1 Propagation in free space and the aperture antenna

This chapter introduces the basic concepts of radio signals travelling from one antenna to another. The aperture antenna is used initially to illustrate this, being the easiest concept to understand. The vital equations that underpin the day-to-day lives of propagation engineers are introduced. Although this chapter is introductory in nature, practical examples are covered. The approach adopted is to deliver the material, together with the most significant equations, in a simplified manner in the first two subsections before providing more detail. Following this, the focus is on developing methods of predicting the received signal power on point-to-point links given vital information such as path length, frequency, antenna sizes and transmit power.

1.1 Propagation in free space: simplified explanation

Radio waves travel from a source into the surrounding space at the 'speed of light' (approximately 3.0×10^8 metres per second) when in 'free space'. Literally, 'free space' should mean a vacuum, but clear air is a good approximation to this. We are interested in the power that can be transmitted from one antenna to another. Because there are lots of different antennas, it is necessary to define a reference with which others can be compared. The isotropic antenna in which the transmitted power is radiated equally in all directions is commonly used as a reference. It is possible to determine the ratio between the power received and that transmitted in linear units, but it is more common to quote it in decibels (dB). Further information on the decibel scale is given in an appendix at the back of this book. If we have an isotropic antenna as transmitter and receiver then the loss in dB is given by

$$\text{loss} = 32.4 + 20 \log d + 20 \log f, \tag{1.1}$$

where d is the path length in kilometres and f is the frequency in MHz. This loss in free space between two reference, isotropic, antennas is known as the 'free-space loss' or 'basic transmission loss'. The difference between the transmitted power and received power on any point-to-point system (the 'link loss') is the free-space loss less any antenna gains plus any miscellaneous losses. To maintain consistency, any antenna gains must be quoted relative to the reference isotropic antenna. Again it is normal to use the decibel scale and the gain is quoted in 'dBi', with the 'i' indicating that we are using an isotropic reference antenna:

$$\text{link loss} = 32.4 + 20\log d + 20\log f - G_t - G_r + L_m, \qquad (1.2)$$

where G_t and G_r are the gains of the transmitting and receiving antennas, respectively, in dBi and L_m represents any miscellaneous losses in the system (such as feeder or connector losses) in dB.

1.2 The aperture antenna: simplified explanation

The antenna forms the interface between the 'guided wave' (for example in a coaxial cable) and the electromagnetic wave propagating in free space. Antennas act in a similar manner irrespective of whether they are functioning as transmitters or receivers, and it is possible for an antenna to transmit and receive simultaneously. The simplest form of antenna to visualise is the aperture antenna. Parabolic dishes used for microwave communications or satellite Earth stations are good examples of aperture antennas. The gain of an aperture antenna increases with increasing antenna size and also increases with frequency. The gain of a circular parabolic dish type of aperture antenna is given by the approximation

$$\text{gain (dBi)} \approx 18 + 20\log D + 20\log f, \qquad (1.3)$$

where D is the diameter of the dish in metres and f is the frequency of operation in GHz. Thus the gain of an antenna will increase by 6 dB if it doubles in diameter or if the frequency of operation doubles. Aperture antennas must be accurately pointed because they generally have narrow beamwidths. The beamwidth is usually measured in degrees. A useful approximation is

$$\text{beamwidth} \approx \frac{22}{Df} \text{ degrees,} \tag{1.4}$$

where D is the diameter of the dish in metres and f is the frequency of operation in GHz. So, as the diameter increases, the beamwidth gets smaller and, similarly, as the frequency increases the beamwidth gets smaller.

The antenna is a reciprocal device, acting equally well as a transmitter as it does as a receiver. A high-quality transmitting antenna will radiate almost all the power entering its feed into the surrounding space. In a lower-quality antenna, a significant amount of this power will be dissipated as heat or reflected back along the feeder to the transmitter. An antenna that reflects a lot of the power back along the feeder is said to be 'poorly matched'.

1.3 Further details and calculations

Assuming that the transmitting antenna is perfect (that is, all the power entering from the feeder cable is radiated into space) makes the calculations simpler. Using linear units, the power entering the antenna from its feed, P_t, is measured in watts. Once it has left the antenna, it creates a power density, P_d, in space that is measured in watts per square metre. One very useful concept in radio wave propagation studies is known as the 'isotropic antenna'. This (fictional) antenna is thought of as radiating power equally in all directions. It is straightforward to predict the power density created by an isotropic antenna at a certain distance. To clarify this, we need to conduct a 'thought experiment'. Consider an isotropic antenna located at the centre of a sphere. Since the antenna radiates equally in all directions, the power density must be equal on all parts of the surface of the sphere. Further, since the sphere encloses the antenna, all of the power radiated by the antenna must pass through this sphere. Combining these two pieces of information, and knowing that the area of a sphere of radius r metres is equal to $4\pi r^2$ square metres, allows us to write the equation

$$P_d = \frac{P_t}{4\pi r^2}. \tag{1.5}$$

This equation reveals a very valuable generalisation in radio wave propagation: the 'inverse square' law. It can be seen that the power density produced by an antenna reduces with the square of the distance. This is true in free space and any deviation from this in 'real-world' environments is a matter of interest to radio-wave scientists.

We now turn our attention to the antenna as a receiver. It is a very common situation for a single antenna to transmit and receive simultaneously. Thus any practical antenna will possess attributes appropriate for transmission and reception of radio waves. When receiving, an antenna is illuminated by a particular power density and has the job of converting this into a received power in its feeder cable. Again a thought experiment is helpful. Consider a radio wave of a particular power density travelling through space. Now consider an aperture such that any power entering is passed into the feeder cable. The power entering the aperture (the received power, P_r) depends on the size of the aperture, A_e square metres (the suffix e taken to stand for 'effective' when referring to a receiving antenna), and the power density of the radio wave:

$$P_r = P_d A_e. \tag{1.6}$$

From our equation linking P_d to P_t we can now write

$$P_r = \frac{P_t A_e}{4\pi r^2}. \tag{1.7}$$

This gives us the power received at distance r metres by an antenna with effective aperture A_e square metres when an isotropic antenna transmits power P_t watts.

At this point we must turn our attention once more to the transmitting antenna. So far, we have considered this antenna to transmit power equally in all directions (isotropically). Practical antennas do not do this. The fact that antennas possess directivity also leads to the phenomenon known as antenna gain.

When most engineers think about a device possessing 'gain' they assume that the device outputs more signal power than it takes in. Antennas, as passive devices, cannot do this. The best they can do is to radiate all the power they receive from the feeder. When we talk about the gain of an antenna, we refer to the fact that the directivity leads to the power being concentrated in particular directions. Thus, if you enclosed the antenna in a sphere, centred at the antenna, you would not see equal power densities at all points on the sphere. The point on the sphere where the power density is greatest indicates the 'principal direction' of the antenna. The power density is given more generally by

$$P_{\mathrm{d}} = \frac{P_{\mathrm{t}} G_{\mathrm{t}}}{4\pi r^2}, \tag{1.8}$$

where G_{t} is the gain of the transmitting antenna in any direction. The principal direction, that of maximum gain (the direction in which the antenna is pointing), is the most usual direction to consider. Using our modified equation for the power density, we can modify the equation for the received power:

$$P_{\mathrm{r}} = \frac{P_{\mathrm{t}} G_{\mathrm{t}} A_{\mathrm{er}}}{4\pi r^2}, \tag{1.9}$$

where the effective aperture of the receiving antenna is now called A_{er} to make it clear that it is the receiving antenna that is being referred to.

Remember that the same antenna will transmit and receive. Thus the same antenna will possess gain as a transmitting antenna and an effective aperture as a receiving antenna. There is a link between the two, which we shall now investigate.

In the thought experiment that follows it is easiest if you allow yourself to imagine the antenna as a parabolic dish type of antenna such as those used for microwave communications or for satellite Earth stations. Such antennas are in fact known as 'aperture antennas' because their 'effective aperture' is very easy to picture; the reflecting dish itself is the aperture. It should be noted that wire antennas (such as those used for terrestrial television reception and on mobile devices) possess an effective aperture, even though this is less obvious from visual examination of the antenna. When an aperture antenna is used to transmit, power is directed towards the parabolic reflector from a feed placed at the focus. This power is then reflected back into space. Aperture antennas are known for producing narrow beams. The 'beamwidth' of an aperture antenna is influenced by the two following rules: the bigger the parabolic dish, the narrower the beam; the higher the frequency, the narrower the beam. It is this ability to focus, or collimate (collimate means 'form into a column' and describes

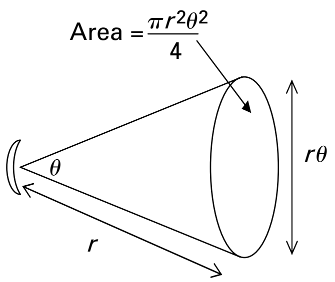

Figure 1.1 The area illuminated by an aperture antenna with a perfectly conical beam.

the way the energy travels without much spreading), the energy into a narrow beam that endows the antenna with gain.

Consider an idealised aperture antenna that radiates the energy in a perfectly conical beam of angle θ radians. If θ is small, a circle of diameter $r\theta$ would be evenly illuminated at distance r. The area of this circle is $\pi r^2\theta^2/4$ (see figure 1.1). Now, an isotropic antenna would illuminate a sphere of area $4\pi r^2$. The aperture antenna illuminates a smaller area. The same amount of power is concentrated into a smaller area. The reduction in area illuminated equals the increase in power density. It is this increase in power density that we call the gain of the antenna. Thus the gain of the antenna in this case is related to its beamwidth:

$$\begin{aligned} \text{gain} &= \frac{4\pi r^2}{\pi r^2\theta^2/4} \\ &= 16/\theta^2. \end{aligned} \tag{1.10}$$

Remember that the beamwidth has been idealised into a perfect cone shape and is not therefore exact for a practical antenna. It is not, however, a bad approximation. Thus, if we are told the beamwidth of an antenna, we can estimate its gain. It is more common to be told the beamwidth of the antenna in degrees.

Example: an antenna has a beamwidth of 3 degrees. Estimate its gain.

Answer:

$$3 \text{ degrees} = 3\times\pi \div 180 = 0.052 \text{ radians;}$$
$$\text{gain} = 16/0.052^2 = 5800.$$

It is common to express the gain in logarithmic units (decibels):

$$\text{gain in dBi} = 10\log_{10}(5800) = 37.7\,\text{dBi}.$$

The additional suffix ‘i’ indicates that the gain has been calculated with respect to an isotropic antenna.

Thus we can estimate that the gain of an antenna with a beamwidth of 3 degrees is 37.7 dBi. However, a further important question is ‘How big would this antenna be?’. In order to answer this question, we need to accept a crucial fact about the isotropic antenna. We need to know the effective aperture of an isotropic antenna. Throughout this book, equations are derived in as straightforward a manner as possible. However, the derivation of the effective aperture of an isotropic antenna, though crucial, is beyond the scope of this book and we must accept it as the proven work of others. The effective aperture of an isotropic antenna, A_{ei}, depends on the wavelength, λ, and is given by

$$A_{\mathrm{ei}} = \frac{\lambda^2}{4\pi}. \tag{1.11}$$

This is actually the area of a circle whose circumference equals λ.

It is more common to talk about the frequency of operation than the wavelength. The wavelength in metres can be converted into the frequency f in MHz by use of the equation

$$\lambda = \frac{300}{f\,(\mathrm{MHz})}\ \text{metres}. \tag{1.12}$$

Thus, if we use the units of square metres for the effective aperture, the effective aperture of an isotropic antenna is given by

$$A_{\mathrm{ei}} = \frac{300^2}{4\pi f^2} = \frac{7160}{f^2}\ \text{square metres}. \tag{1.13}$$

The gain of an antenna is numerically equal to its effective aperture expressed as a multiple of the effective aperture of an isotropic antenna at the frequency of interest. Therefore, knowledge of the effective aperture of an isotropic antenna at a particular frequency is crucial when it comes to determining the gain of an antenna.

A further thought experiment is necessary to clarify a vital feature about antennas: the radiation pattern, and hence directivity and gain, is exactly the same when the antenna is transmitting as when it is receiving. In order to convince yourself of this, imagine an antenna placed in a sealed chamber. Suppose that this antenna is connected to a matched load. This load both receives thermal noise gathered from the antenna and transfers thermal noise to the antenna. The same amount of power is transferred in both directions, thus maintaining equilibrium. If the radiation pattern as a transmitter differed from that as a receiver, the antenna would transfer energy from one part of the chamber to another in violation of the laws of thermodynamics. This fact means that it does not matter whether you consider an antenna as a transmitter or as a receiver; the radiation pattern is the same. If we can calculate its gain as a receiver, then we know its gain as a transmitter.

Let us consider the antenna that we previously calculated to have a gain of 5800. If it is to have this gain as a receiver, the effective aperture must be 5800 times that of an isotropic antenna. The actual aperture size in square metres depends on the frequency. At 5000 MHz, the effective aperture of an isotropic antenna is 2.86×10^{-4} square metres. An antenna with a gain of 5800 would have an effective aperture of 1.66 square metres. If it was a circular aperture, its diameter would be 1.45 metres. This indicates the size of parabolic dish that would be required in order to have a gain of 5800 (and a beamwidth of 3 degrees) at a frequency of 5 GHz. In fact, this would be for a perfect aperture antenna. Practical antennas have a smaller aperture than that calculated from the diameter of the dish. This reduction in aperture is accounted for by the term ‘aperture efficiency’, η, which lies between 0 and 1. The actual size of a circular reflecting parabolic dish of diameter D is $\pi D^2/4$. Thus, considering aperture efficiency, the effective aperture of any practical circular parabolic dish is given by

$$A_\mathrm{e} = \eta \frac{\pi D^2}{4}, \tag{1.14}$$

where D is the diameter of the parabolic dish. This is illustrated in figure 1.2.

If we repeated the calculations at a frequency of 10 GHz, we would find that the diameter of the antenna would be halved for the same gain and beamwidth. This is one main reason why microwave communication

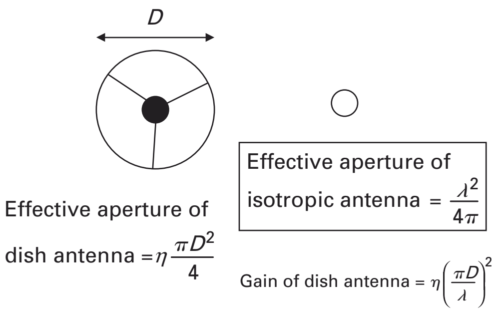

Figure 1.2 The effective aperture of an isotropic antenna and a practical aperture antenna.

has become popular: the higher the frequency, the easier it is to collimate the energy. Some general rules follow.

- If you double the frequency, the gain of an antenna will quadruple.
- If you double the frequency, the beamwidth of an antenna will halve.
- If you double the antenna diameter (keeping the frequency the same), the gain of the antenna will quadruple.
- If you double the antenna diameter (keeping the frequency the same), the beamwidth of an antenna will halve.

In our example, we took an antenna whose beamwidth was 3 degrees. At a frequency of 5 GHz, the required diameter was found to be 1.45 metres. If we multiply the frequency (in GHz) by the diameter (in metres) and by the beamwidth (in degrees) we find $5 \times 1.45 \times 3 \approx 22$. Because of the way in which beamwidth (BW) halves if frequency doubles and so on, we find that this formula is a good general approximation:

$$\mathrm{BW}\,(\text{degrees}) \times D\,(\mathrm{m}) \times f\,(\mathrm{GHz}) \approx 22. \tag{1.15}$$

The fact that the beamwidth reduces as the antenna diameter increases poses significant challenges to the radio-system engineer. The beamwidth gives an indication of how accurately the antenna must be pointed. Antennas on masts are subject to wind forces that will tend to affect the antenna direction. The larger the antenna, the larger the wind force that acts on it. The fact that larger antennas must also be pointed more accurately and cannot therefore be allowed to alter direction by as much as smaller antennas can leads to the necessity for careful consideration before larger antennas are selected.

1.4 Point-to-point transmission

With knowledge about the gain of antennas and the free-space loss between two points it is possible to predict the received signal power for a particular situation and, thence, to design a link to a deliver a particular power to the receiver. For example, suppose that we have a system consisting of two parabolic dish antennas 15 km apart. Each antenna has a diameter of 1 m. Ignoring any feeder or miscellaneous losses, we can estimate the link loss at a frequency of, say, 20 GHz. The gain of the antennas will each be approximately $18 + 20\log(1) + 20\log(20) = 44$ dBi. Therefore the link loss will be given by $32.4 + 20\log(15) + 20\log(20\,000) - 44 - 44 = 53.9$ dB.

Thus, if a received signal power of −40 dBm were required, a transmit power of 13.9 dBm would be needed. If we repeat the exercise for similar-sized antennas at a higher frequency, the loss is found to be less. It must be remembered that this is a prediction for propagation in a vacuum. Although the loss will be nearly as low as this for much of the time in the Earth's atmosphere, atmospheric effects will cause 'fading', leading to the need for a substantial 'fade margin'.

It is possible to do a quick estimate for the necessary receive power in a microwave point-to-point system. It depends upon the bit rate. The required power is approximately $-154 + 10\log(\text{bit rate})$ dBm. Microwave links usually demand high performance and low errors. This in turn means high power. Mobile systems typically operate at a higher bit error rate and can operate at lower power levels, perhaps 10 dB lower than for microwave links of the same bit rate.

1.4.1 Transmitting, directing and capturing power

We now have a picture of an antenna, both in transmit and in receive 'mode'. As a transmitter, the antenna directs energy into space. As a receiver, the antenna presents an aperture to the incident electromagnetic wave in order to gather energy. Our understanding can be enhanced by examining a point-to-point link in a different way. Let us consider our parabolic dish antenna with a perfect, conical, radiation pattern and a beamwidth of 3 degrees. At a distance of 1 kilometre, the antenna would