Appendix A
A Simple Introduction to Antennas

A.1 Introduction: Radiation Resistance and Radiation Patterns

An antenna is a transitional device, or transducer, that forms an interface for energy traveling between a circuit and free space, as depicted in Fig. A.1. It is reciprocal in the sense that it can transfer energy from the circuit to free space (transmission) and from free space to a circuit (reception). We can represent both the circuit and the antenna by their Thévenin equivalents as shown.

In engineering we represent the conversion of energy from electrical to some other non-recoverable form by a resistive load, since the resistor is an element that absorbs real power. We do the same for antennas; the radiation resistance R_r shown in Fig. A.1 models how much power is taken from the transmitter circuit and is radiated non-recoverably into free space. Alternately it is the source resistance of the antenna when it receives a signal, in which case the antenna model will also include a generator to represent the source of energy it is providing to the receiver circuit.

A reactive component X_r of the antenna model is seen in Fig. A.1. Ideally that should be zero because it signifies that energy reflects back from the antenna to the circuit. In the design of an antenna, one object is to make its radiation impedance real because then there can be a smooth transition from the circuit to free space. Ideally the radiation resistance of the antenna should match the output resistance of the circuit, and the circuit's output reactance should be zero, so that maximum power transfer can occur without reflection. In practice it is often difficult to achieve such a match so tuning circuits are sometimes employed between the circuit and the antenna.

Now consider the construction of the antenna. If there were no antenna and the circuit terminated in an open circuited transmission line, then theoretically all the power from the transmitter would be reflected backwards along the line. If the end of the line were flared, or even terminated in a dipole arrangement as shown in Fig. A.2, then a sizable proportion of the energy traveling forward along the transmission line from the transmitter circuit will detach and radiate into free space.

The dipole arrangement shown in Fig. A.2 is a very common form of antenna, particularly if its length from tip to tip is equivalent to a half wavelength of the signal being radiated. We can deduce some of its properties qualitatively, particularly in relation to the directions in which it radiates. For example, if we walked around a

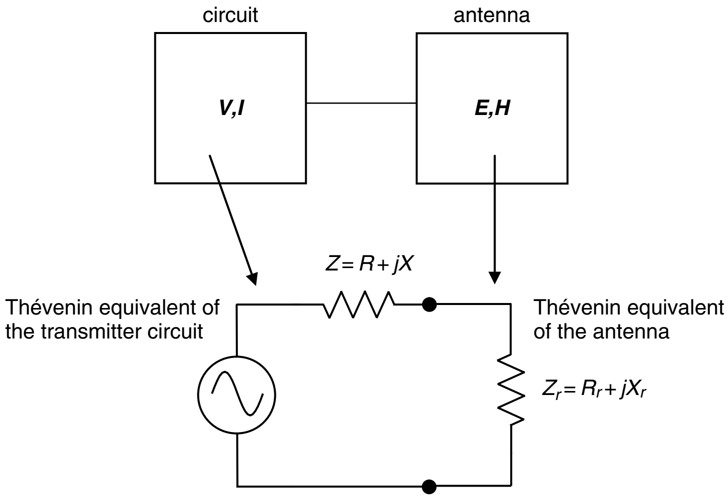


Fig. A.1 The antenna as an interface between a circuit and free space, along with their Thévenin equivalent circuits; the subscript *r* on the antenna model stands for "radiation"

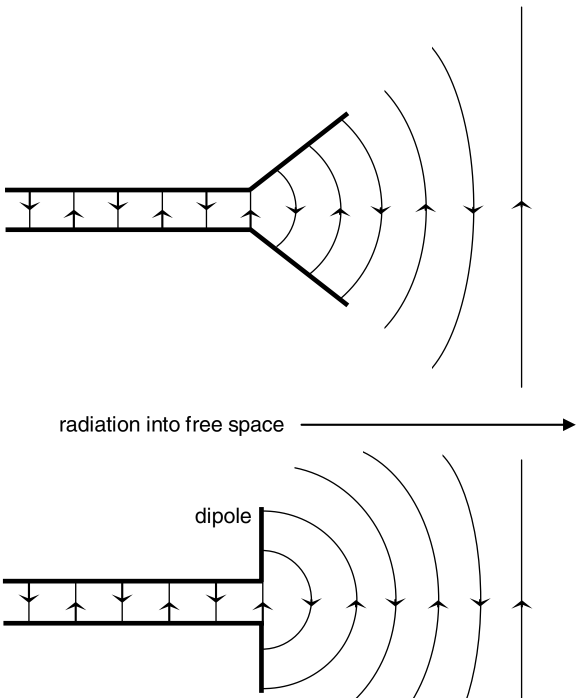


Fig. A.2 Use of a flared horn or a dipole as a transition from a circuit to free space

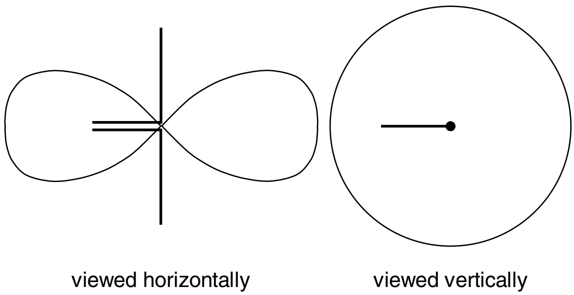


Fig. A.3 The radiation pattern of a dipole antenna

vertically deployed dipole antenna in the horizontal plane it would look no different when viewed from any angle. Therefore we would conclude that its radiating properties will not vary with angle around the dipole. If however we moved around it vertically its aspect would change from its appearance in Fig. A.2 to a single point when viewed directly from above or below. We can conclude therefore that its radiating properties will vary with vertical angle and, indeed, it may not radiate at all in the vertical direction. That is in fact the case for a dipole; it has a *radiation pattern* or *polar pattern* of the form of a doughnut as depicted in Fig. A.3.

As would be imagined, the radiation pattern of an antenna can be quite complicated. That leads to a number of definitions that are helpful in describing and antenna's properties. Figure A.4 shows a typical pattern in one plane, remembering that the full pattern will be a three dimensional figure. It is plotted in Cartesian rather than polar coordinates, which is often the case in practice.

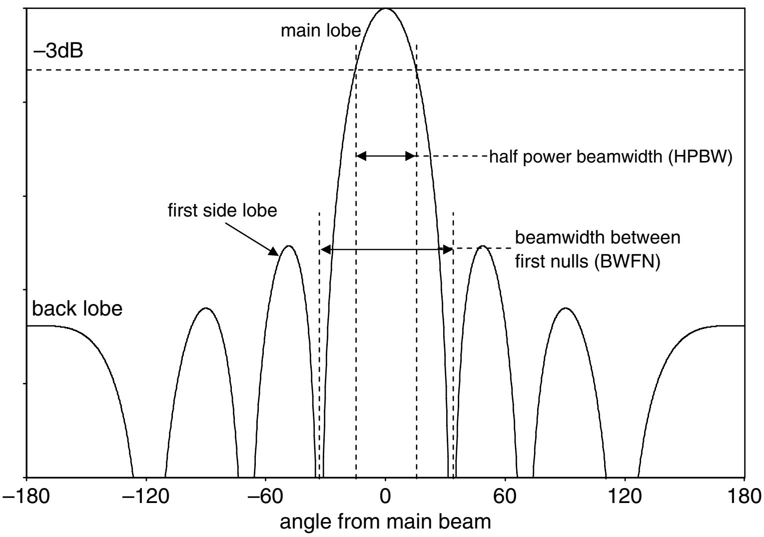


Fig. A.4 Radiation pattern of a fictional antenna plotted in Cartesian coordinates

A.2 The Directivity and Gain of an Antenna

It is now important to describe the antenna quantitatively. We commence with the definition of its *directivity* which, as the name implies, is a measure of how much it concentrates the energy in a certain direction. This is a three dimensional concept since an antenna, in principle, radiates in all directions. Geometric definitions for the derivation of an expression for directivity are given in Fig. A.5, from which it can be seen that the power available over an incremental area α at distance r from the antenna is given by

$$P = p\alpha = \frac{P_t\alpha}{4\pi\, r^2} = \frac{P_t\Omega}{4\pi}$$

giving as the power available per unit of solid angle – also known as *angular power density*

$$p = \frac{P}{\Omega} = \frac{P_t}{4\pi} = p(\theta,\phi)\ \mathrm{Wsr}^{-1} \tag{A.1}$$

Equation (A.1) is an algebraic expression for the three dimensional polar pattern. To proceed to a definition of directivity it is normalized with respect to its maximum to give

$$p_n(\theta,\phi) = \frac{p(\theta,\phi)}{p_{\max}(\theta,\phi)}$$

Integrating this dimensionless, normalized quantity over three dimensions we end up with the angular quantity

$$\Omega_A = \iint_{4\pi} p_n(\theta,\phi)d\theta\, d\phi \tag{A.2}$$

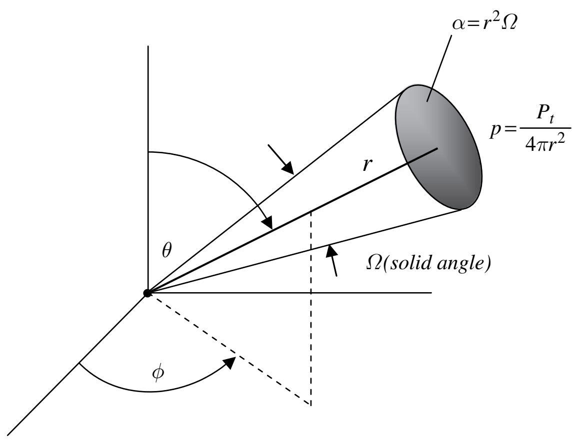


Fig. A.5 Coordinates and definitions for calculating the directivity of an antenna which is called the solid beam angle of the antenna. It is the three dimensional angle through which all the power of the antenna would be transmitted if its polar pattern were uniform over that angle as depicted in Fig. A.6.

We now define the directivity of the antenna as

$$D = \frac{4\pi}{\Omega_A} \tag{A.3}$$

which can be approximated

$$D = \frac{4\pi}{\theta_{HP}\phi_{HP}} \text{ with angles in radians, or}$$

$$D = \frac{41253}{\theta_{HP}\phi_{HP}} \text{ with angles in degrees.}$$

The *gain, G*, of an antenna is closely related to its directivity. They differ only through the efficiency, k, of the antenna, which accounts for ohmic losses in the antenna material. Thus

$$G = kD \qquad 0 < k < 1$$

Although (A.3) allows the directivity and thus gain of an antenna to be derived theoretically it is more usual to measure an antenna's gain. That is done relative to a reference antenna. One reference, even though not experimentally practical, is the isotropic radiator, which has a directivity and thus gain of unity (see Sect. 1.2). In principle, the gain determined from (A.3) is with respect to isotropic. It is more usually expressed in decibels with respect to isotropic, and written as

$$G = 10\log\frac{4\pi}{\Omega_A} = 10\log 4\pi - 10\log(\theta_{HP}\phi_{HP}) = 11 - 10\log(\theta_{HP}\phi_{HP}) \quad \text{dBi}$$

assuming the antenna is ideally efficient.

It is possible to calculate the gain of the dipole antenna when its length is equal to half a wavelength. It is, of course, also possible to construct a half wave dipole,

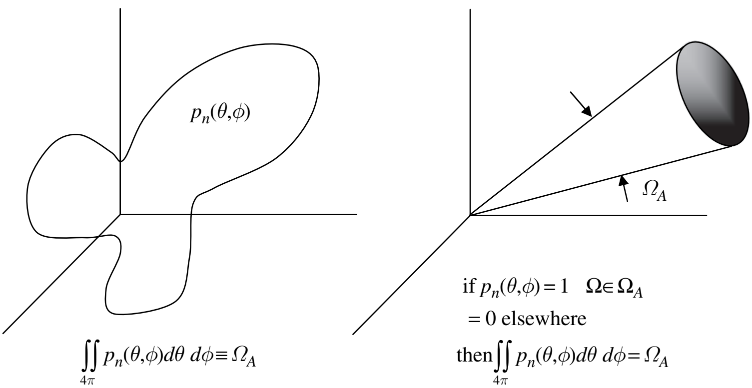


Fig. A.6 Demonstrating the definition of the angular beamwidth of an antenna so that it can be used as a reference antenna when measuring the gain of another antenna. Such a measurement is undertaken by transmitting to the antenna of interest and measuring the received signal, and in fact the full radiation pattern. The experimental antenna is then replaced by the reference dipole and the measurement repeated so that the measurement of the experimental antenna can be normalized to the dipole. The measured gain is then expressed as dB with respect to the half wave dipole, written as $\mathrm{dB}_{\lambda/2}$

The calculated gain of the dipole is 2.16 dBi; thus we have

$$\mathrm{G\ dBi} = \mathrm{G\ dB}_{\lambda/2} + 2.16\mathrm{dB}$$

A.3 The Aperture of an Antenna

Derivation of expressions for the aperture of an antenna requires a field theory analysis. For some antennas, such as a parabolic dish reflector, the aperture concept is straightforward and, provided the dish diameter is much larger than a wavelength, the aperture is related directly to the area presented to an incoming wave front. For other antennas, such as linear structures and even an isotropic radiator, the aperture concept is less straightforward but can, if we know its gain, be determined from

$$A = \frac{\lambda^2 G}{4\pi} \quad \mathrm{m}^2 \tag{A.4}$$

If the antenna has a physical aperture, such as a parabolic reflector, we introduce the concept of aperture efficiency to account for the difference between the physical and electromagnetic apertures:

$$A = k_{\mathrm{aperture}} A_{\mathrm{physical}} \qquad 0 < k_{\mathrm{aperture}} < 1$$

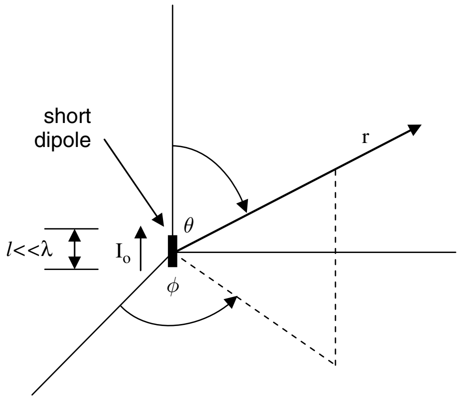
A.4 Radiated Fields

A treatment of the fields radiated by an antenna requires a field theory treatment and is beyond this coverage. It is, however, useful to examine well-known expressions for the fields produced by a so-called short dipole because the fields generated by other antennas can be derived from the short dipole results; it also allows us to understand the concept of near and far fields. Figure A.7 shows the geometry of a short dipole in which distance and direction out from the antenna is described by the radial coordinate *r*.

If the short dipole is carrying a sinusoidal current

$$I_o e^{j\omega t}$$

Fig. A.7 The short dipole



and is so short that there is no distribution of current along its length at any time, then the complete set of fields generated about the short dipole is

$$E_r = \frac{I_o l e^{j(\omega t - \beta r)} \cos\theta}{2\pi\varepsilon_o} \left(\frac{1}{cr^2} + \frac{1}{j\omega r^3} \right) \tag{A.5a}$$

$$E_\theta = \frac{I_o l e^{j(\omega t - \beta r)} \sin\theta}{4\pi\varepsilon_o} \left(\frac{j\omega}{c^2 r} + \frac{1}{cr^2} + \frac{1}{j\omega r^3} \right) \tag{A.5b}$$

$$H_\phi = \frac{I_o l e^{j(\omega t - \beta r)} \sin\theta}{4\pi} \left(\frac{j\omega}{cr} + \frac{1}{r^2} \right) \tag{A.5c}$$

In other words there are transverse components (θ, ϕ) of the magnetic and electric fields. There is also a radial electric field component (r) – i.e. in the direction of propagation. Note however it has a stronger inverse dependence on distance than the transverse components so that if the distance is sufficiently large it disappears and the transverse components themselves become just inverse distance dependent. This is demonstrated by letting r go large in (A.5a–c) to give

$$E_r = 0 \tag{A.6a}$$

$$E_\theta = \frac{j\omega I_o l e^{j(\omega t - \beta r)} \sin\theta}{4\pi\varepsilon_o c^2 r} \tag{A.6b}$$

$$H_\phi = \frac{j\omega I_o l e^{j(\omega t - \beta r)} \sin\theta}{4\pi cr} \tag{A.6c}$$

Thus for large distances the wave is TEM – i.e. transverse electromagnetic. Equations (A.6a–c) describe the so-called *far field* of the antenna. The far fields are inverse distance dependent and the treatment in this book, based on simple power and power density relationships, is valid. In contrast, closer to the antenna (A.5a–c) are needed to describe the field. That is called the *near field* of the antenna. The transition from near to far field is said to occur when the inverse distance terms in

(A.5b,c) are equal to the inverse distance squared terms, assuming that any inverse cubic terms are then negligible. Therefore the near field/far field transition is when

$$\left|\frac{\omega}{cr}\right| = \left|\frac{1}{r^2}\right|$$

which gives

$$r \approx \frac{\lambda}{6}.$$

It is of interest to note from (A.6b,c) that in the far field

$$\left|\frac{E_\theta}{H_\phi}\right| = \frac{1}{\varepsilon_o c} = \frac{\sqrt{\mu_o \varepsilon_o}}{\varepsilon_o} = Z_o$$

the free space impedance, as would be expected.

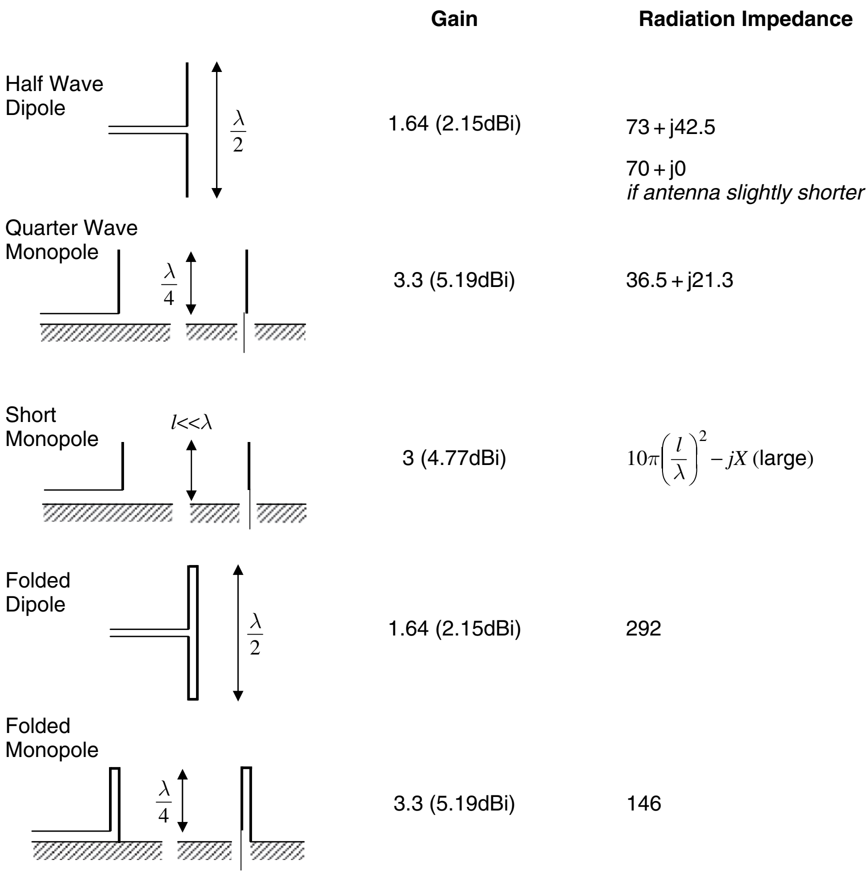


Fig. A.8 Some common linear antennas

A.5 Some Typical Antennas

Figure A.8 shows a number of simple antennas and their characteristics. Figure A.9 shows two common, compound antennas that are built up from combinations of active antennas, of the types shown in Fig. A.8, and passive linear elements.

The folded antennas shown in Fig. A.8 tend to have slightly broader bandwidths than their unfolded counterparts and are often used, particularly the dipole, in more complex structures such as the Yagi-Uda array illustrated in Fig. A.9. The short monopole in Fig. A.8 is commonly used as an AM receiving antenna on motor vehicles.

The log periodic antenna shown in Fig. A.9 is used when operation is necessary over a wide band of frequencies. Although it is more complex in construction than the Yagi, its wide operating bandwidth makes it attractive in many applications.

Figure A.10 shows a number of aperture and slot antennas, along with a bi-cone. Aperture reflectors tend to be used when the wavelength is much smaller than the diameter of the reflector, so they behave somewhat similar to optical reflectors of the same type.

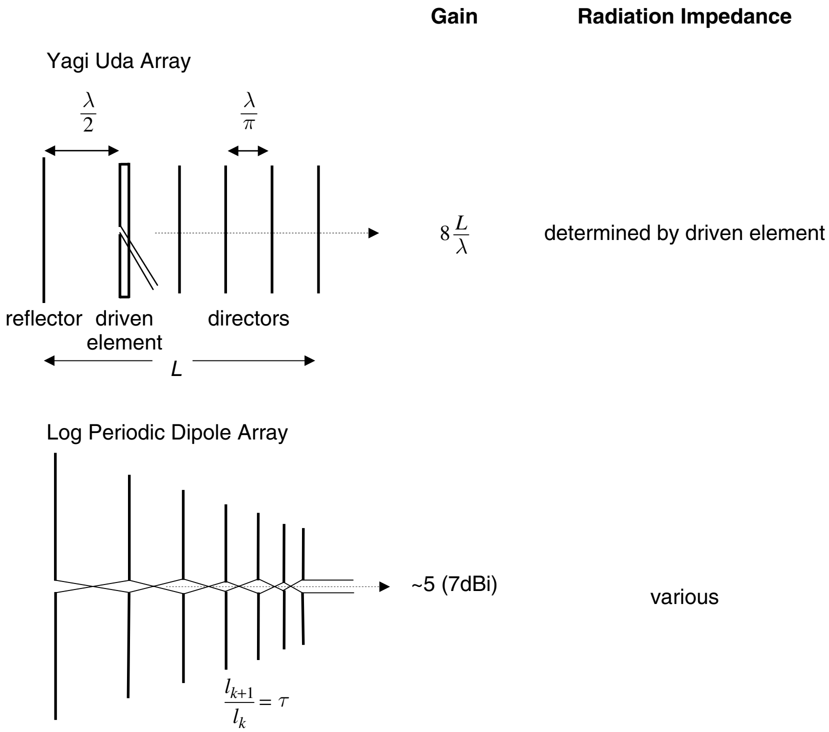


Fig. A.9 Some common compound antennas; the antenna lengths and spacings for the log periodic antenna are in the constant ratio τ as indicated for length

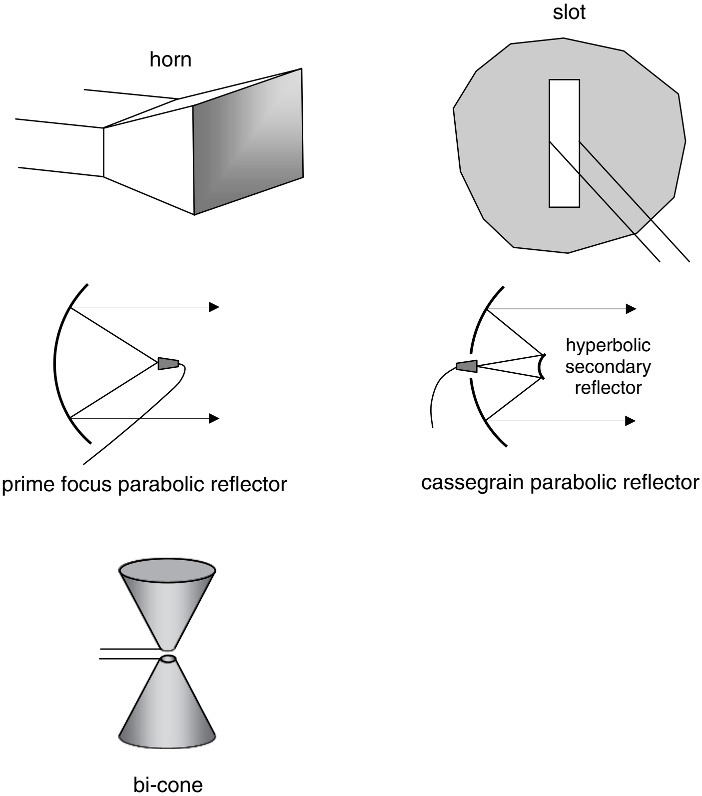


Fig. A.10 Aperture, slot and bi-conical antennas; the bi-cone is broad band and omni-directional in the horizontal plane

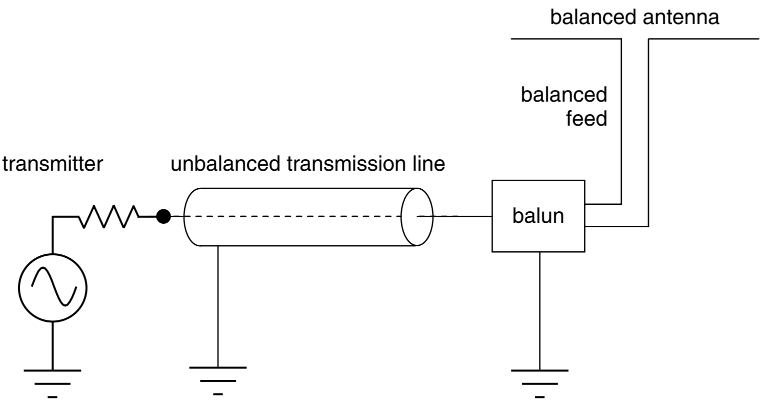


Fig. A.11 Demonstrating the use of a balun to provide the matched transition between an unbalance transmission line and a balanced antenna

A.6 Baluns

With the exception of monopoles, the other antennas in Figs. A.8 and A.9 require balanced feeds. In other words they need to be fed by transmission lines that have neither conductor at earth potential. Yet many of the feed lines in practice are coaxial cables that clearly have one of their conductors – the braid – at earth potential. Coaxial cables are also compatible with many transmitter output circuits that are also unbalanced as noted by the manner in which the Thévenin equivalent is depicted in Fig. A.1. To render the unbalanced transmission line compatible with a balanced antenna a device referred to as a balun is employed, as illustrated in Fig. A.11. Short for *balanced-unbalanced*, this device can be constructed in several forms, each of which not only has to perform the unbalanced-to-balanced transformation but also has to match impedances for maximum power flow and to minimise reflections.

There are many forms of balun, the simplest of which is a transformer. For narrow band operation a simple balun can be constructed from sections of transmission line.

Appendix B
The Use of Decibels in Communications Engineering

Logarithms have two major benefits: they readily summarise numbers that extend over a large range and they simplify multiplication. As a consequence, the decibel (dB), which is defined using base 10 logarithms, is widely used as a convenient measure in many branches of engineering, but especially in communications. Although it can be used with signals generally, it is principally defined in terms of power (or power density). More precisely, the decibel (dB) is defined on the basis of a *reference* power:

$$10\log_{10}\frac{P}{P_{ref}} = x \text{ dB}$$

We say that P is x dB larger than P_{ref}. For example

If P =	then x is
$2P_{\text{ref}}$	3 dB wrt P_{ref}
$10P_{\text{ref}}$	10 dB wrt P_{ref}
$100P_{\text{ref}}$	20 dB wrt P_{ref}
$0.5P_{\text{ref}}$	−3 dB wrt P_{ref}
$0.1P_{\text{ref}}$	−10 dB wrt P_{ref}

Many factors have easily constructed dB equivalents. For example

$$200 = 2\times 100 \rightarrow 3\,\text{dB} + 20\,\text{dB} = 23\,\text{dB},$$

as a result of the additive property of logarithms.
Similarly

$$17\,\text{dB} = 20\,\text{dB} - 3\,\text{dB} \rightarrow 100 \div 2 = 50,$$
$$36\,\text{dB} = 30\,\text{dB} + 6\,\text{dB} \rightarrow 1000\times 4 = 4000.$$

In telecommunications, two common values of P_{ref} are used. The dBs are then given special symbols that imply absolute, as against relative, quantities.

If $P_{\text{ref}} = 1\,\text{W}$, then we use dBW.

If $P_{\text{ref}} = 1\,\text{mW}$, then we use dBm.

Thus we can see the following equivalences:

17 dBm	→ 50 mW	23 dBW	→ 200 W
3 dBm	→ 2 mW	10 dBW	→ 10 W
30 dBm	→ 1 W	20 dBW	→ 100 W
0 dBm	→ 1 mW	0 dBW	→ 1 W
-20 dBm	→ 10μW	-40 dBW	→ 100μW

Decibels can also be used with voltages, but the definition still rests upon power. For example

$$10\log\frac{P}{P_{ref}} = 10\log_{10}\frac{V^2}{V_{ref}^2} = 20\log\frac{V}{V_{ref}} = x\text{ dB}$$

So that if $V = 2V_{\text{ref}}$, then $x = 6\,\text{dB}$.

Appendix C
The Dielectric Constant of an Ionospheric Layer

Equation (3.1) notes that the refractive index of an ionospheric layer is given by

$$n = \sqrt{1 - \frac{81N}{f^2}}$$

We derive that expression below, following the approach of D.J. Angelakos and T.E. Everhart, *Microwave Communication*, McGraw-Hill, New York, 1968.

An ionised region of the atmosphere, such as one of the layers of the ionosphere, will be composed of free ions, electrons and neutral molecules. We assume that the ions, because of their mass, do not respond as well to the passage of an electromagnetic field as the electrons and thus have little effect on it. We will therefore concentrate our attention just on the free electrons, which we assume to be present with density N electrons per cubic metre.

We also assume that the earth's magnetic field has no effect, and that the collisions that occur between electrons and neutral atmospheric constituent molecules can be neglected. Those collisions are significant if we are interested in the attenuation of a wave in transmission through the atmosphere; we mention that below, after the derivation of refractive index.

The response of an individual electron of mass m to an applied electric field $|\mathbf{E}| = E_m \cos \omega t$ Vm^{-1} (resulting from the passage of a radio wave) is given from Newton's law

$$F = ma$$

If the charge on the electron is e and its velocity is v then this last expression can be written

$$eE_m \cos \omega t = m\frac{dv}{dt}$$

which gives for the electron velocity

$$v = \frac{eE_m}{\omega m} \sin \omega t.$$

This movement of electrons gives rise to a conduction current described by the transverse areal current density

$$\mathbf{j}_{\text{cond}} = veN \quad \text{Am}^{-2}$$
$$= \frac{e^2 N E_m}{\omega m} \sin \omega t \tag{C.1}$$

There will also be a displacement areal current density as a result of the dielectric behaviour of the medium, found from

$$\mathbf{j}_{\text{dis}} = \frac{d\mathbf{D}}{dt} = \varepsilon \frac{d\mathbf{E}}{dt} \tag{C.2}$$

in which $\mathbf{D}$ is the electric displacement vector and ε is the permittivity of the medium.

For a plasma as dilute as an ionospheric layer $\varepsilon = \varepsilon_o$, so that

$$|\mathbf{j}_{\text{dis}}| = \varepsilon_o \frac{d\,|\mathbf{E}|}{dt} = -\varepsilon_o \omega E_m \sin \omega t$$

Thus the magnitude of the total current in the layer induced by the passage of a radio wave is

$$|\mathbf{j}_{\text{total}}| = |\mathbf{j}_{\text{cond}} + \mathbf{j}_{\text{dis}}| = \left(\frac{e^2 N}{\omega m} - \varepsilon_o \omega\right) E_m \sin \omega t \tag{C.3}$$

Even though there are free electrons present, we now regard the layer as behaving entirely as a dielectric with permittivity ε – i.e. as though there were no free electrons. The displacement current density is then given just by (C.2). If we call that an *effective* displacement current density and equate it to the actual current density given by (C.3) we have using (C.1)

$$\left|\mathbf{j}_{\text{effective,dis}}\right| = -\varepsilon \omega E_m \sin \omega t = |\mathbf{j}_{\text{cond}} + \mathbf{j}_{\text{dis}}|$$

so that

$$-\varepsilon \omega = \left(\frac{e^2 N}{\omega m} - \varepsilon_o \omega\right)$$

or

$$\varepsilon = \varepsilon_o (1 - \omega_p^2 / \omega^2) \tag{C.4}$$

in which

$$\omega_p = \left(\frac{e^2 N}{m \varepsilon_o}\right)^{1/2} \tag{C.5}$$

is called the plasma frequency of the region with electron density N.

Since

$$e = 1.6 \times 10^{-19}\text{C}$$
$$m = 9.11 \times 10^{-31}\,\text{kg}$$
$$\varepsilon_\text{o} = 8.85\ \text{pFm}^{-1}$$

then

$$\omega_p^2 = 3175N$$

Thus (C.4) becomes

$$\varepsilon = \varepsilon_o(1 - \frac{3175N}{\omega^2}) = \varepsilon_o(1 - \frac{81N}{f^2}) \tag{C.6}$$

from which we recognise that the equivalent dielectric constant (or relative permittivity) of the region is

$$\varepsilon_\text{rel} = (1 - \frac{81N}{f^2})$$

Thus the refractive index of a region of the ionosphere of electron density N and frequency f is

$$n = \sqrt{1 - \frac{81N}{f^2}} \tag{C.7}$$

Recall that this expression, and (C.6), was derived by ignoring losses resulting from electron-neutral collisions. If they were included Rohan[1] shows that (C.6) would be

$$\varepsilon = \varepsilon_o\left(1 - \frac{e^2N}{m\varepsilon_o(\omega^2 + \nu^2)}\right) \tag{C.8}$$

in which ν is the collision frequency of the electrons and neutrals. While the collision frequency is very high in the lower atmosphere because of the neutral density, it is of the order of 1000 or less at the height of the upper ionospheric layers. As a consequence, at the sorts of frequencies normally associated with sky wave propagation $\omega^2 \gg \nu^2$, so that (C.8) reduces to (C.6).

Electron-neutral collisions give rise to losses in the ionosphere, particularly at the lower levels; their effect can be characterised by an equivalent conductivity from which an attenuation constant can be derived. Again, following Rohan, the conductivity of a region of ionisation is

$$\sigma = \frac{e^2N\nu}{m(\omega^2 + \nu^2)}$$

[1] P. Rohan, *Introduction to Electromagnetic Wave Propagation*, Artech, Boston, 1991.

The attenuation constant is then

$$\alpha = \frac{60\pi\sigma}{n} = \frac{60\pi e^2 N\nu}{nm(\omega^2 + \nu^2)}$$

in which n is refractive index. Thus the attenuation of a layer decreases with an increase in operating frequency.

To obtain an idea of the levels of attenuation likely to be encountered by a wave travelling through the D region (above its critical frequency) suppose we choose typical values of $\nu = 10^7\,\mathrm{s}^{-1}$, $N = 10^8$ electrons m^{-3} and $f = 1\,\mathrm{MHz}$ (ie the AM broadcast band). After substituting we have

$$\begin{aligned}\alpha &= 3.8 \times 10^{-5}\,\mathrm{Npm}^{-1} \\ &= 0.33\,\mathrm{dBkm}^{-1}\end{aligned}$$

Thus if the layer were equivalently 25 km thick at that effective electron density then the total attenuation at 1 MHz would be 8.25 dB at vertical incidence and considerably more at oblique incidence. In the evening such a high level of attenuation does not occur because of the absence of the D region. As a consequence it is possible to receive distant AM stations in the evening that are not available during daylight hours.

Index