrimination (often denoted XPD*)*. In the co-polarization scenario, if the polarizations do not match exactly, not all of the incident energy is coupled to the antenna. This is characterized by the polarization loss factor, *F*. Both of these parameters can be determined from the orientation and axial ratio of the wave and the antenna.

Cross-Polarization Discrimination

In some systems, orthogonal polarizations are used to provide two channels in the same frequency band. This may happen for example on a satellite communication system where both left and right circular polarizations are transmitted to the ground at the same frequency, and each earth station must receive the appropriate polarization. Any cross-polarization leakage is considered co-channel interference. Cross-polarization discrimination or isolation can be computed for either linear or circular polarizations. For fixed terrestrial links, it is possible to employ polarization diversity to increase frequency reuse, since the orientation of the polarization vectors can be carefully controlled. For satellite communication systems, polarization diversity requires the use of circular polarization since absolute control of the polarization vector orientation is difficult, due to the geometry and the potential for Faraday rotation in the ionosphere.

An important distinction that must be made is that technically the crosspolarization discrimination is between the incident wave and the receive antenna, not between the transmit and receive antennas [17]. The distinction is due to the fact that there may be environmental effects on the transmitted wave prior to reception. In most practical applications, however, the parameters of the transmit antenna are used either directly or with some allowance for environmental effects. For the linear polarization case, the cross-polarization discrimination is a function of the angle difference between the incident wave polarization vector and the receive antenna polarization vector. This is referred to as the tilt angle and often designated as τ. The amount of cross-polarization discrimination is then given by

$$XPD = \sin^2(\tau)$$

Cross-polarization discrimination for linear polarization can be thought of as a special case of circular or elliptical polarization. For elliptical polarization, the cross-polarization discrimination is a function of the tilt angle τ and of the axial ratios of the incident wave and the receive antenna. For linear polarization, the axial ratio is ideally infinite since the minor axis of the ellipse is zero. Thus linear polarization and circular polarization can both be viewed as special cases of elliptical polarization.

Polarization Loss Factor

The polarization loss factor is the multiplier that dictates what portion of the incident power is coupled into the receive antenna. The polarization loss factor will often be less than one because of polarization mismatch. The received power is given by

$$P_r = FP_i$$

where P_i is the incident power and P_r is the power actually available at the antenna. The polarization loss factor is the complement of cross-polarization discrimination. Just as with XPD, the polarization loss factor for linear polarization is a degenerate case of the elliptical polarization case, where the polarization ellipse collapses into a line. For linear polarization,

$$F = \cos^2(\tau)$$

where τ is the angle between the wave polarization and the receive antenna polarization. F is a power ratio, so it is converted to dB using 10 log(F). It is apparent that when τ is zero, $F = 1$ and all of the incident power is received. When τ is 90 degrees the polarization of the wave and the receive antenna are orthogonal and no power is transferred. From the expressions for XPD and F for linear polarization, it is clear that

$$XPD = 1 - F$$

as expected based on conservation of energy.

The general development of polarization loss factor is based on elliptical polarization. Kraus [19] provides a development of the polarization loss factor as a function of the axial ratios and tilt angle using the Poncaire' sphere, the results of which are summarized here. The coordinates of the polarization vector on the sphere are

$$\text{Longitude} = 2\tau$$

$$\text{Latitude} = 2\varepsilon$$

where τ is the tilt angle as defined earlier and

$$\varepsilon = \tan^{-1}(k/AR)$$

with

k = 1 for left-hand polarization
k = −1 for right-hand polarization
AR = the axial ratio

The great-circle angle to the polarization vector is given by 2γ, where

$$\gamma = \frac{1}{2}\cos^{-1}\left(\cos(2\varepsilon)\cos(2\tau)\right)$$

and the equator-to-great-circle angle is δ as previously defined. The polarization loss factor is defined as

$$F = \cos^2(\gamma_r - \gamma_w)$$

where $2\gamma_r$ and $2\gamma_w$ are the great-circle angles to the receive antenna polarization vector and the wave polarization vector, respectively.

$$\gamma_r = \frac{1}{2}\cos^{-1}\left(\cos(2\varepsilon_r)\cos(2\tau_r)\right)$$

$$\gamma_w = \frac{1}{2}\cos^{-1}\left(\cos(2\varepsilon_w)\cos(2\tau_w)\right)$$

It is common practice to let either τ_r or τ_w be zero and the other term then accounts for the net tilt angle between the wave and the antenna. Making the appropriate substitutions, it is possible to derive an expression for polarization loss as a function of γ_r, γ_w, and τ:

$$F = \cos^2\left\{\tan^{-1}\left(\frac{k}{AR_w}\right) - \frac{1}{2}\cos^{-1}\left[\cos\left(2\tan^{-1}\left(\frac{k}{AR_r}\right)\right)\cos\left(2[\tau_w - \tau_r]\right)\right]\right\} \tag{13}$$

An equivalent and perhaps easier-to-use formulation is given in Ref. 17:

$$F = \frac{\left(1 + AR_w^{\ 2}\right)\left(1 + AR_r^{\ 2}\right) + 4AR_wAR_r + \left(1 - AR_w^{\ 2}\right)\left(1 - AR_r^{\ 2}\right)\cos\left(2[\tau_w - \tau_r]\right)}{2\left(1 + AR_w^{\ 2}\right)\left(1 + AR_r^{\ 2}\right)} \tag{14}$$

In each case, the polarization loss factor applies to the received power. Some interesting observations can be made from the preceding expression. If both the incident wave and the antenna are circular ($AR = 1$), then F defaults to unity as expected, regardless of the tilt angle. On the other hand, if one of the axial ratios is unity and the other is greater than one, there will be some polarization loss regardless of the tilt angle. If both the incident wave and the antenna have identical axial ratios (greater than one), then the polarization loss factor becomes a function of the relative tilt angle only and complete power transfer occurs only when $\tau = 0$ or 180 degrees. The minimum and maximum polarization losses occur at 0/180 and 90/270 degrees, respectively.

For the purpose of link budgeting, the worst-case polarization loss factor is employed. The worst case occurs when $\tau = 90$ degrees. Figure 13 (a) is a plot of the polarization loss factor in dB for a tilt angle of 90 degrees, as a function of incident axial ratio for several different antenna axial ratios. By selecting a curve with the receive antenna axial ratio and finding the wave (transmit) axial ratio along the x axis or vice versa, the polarization loss factor can be read from the y axis of the plot.

In some circumstances, a circularly polarized antenna might be used to receive a linearly polarized wave or vice versa. This may occur when an antenna serves multiple purposes, or if the orientation of the linear polarization is unknown, then use of a circularly polarized receive antenna might be considered. For instance, the orientation of a linearly polarized antenna on an aircraft will vary with the aircraft's attitude. In this case, using a circularly polarized antenna at the other end of the communication link would eliminate the variability in signal strength at the expense of taking a constant loss in signal strength of about 3 dB. The polarization loss factor between linear and circular polarization is generally assumed to result in a 3-dB polarization loss factor, regardless of the orientation of the linearly polarized wave. However, if the circularly polarized antenna is not ideal (i.e., axial ratio greater than 1), then the actual polarization loss factor is a function of the tilt angle.

In this case, the polarization loss factor will be greater than or less than 3 dB depending upon the orientation and the axial ratio. Figure 13 (b) shows the polarization loss factor for larger values of axial ratio along the x-axis. Since a linearly polarized antenna can be represented by an elliptical polarization with a large axial ratio, the far right-hand side of Figure 13 (b) shows how the worst-case polarization loss levels-off when one of the antennas tends toward linear polarization. From this plot, it can be seen that if the circularly polarized antenna (or wave) has an ideal axial ratio of 0 dB, then the worst-case polarization loss is 3 dB. Otherwise, if the circularly polarized antenna (or wave) is not ideal, then the actual circular-to-linear polarization loss factor exceeds 3 dB as shown. A more precise estimate of the polarization loss factor can be obtained by taking the limit of equation (13) as either AR_r or AR_w go to infinity, if desired.

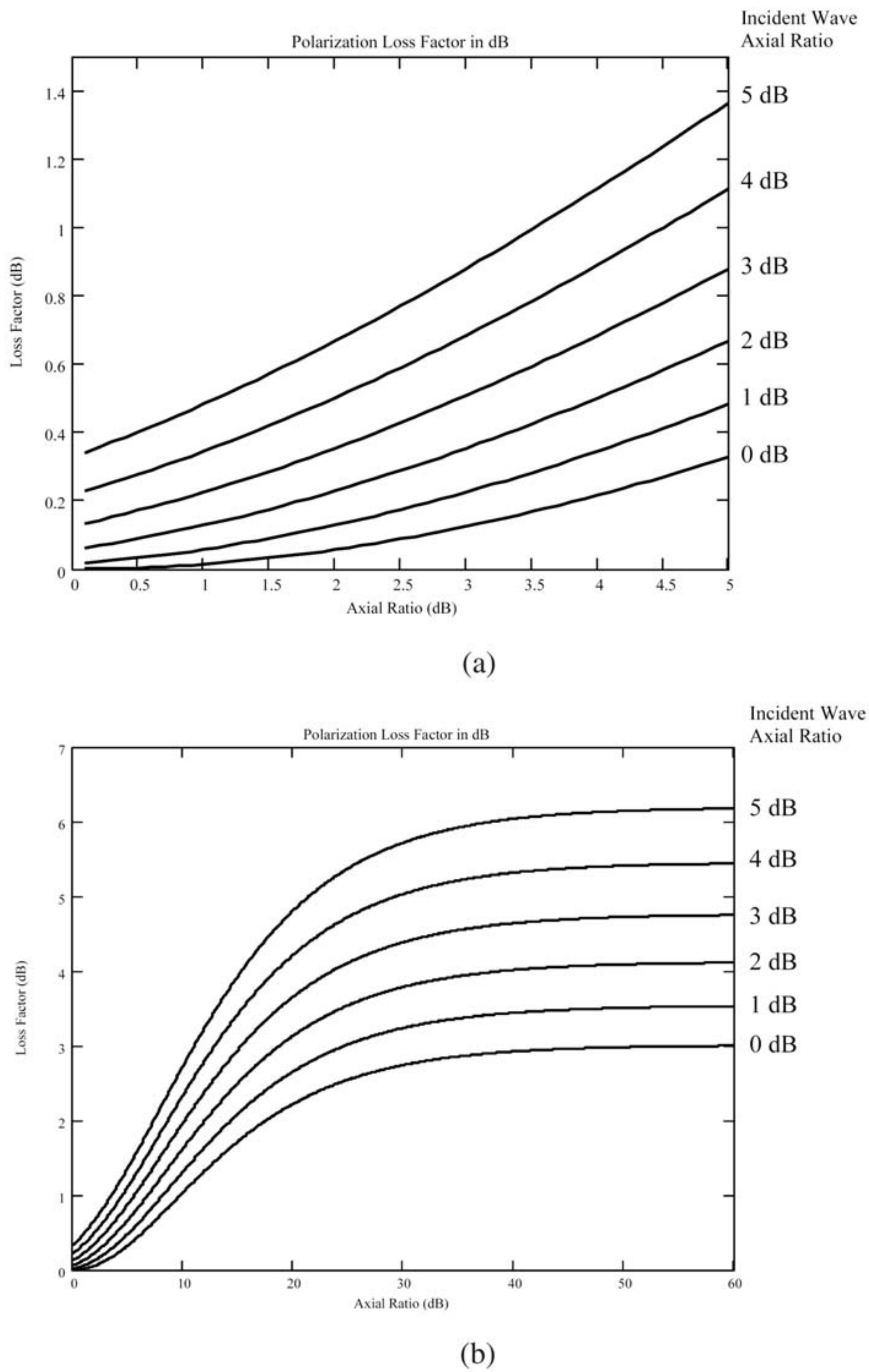


Figure 13 Worst Case Polarization Loss Factor Versus Axial Ratio for CP

Example 5: A desired satellite signal is transmitted with right "circular" polarization from an antenna having a 2-dB axial ratio. If the ground station receive antenna has an axial ratio of 3dB, what are the minimum and maximum polarization loss that can be expected if environmental effects are negligible? If another signal is simultaneously transmitted from the same satellite antenna at the same frequency, but with left circular polarization, what is the worst-case cross-polarization isolation that will occur?

The maximum polarization loss factor can be read from Figure 13 (a) and is

$$F_{max} = -0.35 \text{ dB}$$

The minimum can be found by setting τ equal to 0 or 180 degrees and evaluating either of the expressions for axial ratio. The resulting minimum polarization loss factor is

$$F_{min} = -0.01 \text{ dB}$$

which is negligible. The cross-polarization isolation can be determined by using one minus the worst-case polarization loss factor:

$$\text{XPD}_{min} = 10\log\left(1 - 10^{F/10}\right) = -11.1 \text{ dB}$$

Thus the interfering cross-polarized signal will be only 11 dB below the desired signal. _

ANTENNA POINTING LOSS

In applications where directional antennas are used, it is customary to allow for some antenna misalignment loss when computing the link budget. Antenna misalignment means that the received signal is not received at the peak of the antenna beam and/or the portion of transmit signal that is received did not come from the peak of the transmit antenna beam. This is shown pictorially in Figure 14. While the potential for misalignment is intuitive for mobile systems or systems tracking mobile satellites, it is also an issue for static links. Pointing loss can occur in several ways. The initial antenna alignment is seldom perfect; on a point-to-point link, each end must be aligned. In addition, effects such as wind, age, and thermal effects may all contribute to changing the antenna alignment over time. Variation in atmospheric refraction can produce time-varying pointing error, which can be challenging to identify.

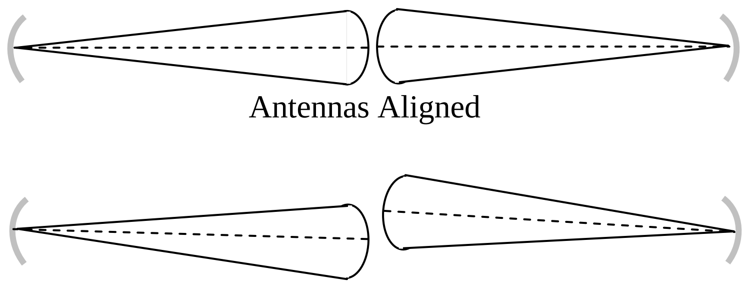


Figure 14 Antenna Alignment and Misalignment for a Point-To-Point Link Using Directional Antennas

For systems with active tracking on one end of the link, the tracking loss must take into account the track loop errors as well as any error introduced by the tracking process. A track loop must move the peak of the antenna beam about the actual track point to generate a tracking error signal for position correction. This process of scanning will result in a certain amount of signal loss (scan loss) that must be factored into the link budget. Most tracking loops require 0.5 to 1 dB of tracking signal variation for stable tracking.

SUMMARY

The antenna serves as the interface between system electronics and the propagation medium. The principle of reciprocity states that an antenna performs the same whether transmitting or receiving, which simplifies analysis and measurement. Antenna gain and directivity define the amplitude of the antenna pattern in the direction of maximum radiation (or reception). The gain of an antenna can be estimated using the concept of effective area for aperture antennas or effective height for linear antennas. Antenna gain, mainlobe width, and sidelobe levels are key concerns in antenna design or analysis.

The performance of an antenna varies with the distance from the antenna. The antenna's far-field, where the distance is greater than twice the square of the antennas largest linear dimension divided by the wavelength, is most often the region where the parameters are defined and specified. The reactive near-field is defined as the region where the distance is less than the wavelength divided by two pi. In the reactive near-field, the antenna pattern does not apply, and in fact the illumination function of the antenna will be present.

Objects in the reactive near-field will couple with the antenna and alter the far-field pattern. The radiating near-field is the region between the reactive near-field and the far-field. In the radiating near-field, the antenna pattern is starting to take shape, but the amplitude in a particular direction may still vary with distance from the antenna. In this region, the radiated wave front is still spherical, versus nearly planar in the far-field.

There are a wide variety of antenna types and development continues. For most applications, antennas are required to transmit/receive either linear (vertical or horizontal) or elliptical (right-hand or left-hand) polarization. Any mismatch between the received wave and the receive antenna may lead to signal loss and reduction in rejection of the orthogonal polarization at the receive side of the link. The polarization loss factor and cross-polarization discrimination are often included in link budgets, particularly for satellite links.

REFERENCES

1. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd ed., Wiley, Hoboken, NJ, 1998 pp. 404–409.

2. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975, pp. 239–250.

3. F. J. Harris, On the use of windows for harmonic analysis with the discrete fourier transform, *Proceedings of the IEEE*,Vol. 66, No. 1, January 1978, pp. 51–83.

4. M. I. Skolnik, *Introduction to Radar Systems*, 3rd ed., McGraw-Hill, New York, 2001, p. 541.

5. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd ed., Wiley, Hoboken, NJ, 1998, p. 77.

6. J. D. Kraus and R. J. Marhefka, *Antennas for All Applications*, 3rd ed., McGraw-Hill, New York, 2002, pp. 30–36.

7. J. D. Kraus and R. J. Marhefka, *Antennas for All Applications*, 3rd ed., McGraw-Hill, New York, 2002, p. 31.

8. C.A. Balanis, *Antenna Theory, Analysis and Design*, 2nd ed., Wiley, New York, 1997, p. 164.

9. M. I. Skolnik, *Introduction to Radar Systems*, 3rd ed., McGraw-Hill, New York, 2001, p. 625.

10. *The ARRL Handbook for Radio Amateurs*, ARRL, Newington, CT, 1994, Table 41, Chapter 35, p. 38.

11. *The ARRL Handbook for Radio Amateurs*, ARRL, Newington, CT, 1994, Figure 24, p. 16-5.

12. C.A. Balanis, *Antenna Theory, Analysis and Design*, 2nd ed., Wiley, New York, 1997, pp. 32–34.

13. K. Siwiak, *Radiowave Propagation and Antennas for Personal Communications*, 2nd ed., Artech House, Norwood, MA, 1998, pp. 352–353.

14. W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*, 2nd ed., Wiley, Hoboken, NJ, 1998, pp. 187–195.

15. J. Liberti, Jr. and T. S. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*, Prentice-Hall, Upper Saddle River, NJ, 1999.

16. J. D. Kraus and R. J. Marhefka, *Antennas for All Applications*, 3rd ed., McGraw-Hill, New York, 2002, p. 45.

17. J. S. Hollis, T. J. Lyon, and L. Clayton, *Microwave Antenna Measurements*, 2nd ed., Scientific Atlanta, Atlanta, GA, 1970, Chapter 3.

18. J. D. Kraus and R. J. Marhefka, *Antennas for All Applications*, 3rd ed., McGraw-Hill, New York, 2002, p. 46.

19. J. D. Kraus and R. J. Marhefka, *Antennas for All Applications*, 3rd ed., McGraw-Hill, New York, 2002, pp. 47–52.