8 Antennas

We have already seen that antennas are important component of a radio system and, in Chapter 2, have already discussed the most basic of all antennas, the dipole. The radio electronics merely produce an oscillating voltage and it is the function of the antenna to turn this voltage into the desired radio wave. Consequently, the efficiency with which an antenna performs this function is of great importance to radio technology. In the following chapter we will further develop the theory of antennas and, in particular, will look at the factors that control the performance of these devices. Further, we will study how different forms of antenna can meet the requirements of different radio technologies. We will look at electrically small antennas for situations where size is important (in personal communications for example), broadband antennas for situations where multiple frequencies are to be used (in communications via the ionosphere for example) and array antennas for situations where the radiation needs to be steered in a particular direction (in radar for example).

8.1 The Electric Dipole

In Chapter 2 we introduced the idea of a dipole antenna, an antenna which consisted of a metallic rod with an RF signal driven into its centre. An electric field develops between the two sides of the antenna and so it is often known as an *electric dipole*. Obviously, this antenna must be driven by a radio transmitter and the power will be transferred from the transmitter to the antenna through a transmission line. The dipole is a balanced load and so, from the considerations of the previous chapter, it needs to be fed by a balanced transmission line (if not a BALUN will need to be used). We will assume that the antenna is fed by a parallel-wire transmission line and then the dipole antenna can be pictured as a transmission line that has been opened out (see Figure 8.1) to connect into free space. The speed of propagation c on parallel wires does not depend on the distance between the wires and so will be the same when the wires are opened out (this will be the speed of light in free space). Consequently, the current on the dipole will satisfy a wave equation with wave speed c. Let the arms of the antenna have length l (see Figure 8.2a) and assume the antenna is driven by a time-harmonic current source at the dipole centre ($\mathcal{I}_0(t) = \Re\{I_0 \exp(\mathrm{j}\omega t)\}$). Noting that the current will be zero at the end of each dipole arm, the current will be distributed on the dipole according to

$$\mathcal{I}_0(t,z) = \Re\left\{I_0 \exp(\mathrm{j}\omega t)\frac{\sin(\beta(l-|z|))}{\sin\beta l}\right\}, \tag{8.1}$$

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

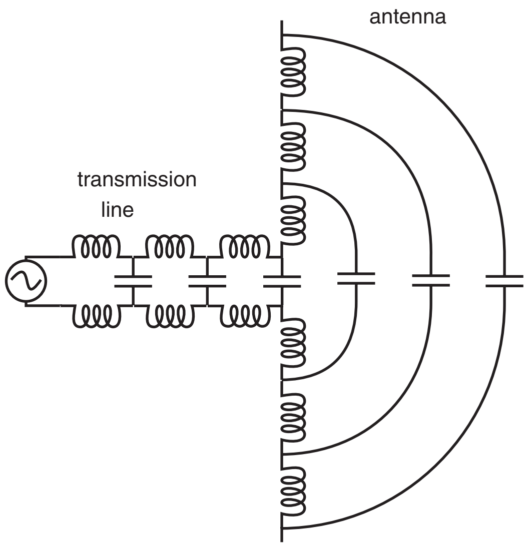


Fig. 8.1 Circuit model of a dipole.

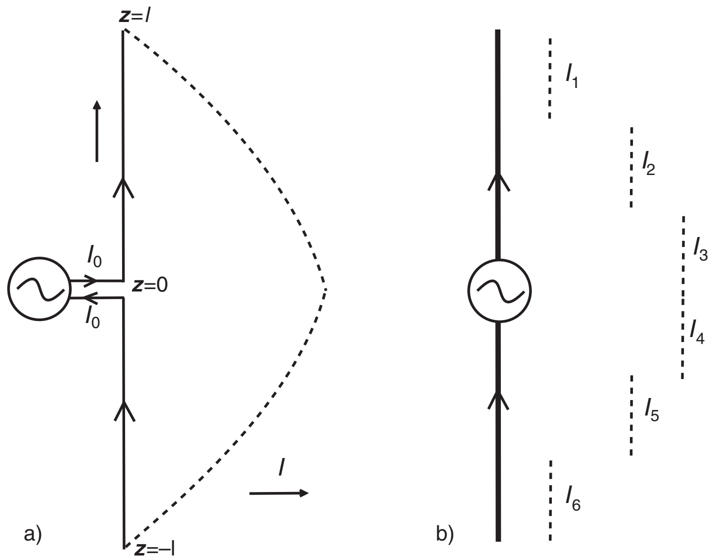


Fig. 8.2 Current distribution along a dipole and its approximation by ideal dipoles.

where $\beta = 2\pi/\lambda$ (note that the complex amplitude of current on the dipole current is given by $I = I_0 \sin(\beta(l-|z|))/\sin\beta l$).

In Chapter 2 we considered an *ideal dipole*, a harmonically oscillating current that is uniformly distributed along the length of the dipole. At large distances from this dipole, the field is given by $\mathcal{E}_t = \Re\{\exp(\mathrm{j}\omega t)E_t\}$, where

$$E_t(R,\theta) = \mathrm{j}\omega\mu_0 I\Delta z \sin\theta \frac{\exp(-\mathrm{j}\beta R)}{4\pi R} \tag{8.2}$$

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

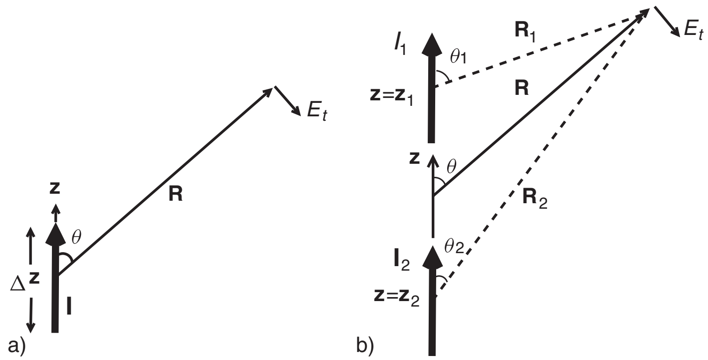


Fig. 8.3 Single and dual ideal dipole elements.

in which I is the complex amplitude of the current, Δz is the dipole length, R is the distance from the dipole and θ is the angle between the dipole axis and the radial vector to the observation point (see Figure 8.3a). The field is a spherical wave that travels outwards from the dipole. In the locality of the observation point P, the field will behave like a plane wave which propagates in the radial direction. The electric field is transverse to this direction (in the plane of the dipole and point P) with magnitude $\mathcal{E}_t$ and the magnetic field is orthogonal to both the electric field and propagation direction with magnitude $\mathcal{B}_t = \mathcal{E}_t/c$. (Note that the electric field, magnetic field and propagation direction form an orthogonal triad that obeys the right-hand screw rule.) For a general ideal dipole with current I in the direction of unit vector $\hat{\mathbf{I}}$, the electric field will be given by $\boldsymbol{\mathcal{E}} = \Re\{\exp(\mathrm{j}\omega t)\mathbf{E}\}$ where

$$\mathbf{E}(R,\theta) = \mathrm{j}\omega\mu_0 I \Delta l \hat{\mathbf{R}} \times (\hat{\mathbf{R}} \times \hat{\mathbf{I}}) \frac{\exp(-\mathrm{j}\beta R)}{4\pi R}, \tag{8.3}$$

with $\hat{\mathbf{R}}$ a unit vector in the radial direction and Δl the dipole length. On noting the vector identity $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{BA} \cdot \mathbf{C} - \mathbf{CA} \cdot \mathbf{B}$, we can put this in the more meaningful form

$$\mathbf{E}(R,\theta) = \mathrm{j}\omega\mu_0 I \Delta l (\hat{\mathbf{R}}\hat{\mathbf{R}} \cdot \hat{\mathbf{I}} - \hat{\mathbf{I}}) \frac{\exp(-\mathrm{j}\beta R)}{4\pi R}, \tag{8.4}$$

where $\hat{\mathbf{R}}\hat{\mathbf{R}} \cdot \hat{\mathbf{I}} - \hat{\mathbf{I}}$ is a vector in a direction transverse to $\hat{\mathbf{R}}$ having magnitude $\sin\theta$ where θ is the angle between vectors $\hat{\mathbf{R}}$ and $\hat{\mathbf{I}}$ (note that $\hat{\mathbf{I}} - \hat{\mathbf{R}}\hat{\mathbf{R}} \cdot \hat{\mathbf{I}}$ is essentially $\hat{\mathbf{I}}$ with its component in the $\hat{\mathbf{R}}$ direction removed).

The field of a general current distribution can be made up of a combination of ideal dipoles (the field of a set of dipoles is simply the sum of the fields of the individual dipoles). Consider the two elements (both of length Δz) shown in Figure 8.3b, then the field will be given by

$$E_t(R,\theta) = \mathrm{j}\omega\mu_0 \Delta z \left(I_1 \sin\theta_1 \frac{\exp(-\mathrm{j}\beta R_1)}{4\pi R_1} + I_2 \sin\theta_2 \frac{\exp(-\mathrm{j}\beta R_2)}{4\pi R_2} \right). \tag{8.5}$$

From the law of cosines for triangles, we will have that

$$R_1 = \sqrt{R^2 + z_1^2 + 2|z_1|R\cos\theta}$$

$$\text{and} \quad R_2 = \sqrt{R^2 + z_2^2 + z_1^2 - 2|z_2|R\cos\theta}. \tag{8.6}$$

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

Since we are considering the field far away from the antenna we will have $|z_1| \ll R$. Consequently, on noting that $\sqrt{1+x} \approx 1 + x/2$ for small x, $R_1 = R\sqrt{1 + z_1^2/R^2 + 2|z_1|\cos\theta/R}$ can be approximated by $R_1 \approx R + |z_1|\cos\theta$ and, in a similar fashion, $R_2 \approx R - |z_2|\cos\theta$. By the sine rule for triangles, $\sin\theta_1/R = \sin\theta/R_1$ and so, to the leading order in $1/R$, $\theta_1 \approx \theta$ (in a similar fashion $\theta_2 \approx \theta$). We will now have that

$$\begin{aligned} E_t(R,\theta) \approx j\omega\mu_0\Delta z\,(I_1 \exp(j\beta z_1\cos\theta) \\ + I_2\exp(j\beta z_2\cos\theta))\sin\theta\frac{\exp(-j\beta R)}{4\pi R}. \end{aligned} \tag{8.7}$$

In general, for elements at $z_1, z_2, \ldots, z_N$ with currents $I_1, I_2, \ldots, I_N$,

$$E_t(R,\theta) \approx \mathrm{j}\omega\mu_0\Delta z\left(\sum_{i=1}^{N} I_i \exp(\mathrm{j}\beta z_i\cos\theta)\right)\sin\theta\frac{\exp(-\mathrm{j}\beta R)}{4\pi R}. \tag{8.8}$$

We can approximate the dipole as a series of current elements (see Figure 8.2b) and then, in the limit that the number of current elements tends to infinity, the sum will become an integral, i.e.

$$E_t(R,\theta) = \mathrm{j}\omega\mu_0\left(\int_{-l}^{l} I(z)\exp(\mathrm{j}\beta z\cos\theta)dz\right)\sin\theta\frac{\exp(-\mathrm{j}\beta R)}{4\pi R}. \tag{8.9}$$

For a general dipole

$$I(z) = I_0\frac{\sin(\beta(l-|z|))}{\sin\beta l} \tag{8.10}$$

and so

$$E_t(R,\theta) = 2\mathrm{j}\eta_0 I_0 g(\theta)\sin\theta\frac{\exp(-\mathrm{j}\beta R)}{4\pi R}, \tag{8.11}$$

where

$$g(\theta) = \frac{cos(\beta l\cos\theta) - \cos(\beta l)}{\sin^2\theta\sin\beta l}. \tag{8.12}$$

Two special cases are the *short dipole* ($l \ll \lambda$) for which $g(\theta) = \beta l/2$ and the *half-wave dipole* ($l = \lambda/4$) for which $g(\theta) = \cos(\frac{\pi}{2}\cos\theta)/\sin^2\theta$.

We consider the average power P_{rad} that is radiated by a dipole antenna. From (2.37), when the antenna is driven by a time-harmonic current, this is

$$P_{\mathrm{rad}} = \frac{1}{2\eta_0}\int_S |E_t|^2\mathrm{d}S. \tag{8.13}$$

Then, from (2.38),

$$P_{\mathrm{rad}} = \frac{1}{2\eta_0}\int_0^{\pi} |E_t|^2 2\pi R^2\sin\theta\mathrm{d}\theta \tag{8.14}$$

when the surface S is a sphere of radius R. In the case of an ideal dipole with length Δz, this can be integrated to yield $P_{\mathrm{rad}} = 40\pi^2\Delta z^2|I_0|^2/\lambda^2$. Further, for a general dipole

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

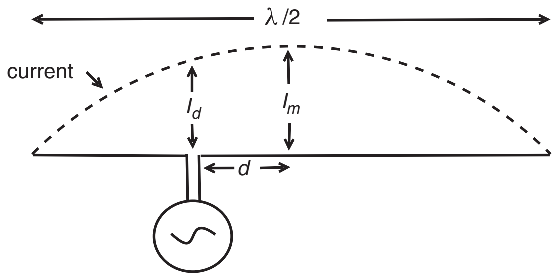


Fig. 8.4 The offset fed dipole.

with current distribution 8.10,

$$P_{\rm rad} = \frac{\eta_0 |I_0|^2}{4\pi} \int_0^{\pi} g(\theta)^2 \sin^3\theta \, d\theta. \tag{8.15}$$

For the case of a short dipole, the integral can be performed analytically to yield $P_{\rm rad} = 40\pi^2 l^2 |I_0|^2/\lambda^2$. In general, the integral needs to be integrated numerically and, in the case of a half-wave dipole, this leads to the result $P_{\rm rad} = 1.218\lambda |I_0|^2/4\pi$. If all the power that is supplied to the antenna is radiated, the source of the antenna will see a load resistance $R_{\rm rad} = 2P_{\rm rad}/I_0^2$. Consequently, for the ideal dipole $R_{\rm rad} = 80\pi^2 \Delta z^2/\lambda^2$ ohms, for the short dipole $R_{\rm rad} = 80\pi^2 l^2/\lambda^2$ ohms and for the half-wave dipole $R_{\rm rad} \approx 73$ ohms. It turns out that dipoles can also exhibit some reactance X_A. For a half-wave dipole $X_A \approx 43$ ohms and for a short dipole $X_A = -120(\ln(l/a) - 1)/\beta l$ ohms where a is the radius of the dipole rod. If the half-wave dipole is slightly shortened (by about 5 per cent) it is possible to reduce X_A to zero.

Thus far, we have only considered dipoles that are fed at their centre. This does not need to be the case and a dipole can be fed at any point. However, the impedance that the dipole presents to a source will change with the position at which it is fed. For a half-wave dipole, the current at a point displaced distance d from the centre will be

$$I_d = I_m \cos\left(2\pi \frac{d}{\lambda}\right). \tag{8.16}$$

The radiated power $P_{\rm rad}$ will be the same wherever the antenna is fed and so, at distance d from the centre, the input resistance R_d will satisfy $P_{\rm rad} = \frac{1}{2} R_d I_d^2 = \frac{1}{2} R I_m^2$ where R_m is the input resistance at the antenna centre. As a consequence,

$$R_d = R_m \frac{1}{\cos^2\left(2\pi \frac{d}{\lambda}\right)} \tag{8.17}$$

and from which we see that the resistance is minimum at the centre and increases towards the dipole ends. (Although an infinite resistance is predicted for a dipole end, in reality this will be a finite, but very large, value.)

For a perfectly conducting antenna, loss will only occur through radiation, i.e. the loss resistance $R_{\rm rad}$. However, if the material of a dipole antenna is imperfectly conducting, power will be dissipated as heat in the antenna structure. For DC, the resistance of a wire of length Δz is given by $R = \Delta z/\pi a^2 \sigma$ where σ is the conductivity

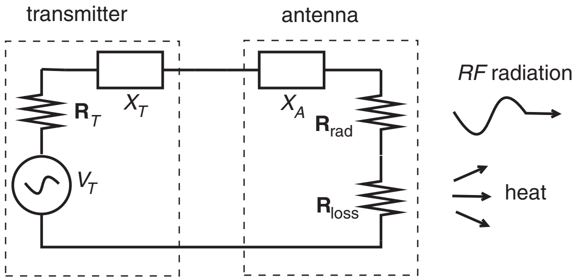


Fig. 8.5 Circuit model of a transmit antenna.

of the wire. At radio frequencies, however, things are not quite so simple due to a phenomenon known as the *skin effect*. Radio waves are heavily attenuated in metals and so current is unable to penetrate very far into the metal, being confined to a thin layer at the surface (thickness $\delta = \sqrt{2/\omega\mu_0\sigma}$). The resistance of a wire of length Δz will therefore be $R = \Delta z/2\pi a\delta\sigma$ and the average power lost in this wire will be $\Delta P_{loss} = (\Delta z/2\pi a\delta\sigma)|I|^2/2$. ($\sigma = 3.5 \times 10^7$ siemens per metre in the case of aluminium, a common material used in the construction of antennas.) Summing such losses across the dipole, and taking the limit $\Delta z \to 0$, we obtain that

$$P_{loss} = \int_{-l}^{l} \frac{|I|^2}{4\pi a\delta\sigma} dz. \tag{8.18}$$

For a short dipole $P_{loss} = lI_0^2/6\pi a\delta\sigma$ and for a half-wave dipole $P_{loss} = lI_0^2/4\pi a\delta\sigma$. The loss resistance of the dipole will be $R_{loss} = 2P_{loss}/|I_0|^2$ which, for a short dipole, is given by $R_{loss} = l/3\pi a\delta\sigma$ and for a half-wave dipole by $R_{loss} = l/2\pi a\delta\sigma$. The antenna and transmitter can now be modelled as the circuit of Figure 8.5 and from this it can be seen that the maximum power will be transferred from the transmitter to the antenna when $X_A = -X_T$ and $R_T = R_{rad} + R_{loss}$ i.e. when there is a conjugate match.

The fact that some of the power that is fed to the antenna will be lost as heat on the antenna body, and not radiated, brings us to the concept of *antenna efficiency*. This is a measure of how well the antenna radiates the power that it accepts and is defined by

$$\text{efficiency} = \frac{P_{rad}}{P_{rad} + P_{loss}} = \frac{R_{rad}}{R_{rad} + R_{loss}}. \tag{8.19}$$

It is clear that, for an efficient antenna, the radiation resistance will need to be much larger than the loss resistance. Although this will usually be the case for an antenna such as a half-wave dipole, this will not be the case for antennas that are short in comparison to a wavelength. It will be noted that, for a short dipole, $R_{loss} \propto l$ and that $R_{rad} \propto l^2$. This is bad news as it implies that the efficiency will tend to zero as the length of the dipole tends to zero. The basic message is that small antennas can be very inefficient.

8.2 The Magnetic Dipole

We have already seen that loops were used as receiving antennas in the early days of radio. If a harmonic EM wave impinges on such a loop, Faraday's law implies that the magnetic field $\mathcal{B}$ of the wave will induce an open circuit voltage $V = j\omega A\mathcal{B} \cdot \mathbf{n}$ where A

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

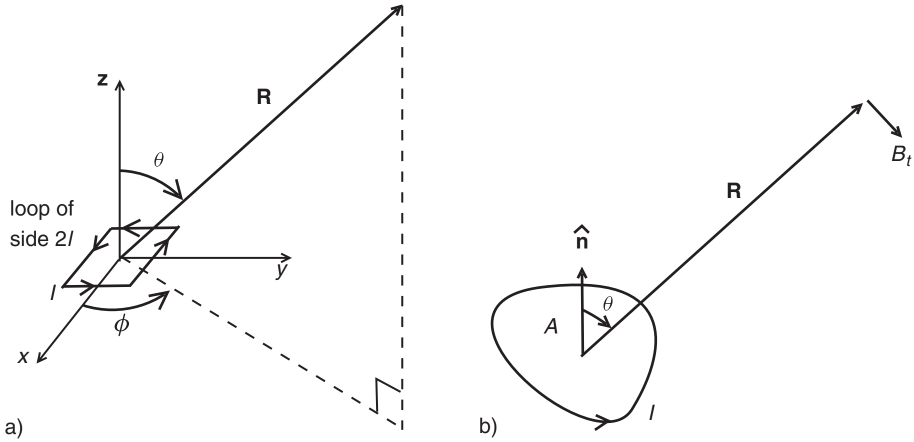


Fig. 8.6 Magnetic dipole antenna.

is the area of the loop and $\mathbf{n}$ is a unit vector perpendicular to the loop (we assume the loop has dimensions much less than a wavelength). If, on the other hand, we use the loop for transmission, we will drive a current I into the loop and, in the case of a rectangular loop, we can model this as four electric dipoles of length $2l$ (see Figure 8.6a).

For observation points that are a large distance away from the rectangular loop, the difference between a dipole that constitutes a side and one located at the origin is merely a phase factor (as we have seen in our considerations of the electric dipole). The sides that are parallel to the x axis will generate a transverse field $\mathbf{E}^x$ that lies in the plane of the x axis and the radial direction. In a similar fashion, the sides that are parallel to the y axis will generate a transverse field $\mathbf{E}^y$ that lies in the plane of the y axis and the radial direction. From (8.4), we will have that

$$\mathbf{E}^x(R,\theta,\phi) = \mathrm{j}\omega\mu_0 2lI(\hat{\mathbf{R}}\hat{\mathbf{R}}\cdot\hat{\mathbf{x}} - \hat{\mathbf{x}})\left(\exp(-\mathrm{j}\beta l\hat{\mathbf{R}}\cdot\hat{\mathbf{y}}) - \exp(\mathrm{j}\beta l\hat{\mathbf{R}}\cdot\hat{\mathbf{y}})\right)\frac{\exp(-\mathrm{j}\beta R)}{4\pi R} \tag{8.20}$$

and

$$\mathbf{E}^y(R,\theta,\phi) = \mathrm{j}\omega\mu_0 2lI(\hat{\mathbf{R}}\hat{\mathbf{R}}\cdot\hat{\mathbf{y}} - \hat{\mathbf{y}})\left(\exp(\mathrm{j}\beta l\hat{\mathbf{R}}\cdot\hat{\mathbf{x}}) - \exp(-\mathrm{j}\beta l\hat{\mathbf{R}}\cdot\hat{\mathbf{x}})\right)\frac{\exp(-\mathrm{j}\beta R)}{4\pi R}. \tag{8.21}$$

(Note that, in terms of the polar coordinates θ and ϕ, $\hat{\mathbf{R}} = \sin\theta\cos\phi\hat{\mathbf{x}} + \sin\theta\sin\phi\hat{\mathbf{y}} + \cos\theta\hat{\mathbf{z}}$.) We will assume the lengths of the sides are much less than a wavelength and, by means of the approximation $\exp x \approx 1 + x$, the total field ($\mathbf{E} = \mathbf{E}^x + \mathbf{E}^y$) will reduce to

$$\begin{aligned}\mathbf{E}(R,\theta,\phi) &= \omega\mu_0\beta AI\left((\hat{\mathbf{R}}\hat{\mathbf{R}}\cdot\hat{\mathbf{x}} - \hat{\mathbf{x}})\hat{\mathbf{R}}\cdot\hat{\mathbf{y}} - (\hat{\mathbf{R}}\hat{\mathbf{R}}\cdot\hat{\mathbf{y}} - \hat{\mathbf{y}})\hat{\mathbf{R}}\cdot\hat{\mathbf{x}}\right)\frac{\exp(-\mathrm{j}\beta R)}{4\pi R} \\ &= \omega\mu_0\beta AI\hat{\mathbf{z}}\times\hat{\mathbf{R}}\frac{\exp(-\mathrm{j}\beta R)}{4\pi R}.\end{aligned} \tag{8.22}$$

For a general plane loop, with area A and unit normal $\hat{\mathbf{n}}$, the electric field will be

$$\mathbf{E}(R,\theta,\phi) = \omega\mu_0\beta AI\hat{\mathbf{n}}\times\hat{\mathbf{R}}\frac{\exp(-\mathrm{j}\beta R)}{4\pi R}. \tag{8.23}$$

Further, at large distances from the source, the field will behave as a plane wave travelling in the direction of $\hat{\mathbf{R}}$ and so $\mathbf{B} = -\hat{\mathbf{R}} \times \mathbf{E}/c$. As a consequence,

$$\mathbf{B}(R,\theta,\phi) = \mu_0\beta^2 AI\hat{\mathbf{R}} \times (\hat{\mathbf{R}} \times \cdot\hat{\mathbf{n}})\frac{\exp(-j\beta R)}{4\pi R}. \tag{8.24}$$

From (8.24) and (8.3), it can be seen that the magnetic field for a loop is proportional to the electric field of a short dipole in the direction $\hat{\mathbf{n}}$. Because of this, a current loop is often known as a *magnetic dipole*. By analogy with the electric dipole, there is only a transverse component of magnetic field (see Figure 8.6b) which can be written as

$$B_t(R,\theta) = -\mu_0\beta^2 AI\sin\theta\frac{\exp(-\mathrm{j}\beta R)}{4\pi R}, \tag{8.25}$$

where θ is the angle between the loop normal and the radial vector to the observation point (see Figure 8.6b)

We would now like to calculate the average power radiated by a magnetic dipole and note that, from (8.13) and the relationship between $\mathbf{E}$ and $\mathbf{B}$, that

$$P_{\mathrm{rad}} = \frac{c^2}{\eta_0}\int_S |B_t|^2 \mathrm{d}S. \tag{8.26}$$

From (8.25) and (8.26) we obtain that $P_{\mathrm{rad}} = 40\beta^4 A^2|I|^2$ and hence that the radiation resistance will be given by $R_{\mathrm{rad}} = 20\beta^4 A^2$. As far as reactance is concerned, the loop is an inductance with $X_A = \omega\mu_0 L(\ln(4L/\pi a) - 1.75)/2\pi$ where L is the length of wire in the loop. The loss resistance of the loop is given by $R_{\mathrm{loss}} = L/2\pi a\delta\sigma$.

8.3 Reciprocity

The transmission and reception of radio signals can be viewed as a problem in circuit theory with the ports of the transmit and receive antennas regarded as those of a two-port network (see Figure 8.7a), i.e. we can view the problem in terms of an impedance matrix Z with

$$\begin{aligned} V_1 &= Z_{11}I_1 + Z_{12}I_2 \\ V_2 &= Z_{21}I_1 + Z_{22}I_2. \end{aligned} \tag{8.27}$$

We have already seen, in the case of transmission lines, that a propagation medium can be modelled as a combination of series inductors, shunt capacitors and resistors (both series and shunt). Indeed, a general propagation medium can also be modelled as a network of these passive circuit elements and for such a network we will always have that $Z_{12} = Z_{21}$. Consider port 1 driven by current I_1 and then the open circuit ($I_2 = 0$) voltage V_{21} in terminal 2 will be given by $V_{21} = Z_{21}I_1$. Now consider port 2 driven by current I_2 and then the open circuit ($I_1 = 0$) voltage V_{12} in terminal 1 will be given by $V_{12} = Z_{12}I_2$. Since $Z_{12} = Z_{21}$, we will therefore have that

$$\frac{V_{12}}{V_{21}} = \frac{I_2}{I_1} \text{ or } V_{12}I_1 = V_{21}I_2. \tag{8.28}$$

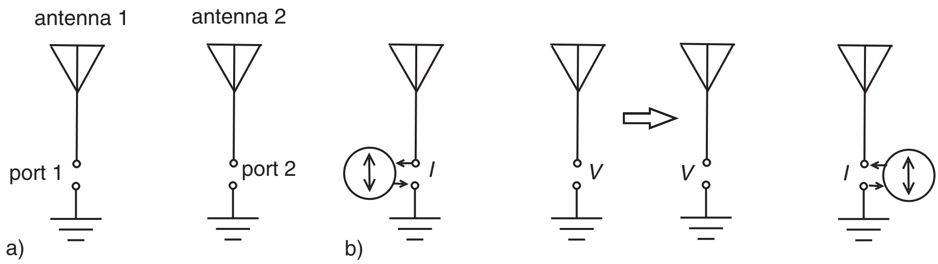


Fig. 8.7 Reciprocity of antennas.

If the ports are driven, in turn, by the same current I, (8.28) implies that the corresponding open circuit will exhibit the same voltage V (see Figure 8.7b). This is known as *reciprocity* and has important consequences for antennas. In particular, it implies that we can infer the two-way performance of an antenna combination from its one-way performance.

In general, the electric field of a radiating source will behave as

$$\mathbf{E} = \mathrm{j}\omega\mu_0 I \mathbf{h}_{\text{eff}} \frac{\exp(-\mathrm{j}\beta R)}{4\pi R}, \tag{8.29}$$

where $h_{\text{eff}} = |\mathbf{h}_{\text{eff}}|$ has the dimensions of length and is known as the *effective length* of the antenna. It is clear that $\mathbf{h}_{\text{eff}}$ can depend on the direction of observation (e.g. in the case of a short dipole $\mathbf{h}_{\text{eff}} = l\hat{\mathbf{R}} \times (\hat{\mathbf{R}} \times \hat{\mathbf{t}})$ and, in the case of a small loop, $\mathbf{h}_{\text{eff}} = -\mathrm{j}\beta A\hat{\mathbf{n}} \times \hat{\mathbf{R}}$). We will consider a two-port system in which one port is provided by an antenna with general effective length $\mathbf{h}_{\text{eff}}$ and the other by a small loop antenna with area A and unit normal $\hat{\mathbf{n}}$. From Faraday's law, the complex amplitude of the open circuit voltage in the terminals of the loop will be

$$\begin{aligned}
V &= \mathrm{j}\omega A\mathbf{B} \cdot \hat{\mathbf{n}} = -\frac{\mathrm{j}\omega A}{c}\hat{\mathbf{R}} \times \mathbf{E} \\
&= \frac{\omega^2\mu_0 IA}{c}(\hat{\mathbf{R}} \times \mathbf{h}_{\text{eff}}) \cdot \hat{\mathbf{n}} \frac{\exp(-\mathrm{j}\beta R)}{4\pi R} \\
&= \frac{\omega^2\mu_0 IA}{c}(\hat{\mathbf{n}} \times \hat{\mathbf{R}}) \cdot \mathbf{h}_{\text{eff}} \frac{\exp(-\mathrm{j}\beta R)}{4\pi R} \\
&= \mathbf{h}_{\text{eff}} \cdot \left(\omega\mu_0\beta IA\hat{\mathbf{n}} \times \hat{\mathbf{R}} \frac{\exp(-\mathrm{j}\beta R)}{4\pi R}\right)
\end{aligned} \tag{8.30}$$

on noting the vector identity $\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B})$. Consequently, $V = \mathbf{h}_{\text{eff}} \cdot \mathbf{E}_{\text{loop}}$ where $\mathbf{E}_{\text{loop}}$ is the electric field of the loop (see (8.23)). Then, by the reciprocity theorem, we also know that this is the open-circuit voltage at the input of the antenna with effective length $\mathbf{h}_{\text{eff}}$ when a current I is driven into the loop antenna. In general, if an antenna of effective length $\mathbf{h}_{\text{eff}}$ is placed in a time-harmonic electric field $\mathbf{E}$, then an open-circuit voltage

$$V = \mathbf{h}_{\text{eff}} \cdot \mathbf{E} \tag{8.31}$$

will appear across its terminals. Once again, the transmitting capability of an antenna is related to its receiving capability. We can now form the circuit model of a receive

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

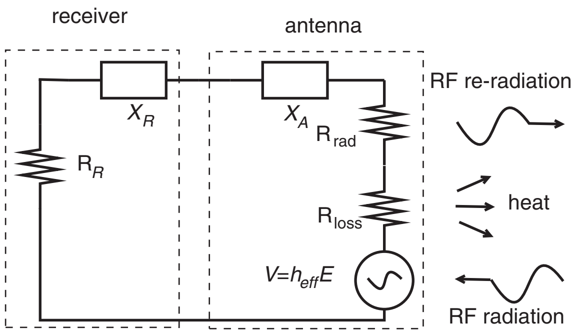


Fig. 8.8 Circuit model of a receive antenna.

antenna shown in Figure 8.8. It will be noted that the antenna has now become a source to the receiver circuit, but there are losses in this source due to heating of the antenna and the reradiation of some of the incoming radio wave. From the expression (8.31), if $\mathbf{h}^1_{\rm eff}$ and $\mathbf{h}^2_{\rm eff}$ are the effective lengths of two antennas, their mutual impedance will be

$$Z_{12} = \mathrm{j}\omega\mu_0 I \mathbf{h}^1_{\rm eff} \cdot \mathbf{h}^2_{\rm eff} \frac{\exp(-\mathrm{j}\beta R_{12})}{4\pi R_{12}}, \tag{8.32}$$

where R_{12} is the distance between the two antennas. Obviously, Z_{11} and Z_{22} will be given by the input impedances of the respective antennas and so we have the complete impedance matrix of a pair of antennas regarded as a two-port device.

8.4 Gain

We have seen that even the simplest of antennas has fields that are highly dependent upon the direction from which they are observed. A useful description of this property is what is known as the *directivity* of an antenna. This is the power radiated in a given direction when scaled upon the average power radiated in all directions. We can evaluate the gain over a sphere at a large distance R from the antenna, and then

$$D(\theta,\phi) = \frac{\bar{\mathbf{E}}(\theta,\phi)\cdot\mathbf{E}(\theta,\phi)}{2\eta_0}\frac{4\pi R^2}{P_{\rm rad}}. \tag{8.33}$$

In the case that the antenna is a short dipole, the directivity will given by $D(\theta,\phi) = \frac{3}{2}\sin^2\theta$ and this will be the same for a small magnetic loop. In the case of a half-wave dipole, the directivity will be given by

$$D(\theta,\phi) = \frac{5}{3}\left(\frac{\cos(\frac{\pi}{2}\cos\theta)}{\sin\theta}\right)^2. \tag{8.34}$$

Directivity tells us something above the direction in which an antenna is sending the power, but it is incomplete as a measure of antenna performance. In particular, it does not tell us anything about the efficiency of an antenna. A more useful measure is known as *gain*. This is the power radiated in a given direction when scaled upon the power accepted by the antenna. The gain G can be related to the directivity D and efficiency

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

e of the antenna through the relation $G = eD$. For an antenna with size of the order of a wavelength (a half-wave dipole for example), the gain is usually fairly close to the directivity. However, as we have already noted, e can be quite low for antennas that are very small in comparison to a wavelength. Due to reciprocity, gain will also provide a measure of how an antenna responds to an incoming signal when operating in the receive mode and we will see further evidence of this when we discuss aperture antennas.

Gain is often represented as a surface that is centred on some origin with the distance of the surface from the origin in a given direction being the value of gain in that direction. In the case of dipoles, the gain surface will be rotationally symmetric about the axis of the dipole and so can be represented by a slice through this surface. Figure 8.9 shows such a slice for a half-wavelength dipole and Figure 8.10 for a 5/4-wavelength dipole. It will be noted that the longer dipole is far more directional than the shorter one. In

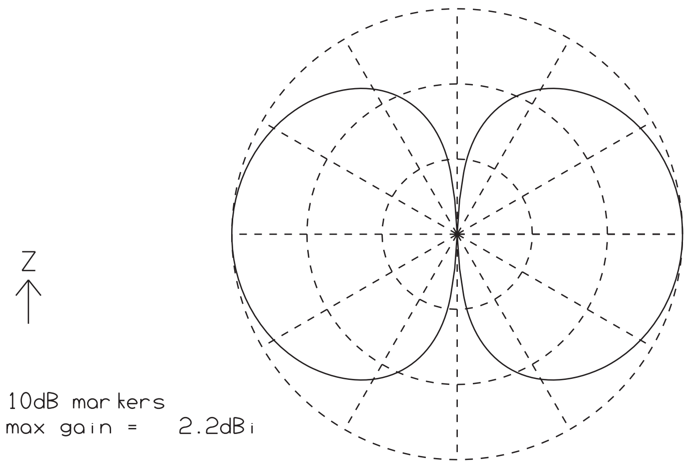


Fig. 8.9 Gain pattern of a half-wavelength dipole.

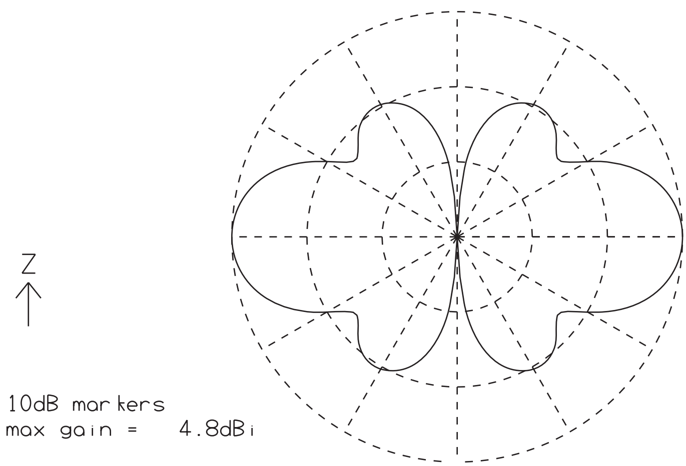


Fig. 8.10 Gain pattern of a 5/4-wavelength dipole.

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

the Figures, the gain G is expressed in terms of *decibels* (or dB for short), a logarithmic scale in which the value of the gain is given by $10\log_{10} G$ (note that the notation dBi is often used to signify dB with respect to an isotropic antenna). Such a scale is far more convenient when a large range of gain values needs to be considered.

8.5 The Monopole Antenna

Thus far, we have assumed that an antenna is completely isolated from other objects, but this is rarely the case. Objects closer than a wavelength can often have a significant impact, affecting both the gain pattern and the antenna input impedance. Consider a radiating current element that is placed above a perfectly conducting ground plane (see Figure 8.11), then there will be a reflection of some of the radiation by the plane (the mirror reflection that we have in the case of light). We could remove the plane and model its effect by an image current element and then the field in the upper space would be the same as the if the ground were still there.

Now consider a vertical rod with its end just above the conducting plane and driven against the plane (see Figure 8.12a). This is known as a monopole antenna and will constitute an unbalanced load to the source that drives it. The feeder will also need to be

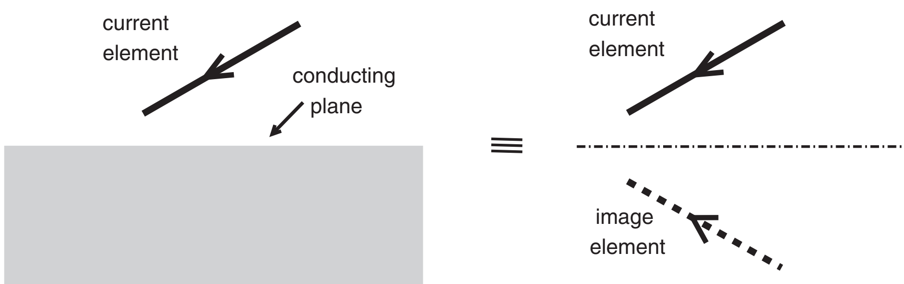


Fig. 8.11 Current element over a conducting plane and equivalent system that uses an image element.

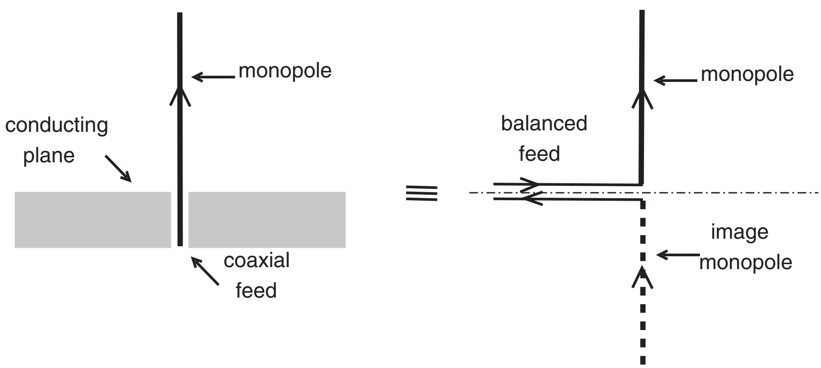


Fig. 8.12 Monopole and equivalent dipole system.

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

unbalanced and this will usually be a coaxial transmission line. Indeed, the monopole antenna can be regarded as the end of a coaxial line in which the centre conductor has become the monopole element and the outer conductor opened out to form the conducting plane. If we replace the conducting plane by an image of the monopole, this will constitute an effective dipole that yields the same field as the monopole in the upper space (see Figure 8.12b). Further, the coaxial feeder will now be replaced by a balanced feeder whose centre is at the same position as the conducting plane. As a consequence, the voltage in the monopole feed V_{mon} will be half that of the effective dipole V_{dip}. However, the current in the monopole I_{mon} will be the same as that in the dipole I_{dip}. The input impedance of the monopole Z_{mon} will therefore be

$$Z_{\text{mon}} = \frac{V_{\text{mon}}}{I_{\text{mon}}} = \frac{\frac{1}{2}V_{\text{dip}}}{I_{\text{dip}}} = \frac{1}{2}Z_{\text{dip}}, \tag{8.35}$$

where Z_{dip} is the dipole input impedance. Consequently, for a short monopole

$$Z_{\text{mon}} = R_{\text{rad}} + R_{\text{loss}} + \mathrm{j}X_A = 40\left(\frac{\pi l}{\lambda}\right) + \frac{l}{6\pi a\delta\sigma} - \mathrm{j}\frac{60}{\beta l}\left(\ln\left(\frac{l}{a}\right) - 1\right). \tag{8.36}$$

The directivity of a monopole will not be the same as that of the equivalent dipole. The radiation into the upper space will be the same, but there will be no radiation into the lower space and hence the total radiated power will be half that of the dipole. As a consequence, the directivity of a monopole will be twice that of the equivalent effective dipole.

In reality, the conducting plane will not be perfect and the image dipole will be distorted. In this case it is simpler to describe the effect of the conducting plane in terms of what is known as a *reflection coefficient*. The electric field of a short monopole (see Figure 8.13a) will be given by

$$E_t(R,\theta) = \mathrm{j}\omega\mu_0 Il(1 + R_g)\sin\theta\,\frac{\exp(-\mathrm{j}\beta R)}{4\pi R}, \tag{8.37}$$

where R_g is the reflection coefficient of the conducting plane and R is the distance from the base of the monopole to the observation point. For a perfectly conducting plane we have $R_g = 1$ but, for the more realistic case of a finite conducting plane, the reflection coefficient R_g will depend on angle θ. For an observation point close to the ground,

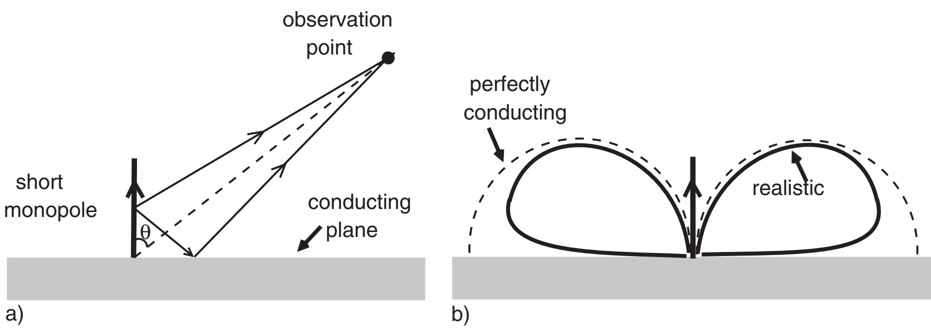


Fig. 8.13 A short monopole and its gain pattern.

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

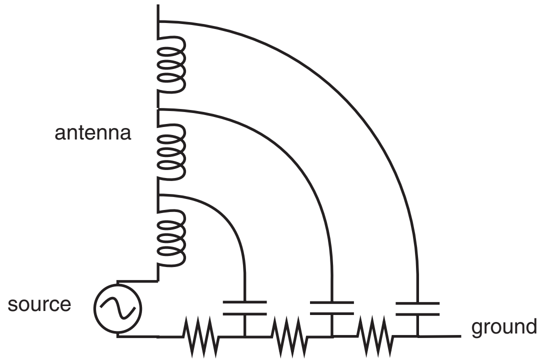


Fig. 8.14 Circuit model of a monopole over a lossy ground.

we will have $R_g \approx -1$ changing to $R_g \approx 1$ at the highest elevations. (Note that we will discuss the reflection coefficient further in the next chapter.)

The effect of a finitely conducting plane upon gain is to make this quantity zero at zero elevation (the elevation of maximum gain in the case of a perfectly conducting plane) and to reduce its value at higher elevations (see Figure 8.13b). Figure 8.14 shows a circuit model of a monopole over a finitely conducting plane. From this it will be seen that the return path for the capacitance is through the resistors that represent the finitely conducting ground. It is the loss in these resistances that reduces the efficiency of the monopole and hence its gain.

8.6 Reducing the Size of Antennas

At lower frequencies, such as those used in broadcasting, a wavelength can be physically large (several hundred metres) and antennas, such as a half-wave dipole, will be far too cumbersome for practical purposes. The simplest approach might be to use a short monopole together with a matching circuit that ensures maximum power is transferred to the antenna (see Figure 8.15a). Another approach might be to use a small resonant loop with suitable matching (see Figure 8.15b). However, as we have seen above, small antennas have a small radiation resistance that can be of the same order as the loss resistance and therefore can be very inefficient. On the other hand, antennas operated close to resonance, such as the half-wave dipole, tend to have a much larger radiation resistance and therefore a much higher efficiency.

If we look at the circuit model of the dipole (see Figure 8.1), we can see a possible approach to achieving a shorter resonant dipole. For a given frequency, we can lower the electrical wavelength $\lambda = c/\omega$ by lowering the propagation speed of the wave on the antenna. Treating the dipole as an opened-out transmission line, we have that $c = 1/\sqrt{LC}$ and so we will need to increase the inductance L, and/or the capacitance C, per unit length. We could increase L by replacing the straight wire with a helical winding (see Figure 8.16a), i.e. the dipole arms become solenoids.

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

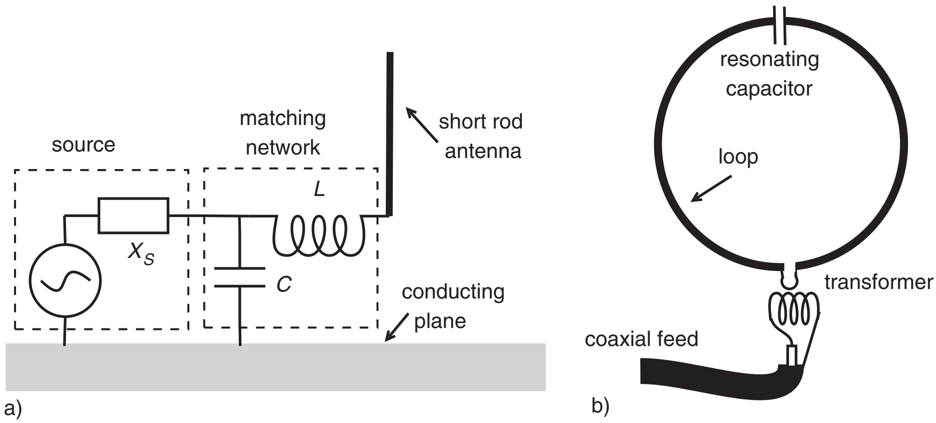


Fig. 8.15 Short monopole and small loop with matching networks.

We can improve on this by making the distribution of current along the wire more uniform and this is done by concentrating the additional inductance at a point further along each dipole arm (see Figure 8.16b). Since we have lumped all the inductance increase into a short section, the wave speed in this section is very low and most of the current drop will occur here. This means that most of the antenna length will support only high current with a consequent increase in effective length. The downside is that we need a lot more additional inductance the further we go out on the arm. (This follows from the fact that, as a parallel wire transmission widens, the capacitance gets weaker and the inductance stronger.) Consequently, to get the necessary level of inductance, the solenoid may need to be made of very thin wire and this will lead to a large increase is loss resistance.

An alternative is to increase the capacitance by means of what is known as a capacity hat (see Figure 8.16c). This can be a large conducting disc at the end of each arm, the disc being orthogonal to the arm. Such an approach overcomes the problem of the loss resistance in a lumped inductance. The downside is that the capacity hats can be quite cumbersome. Despite this, such an approach found favour in the early days of radio. The capacity hat is often made less cumbersome by replacing the conducting discs with a set of radial rods at the ends of the dipole arms.

Besides low efficiency, small antennas bring with them the problem of low bandwidth. The Q of an antenna will be given by $Q = |X_A/(R_{\text{rad}} + R_{\text{ohm}})|$ and this is related to bandwidth B through $B = \omega/Q$. We consider the case of a dipole (length $2l$) with a capacity hat (see Figure 8.16). For the case where $l \ll \lambda$, the current will be uniformly distributed and the antenna will behave as an ideal dipole. The radiation resistance of an ideal dipole is $R_{\text{rad}} = 80\beta^2 l^2$. Further, the reactance will be dominated by the capacitance of the capacity hats and will be given by $X_A = 2l\eta_0/A\beta$. As a consequence, $Q = 3\pi/Al\beta^3$ where A is the area of each capacity hat. In general, a result due to Wheeler (1947) states that Q will be limited by $Q \geq \alpha^{-3}\beta^{-3}$ where α is the radius of the smallest sphere that contains the antenna. This result sets a severe limit on the bandwidth of small antennas.

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

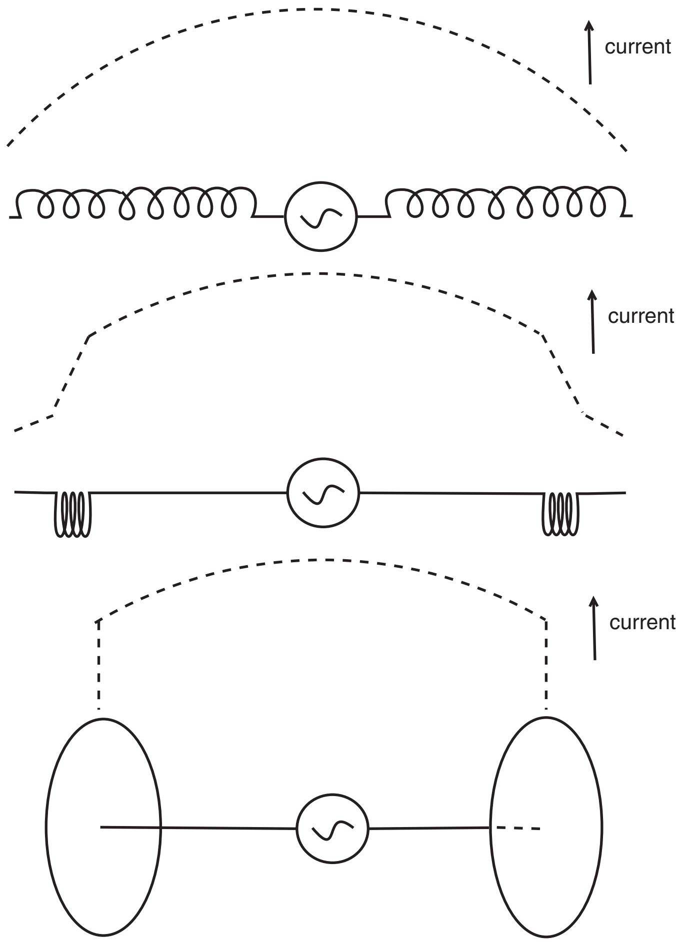


Fig. 8.16 Shortening a dipole antenna.

8.7 Broadband Antennas

A famous result of Rumsey (1966) states that:

If a structure upon scaling by $1/\tau$ is equal to itself, then it will have the same properties at frequencies ω and $\tau\omega$.

It is easiest to see this in the one-dimensional case, i.e. the transmission line (Jones, 1994). From (7.7), the time-harmonic equations for the current I and voltage V are

$$\frac{\mathrm{d}V}{\mathrm{d}z} = -\mathrm{j}\omega L(z)I \text{ and } \frac{\mathrm{d}I}{\mathrm{d}z} = -\mathrm{j}\omega C(z)V, \tag{8.38}$$

where z is the coordinate along the line axis. (Note that C and L are now position-dependent, as would be the case in a transmission line that is opened out into a dipole antenna.) We would like to know the conditions under which the current and voltage at position z, and frequency ω, will be the same as those at position $z' = z/\tau$,

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

and frequency $\omega' = \tau\omega$, i.e.

$$V(z,\omega) = V(z',\omega') \text{ and } I(z,\omega) = I(z',\omega'). \tag{8.39}$$

Noting that $\mathrm{d}/\mathrm{d}z' = \tau\mathrm{d}/\mathrm{d}z$, will have from (8.38) that

$$\frac{\mathrm{d}V}{\mathrm{d}z'} = -\mathrm{j}\omega' L(\tau z')I \text{ and } \frac{\mathrm{d}I}{\mathrm{d}z'} = -\mathrm{j}\omega' C(\tau z')V. \tag{8.40}$$

For Eqs. (8.38) to (8.40) to be consistent we require

$$L(z') = L(\tau z') \text{ and } C(z') = C(\tau z'), \tag{8.41}$$

i.e. L and C should be invariant under the change of scale. As a consequence, the structure of the transmission line needs to be scale-invariant. Consider a parallel-wire transmission line that is opened out (see Figure 8.18), then the radiation from the separate lines will no longer totally cancel and the transmission line will act as an antenna. For the parallel-wire transmission line $L = \mu_0 \cosh^{-1}(D/2a)/\pi$ and $C = \pi\epsilon_0/\cosh^{-1}(D/2a)$, where D is the distance between the wires and a is the wire radius. Consequently, for the transmission to be invariant under the change of scale, we need D/a to be invariant under this change. The simplest way to achieve this is to choose D/a to be constant along the wire, i.e. the wires are conical in shape. The resulting antenna will then have the same properties at all frequencies.

The radiation will introduce loss into the transmission line and so it is better described by

$$\frac{\mathrm{d}V}{\mathrm{d}z} = -R(z)I - \mathrm{j}\omega L(z)I \text{ and } \frac{\mathrm{d}I}{\mathrm{d}z} = -\mathrm{j}\omega C(z)V, \tag{8.42}$$

where R is the radiation resistance of a unit length of the transmission line. Only current orthogonal to the line axis will contribute (radiation from the horizontal components still cancels) and so, from previous considerations of an ideal dipole, $R \approx 80\beta^2 \sin^2\gamma$ where γ is the angle between the wire axis and the z axis (γ is assumed to be small). The current distribution on the upper wire of the transmission line will now take the form

$$I(z) = I_0 \exp(-\mathrm{j}\beta z - \alpha z), \tag{8.43}$$

where z is now the distance along the wire (the current on the lower wire is $-I$). Assuming the radiation loss to be small, the propagation constant along the line will be $\beta \approx \omega\sqrt{LC} = \omega/c_0$, the impedance will be $Z_0 \approx \sqrt{L/C} = \eta_0/\pi \cosh^{-1}(D/2a)$ and the attenuation constant will be $\alpha \approx R/2Z_0 = \beta^2 \sin^2\gamma/3\cosh^{-1}(D/2a)$ (on noting that $\eta_0 = 120\pi$). The important thing to note is that the current will decay along the wire due to radiation and that, as a consequence, it is possible to truncate the antenna. Over a length $l_{\text{atten}} = 3\cosh^{-1}(D/2a)/\beta^2 \sin^2\gamma$ the current amplitude will drop by the factor $1/e \approx 0.36788$. Consequently, once we know the lowest frequency ω_L at which the antenna is required to operate, we can then truncate it at a length for which the missing section will carry negligible current for the frequencies above ω_L. We will then have an antenna with approximately the same properties for frequencies above ω_L.

When the wires spread out to form a pair of cones with a common axis (i.e. $\gamma = \pi/2$), we have what is known as a biconical dipole (see Figure 8.17). This dipole has an input

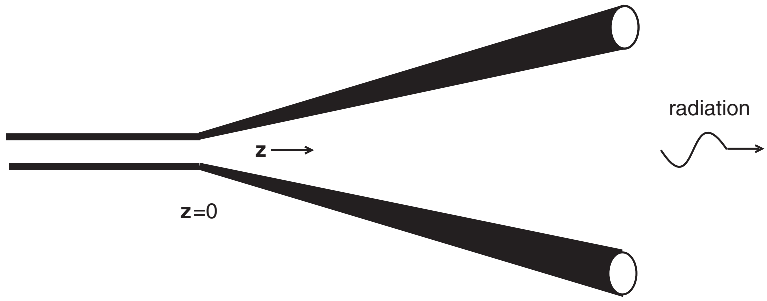


Fig. 8.17 Radiating transmission line.

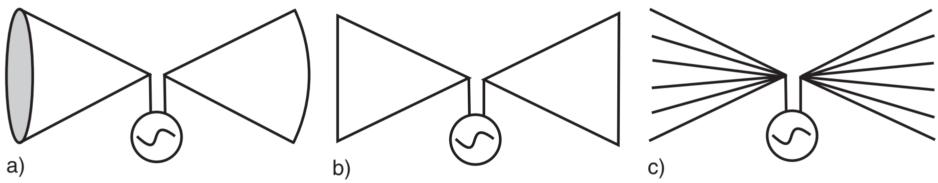


Fig. 8.18 Biconical and bow-tie antennas.

impedance $Z_{\text{in}} = 120 \ln(\cot(\theta_h/2))$ where θ_h is the cone half angle. Once again, because of radiation losses, the current will be effectively zero after a distance of just over half a wavelength along each arm of the dipole. Consequently, the antenna radiation pattern will behave like a normal dipole up to a frequency where its length is just over a wavelength and thereafter its pattern will stabilise.

Another antenna that satisfies the above scaling condition is the infinite bow-tie dipole, the planar version of the biconical dipole. Once again, this can only be realised in its truncated form (see Figure 8.17b). A metal cone, or bow tie, can be rather cumbersome, but fortunately we can replace such structures by a series of rods that spread out from the feed (see Figure 8.17c for the case of a finite bow tie). In general, it is found that metal objects that contain gaps of less than $\lambda/10$ behave as if impervious to electromagnetic waves. Further, from the above considerations, currents are only significant out to a finite distance from the antenna feed. Consequently, providing the rods are sufficiently dense, the fact that they spread out from the feed does not stop them acting as if impervious.

It is also possible to form a conical monopole (see Figure 8.19a). As with the rod monopole, the input impedance is half that of the equivalent dipole and the gain is double. A further development of the monopole is the discone antenna. In this, the ground plane is replaced by a disc (see Figure 8.19b). As with the biconical dipole, it is possible to replace the cone, and the disc, by rods that spread out radially from the feed. In the case of the disc, it is usually sufficient for the rods to spread out to the same extent as the cone base.

A further variety of antenna that has broadband properties is the *travelling-wave antenna*. Consider a parallel-wire transmission line that has been opened out to form a

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

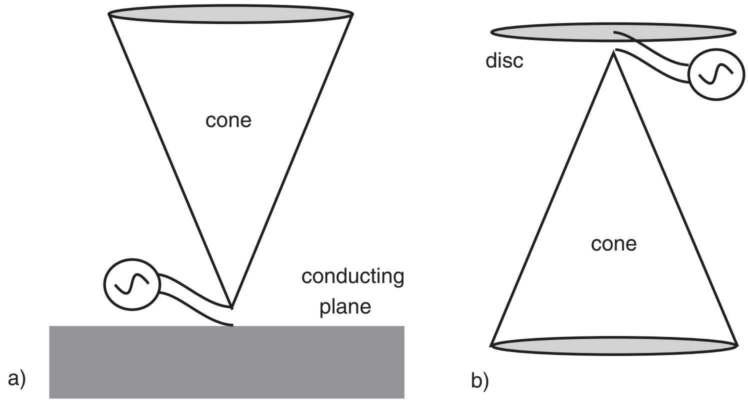


Fig. 8.19 Conical monopole and discone antennas.

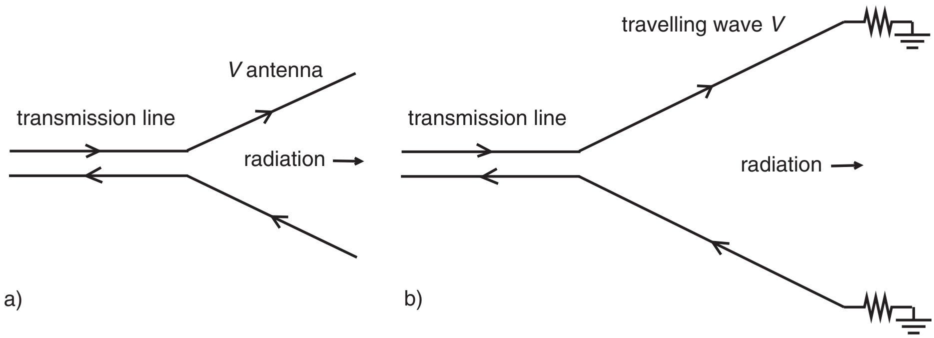


Fig. 8.20 V and travelling-wave V antennas.

V-shaped antenna (see Figure 8.20a), this will have a specific frequency at which it is resonant (i.e. it presents a real impedance to the transmission line) and away from this frequency the antenna will present some reactive impedance. From transmission-line theory, the resonance will occur at a frequency for which the arms of the V are approximately one quarter of a wavelength long. Away from this frequency, the waves travelling into the antenna will be reflected back at the antenna ends and cause a reactive impedance at the antenna input. In the case of a transmission line, we can prevent these reflections by placing a matched load across the end of the line and, in a travelling-wave antenna, we can do the same. As shown in Figure 8.20b, we load each arm of the antenna by connecting it to the ground through half the characteristic impedance (we can do this since the antenna is balanced). The antenna will then present a real impedance at all frequencies. The downside of such an antenna, however, is that it can be lossy at low frequencies. At the higher frequencies the decay in current due to radiation means that very little power will reach the loads. However, as the frequency drops, more and more power will reach the loads and so less and less will be radiated.

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

8.8 Array Antennas

So far we have only considered some fairly simple antenna elements, but it is possible to manufacture antennas with more complex gain patterns through arrays of such elements. Some insight into this process can be gained by considering a two-element array in the receive mode (remember the reciprocity between the behaviour of antennas in receive and transmit modes). Consider the configuration shown in Figure 8.21 and a harmonic wave of complex amplitude E_r arriving at the rear element from the rear. For an antenna with effective length h_{eff}, this will result in a voltage $h_{\text{eff}} E_r \exp(-\mathrm{j}\beta(L+S+D))$ at the combiner due the front element and a voltage $h_{\text{eff}} E_r \exp(-\mathrm{j}\beta L)$ arriving at the combiner due to the rear element (the signals travel path lengths $L+S+D$ and L respectively). The combined voltage at the output of the combiner will now be

$$V_r = h_{\text{eff}} E_r \exp(-\mathrm{j}\beta L)(\exp(-\mathrm{j}\beta(S+D))+1). \tag{8.44}$$

In a similar fashion, for a field E_f arriving at the front from the front,

$$V_f = h_{\text{eff}} E_f \exp(-\mathrm{j}\beta L)(\exp(-\mathrm{j}\beta S)+\exp(-j\beta D)). \tag{8.45}$$

A harmonic wave of amplitude E_s arriving from the side of the array will reach both elements simultaneously and this will result in a voltage

$$V_s = h_{\text{eff}} E_s \exp(-\mathrm{j}\beta L)(\exp(-\mathrm{j}\beta S)+1) \tag{8.46}$$

at the output of the combiner. We first make the feeds to the combiner of equal length (i.e. $S=0$) in order to generate a zero-phase increment between the elements. Then, if we make the separation between the antennas a half wavelength (i.e. $D=\lambda/2$), we will have $V_r = h_{\text{eff}} E_r \exp(-\mathrm{j}\beta L)(\exp(-\mathrm{j}\pi)+1) = 0$, $V_f = h_{\text{eff}} E_f \exp(-\mathrm{j}\beta L)(1+\exp(-\mathrm{j}\pi)) = 0$ and $V_s = 2h_{\text{eff}} E_s \exp(-\mathrm{j}\beta L)$ (note that $\beta = 2\pi/\lambda$). In other words, we have zero response to the front and rear, but enhanced response to the sides. From the terminology of battleships, the array is known as a *broadside array* since, in the transmit mode, it fires off orthogonal to the axis of the array. Greater detail can be found in Figure 8.22 which

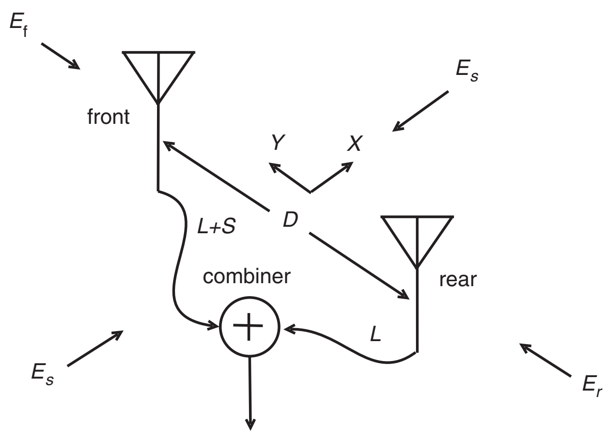


Fig. 8.21 Array of two elements in receive configuration.

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

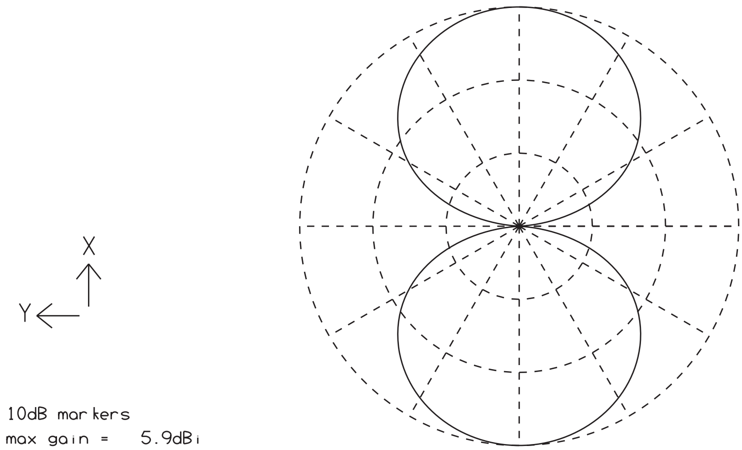
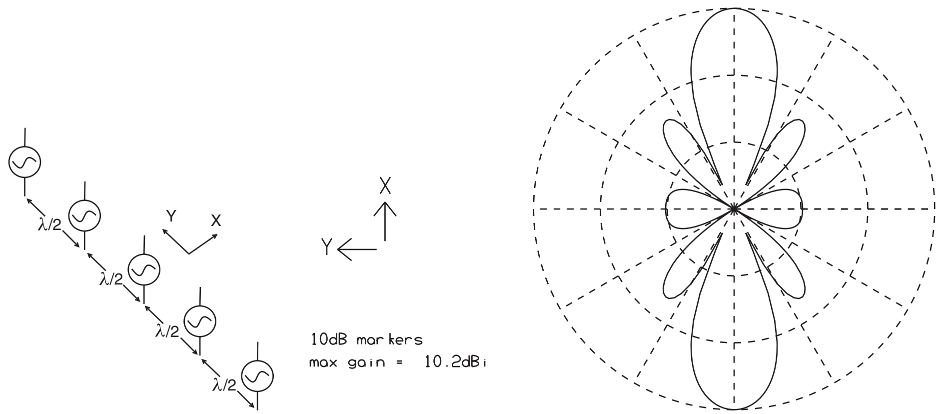


Fig. 8.22 Gain pattern of a two-dipole broadside array with $\lambda/2$ spacing.



a) array b) array gain

Fig. 8.23 Gain pattern of a five-dipole broadside array with $\lambda/2$ spacing.

shows the gain pattern derived from a numerical simulation of the array. Figure 8.23 shows the gain pattern for a broadside array with five elements ($D = \lambda/2$ and a zero phase increment between the elements). From this simulation, it will be noted that there are now several *side lobes* in addition to the main lobe and these can have significant amplitude. Importantly, it will be noted that there has been a significant narrowing of the main lobe with a consequent increase in gain. Obviously, if the angle through which the major part of the radiation flows becomes narrower, the density of power within this angle is greater. For a broadside array, the angle between the nulls either side of the main lobe will be λ/L radians where $L = ND$ is the effective length of the array, N is the number of elements and D is their spacing. Further, the gain will increases by the multiplicative factor $2ND/\lambda$. At this point one might be tempted to increase the spacing between the elements in order to make a narrower beam and increase the gain. However, this can lead to unwanted consequences as illustrated in Figure 8.24 for the case of a wavelength

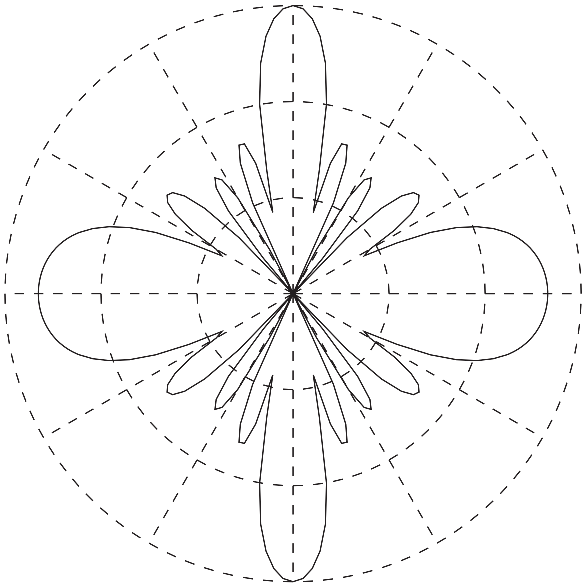


Fig. 8.24 The effect of increasing spacing to λ.

spacing between the elements. Although the beam has indeed narrowed and the gain increased, there is now strong lobes from either end of the array. These lobes, known as *grating lobes*, can give the array strong response in undesired directions. It turns out that $\lambda/2$ is the practical limit of element spacing if grating lobes are to be avoided.

For the two-element array, if we now make the feeds to the combiner differ in length by half a wavelength (i.e. $S = \lambda/2$), there will be a phase increment of π between the elements and we will find that $V_s = 0$, $V_f = 2h_{\text{eff}}E_f \exp(-\mathrm{j}\beta(L + \lambda/2))$ and $V_r = 2h_{\text{eff}}E_r \exp(-\mathrm{j}\beta L)$. We now have zero response to the sides and equal magnitude response to the fore and rear. From the terminology of battleships, the array is known as an *endfire array* since, in the transmit mode, it now fires off along the axis of the array. Greater detail can be found in Figure 8.25 which shows the gain pattern derived from a numerical simulation of the array. The important thing to note is that, by varying the phase increment between then elements from 0 to π, we have *steered* the array from broadside to endfire. As a consequence, although we could steer the main lobe by physically rotating a broadside array in the *xy* plane, we can also steer the array by changing the phase increment between the array elements. Besides adjusting the phase across the array, the distribution of current amplitude across the array can also be adjusted as a further means of manipulating the array gain pattern. Whilst the phase can be used to direct the array main lobe, the current amplitude can be used to control the level of side lobes and hence make a much more directionally selective antenna. Figure 8.26 shows the gain pattern of a five-element broadside array with a Dolph-Chebyshev taper (i.e. current amplitudes distributed in proportion to 1:2.41:3.14:2.41:1 across the array) and from which it will be noted that the side lobes have almost been eliminated. The downside, however, is that the main lobe is now slightly wider.

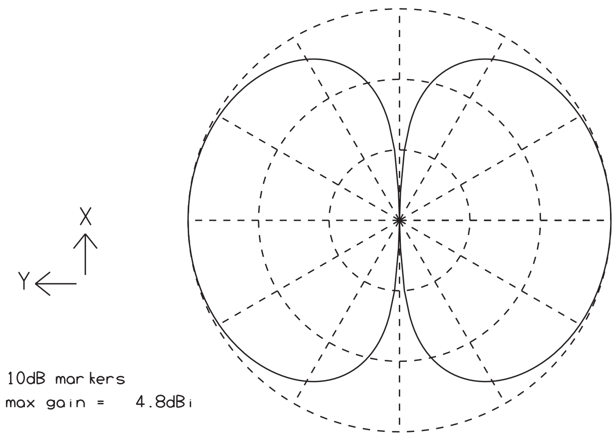


Fig. 8.25 Gain pattern of a two-dipole endfire array with $\lambda/2$ spacing.

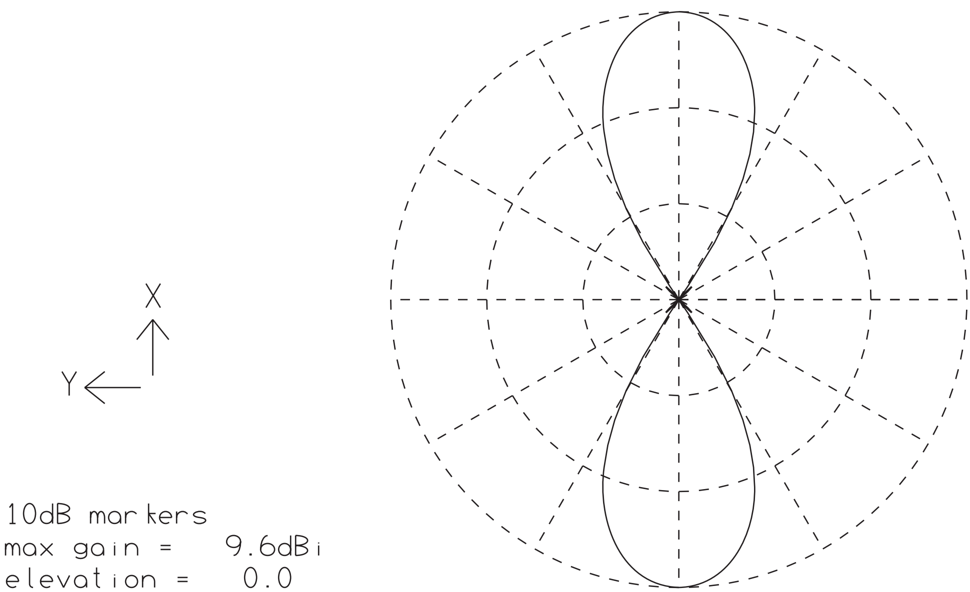


Fig. 8.26 Gain pattern of a five-element broadside array with a Dolph-Chebyshev taper.

We have already seen that there are problems with element spacing greater than $\lambda/2$. However, there turns out to be some utility in making the spacing less than $\lambda/2$, even though there will be some loss in gain and an increase in beamwidth. Consider the two element receive array and make a separation of a quarter wavelength between the array elements (i.e. $D = \lambda/4$). We also make the feeds to the combiner differ in length by a quarter wavelength (i.e. $S = \lambda/4$) to obtain a $\pi/2$ phase increment between the array elements. Then, from (8.44) to (8.46), we obtain that $V_r = 0$, $V_f = -2jh_{\text{eff}}E_f \exp(-j\beta L)$ and $V_s = (1-\text{j})h_{\text{eff}}E_s \exp(-\text{j}\beta L)$, i.e. we have a zero response to the rear, an enhanced response to the fore and a reduced response to the sides. Essentially, the array has only

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

one main lobe and this can be seen in Figure 8.27 which shows the gain pattern derived from a numerical simulation. Unlike the above simple theory, the simulations exhibit a non-zero gain to the rear which is, nevertheless, small in comparison to the main lobe. This effect arises due to inter-element coupling that is taken into account by the simulations, but not by the simple theory.

As we have seen above, inter-element interactions can have a significant effect upon the radiation pattern of arrays and must be given serious consideration, especially when the elements are close in comparison to a wavelength. The effect of element interaction can be seen from the simple case of two interacting half-wave dipoles. Figure 8.28 details the interaction of two parallel dipoles through their mutual impedance Z_{12} (itself equal to Z_{21} by symmetry). The mutual impedance is shown as a function of distance between the dipole centres and it will be noted that Z_{12} can be of a similar order to the self impedances Z_{11} (itself equal to Z_{22} by symmetry). It can be seen that the interaction can be quite strong for separations less than half a wavelength. The strong mutual interaction can be used to create arrays in which only one element is directly driven and several *parasitic elements* are driven through their mutual interaction with the driven element. Figure 8.30a shows an example of such an array, known as a Yagi-Uda array after its inventors. The array that is shown has three elements, known as the reflector, driven and director elements respectively. We can estimate the currents excited in the reflector and director element through their mutual impedance with the driven element. Assuming the input of the driven element to be port 1, driven by voltage V_1, there will be zero voltage V_2 at the centre of a parasitic element that is taken to be port 2. We will then have that

$$\begin{aligned} V_1 &= Z_{11}I_1 + Z_{12}I_2 \\ 0 &= Z_{21}I_1 + Z_{22}I_2, \end{aligned} \tag{8.47}$$

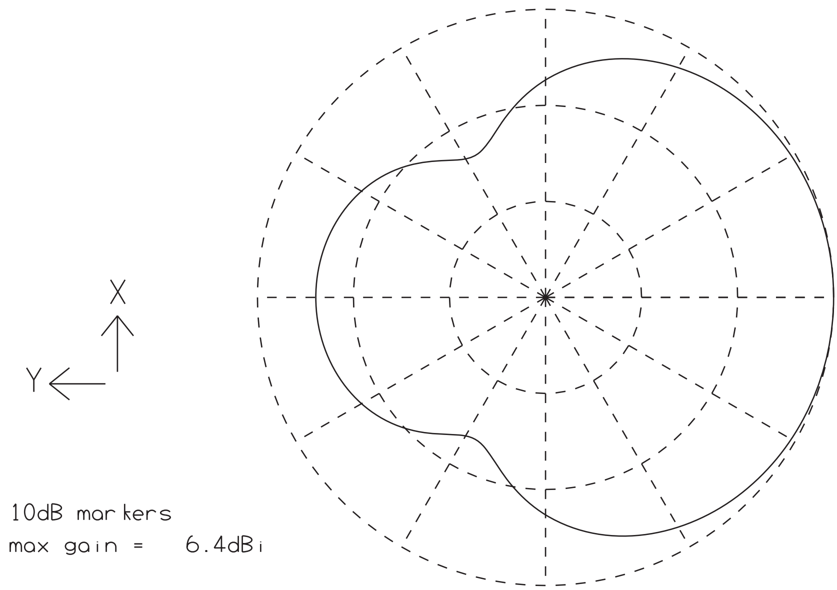


Fig. 8.27 Gain pattern of a two-dipole array with $\lambda/4$ spacing and $\pi/2$ phase increment.

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

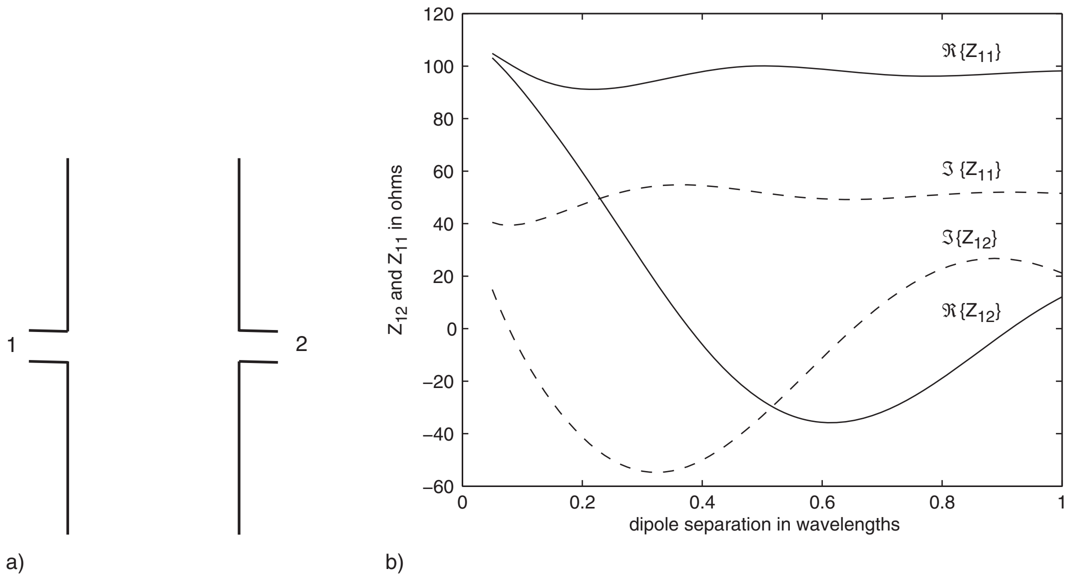


Fig. 8.28 Impedance and mutual impedance of parallel dipoles at various separations.

where I_1 and I_2 are the currents flowing into ports 1 and 2 respectively. Then, from the above equations, we have that $I_2 = -(Z_{12}/Z_{22})I_1$. For element spacings of around a quarter wavelength, we have already seen that we can produce an endfire array with only one main lobe. Further, from Figure 8.28, the mutual impedance will be of the order of self impedance and so the current flowing in the parasitic element will be of the same order of magnitude as that flowing in the driven element. However, to obtain endfire behaviour, we need to ensure the correct phase increment between the driven and parasitic elements. We can achieve this by adjusting the length of the parasitic elements (see Figure 8.29). In order to produce radiation in the direction from the reflector to director elements, the reflector element needs to be larger than the driven element and the director element needs to be shorter.

Figure 8.31 shows the gain pattern in the plane of the antenna for an array with quarter-wavelength spacings between the elements and parasitic element lengths that vary by five per cent from the length of the driven element. It will be noted that, when compared with a dipole, there is substantial narrowing of the main lobe and a consequent increase in gain. This performance can be further enhanced by introducing additional director elements.

Being essentially an extension of the dipole, the Yagi-Uda array is inherently narrowband. To obtain a broadband array, we could create an array of broadband elements such as bicones or discones. However, the Rumsey result of the previous section introduces another possibility. If we consider a structure that is equal to itself upon scaling by $1/\tau$, we will have a structure that has the same properties at frequencies $\tau\omega, \tau^2\omega, \tau^3\omega, \ldots$ (the antenna is said to be a *log-periodic* structure). Providing that τ is close to 1, this structure will have almost the same properties on all frequencies. A

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

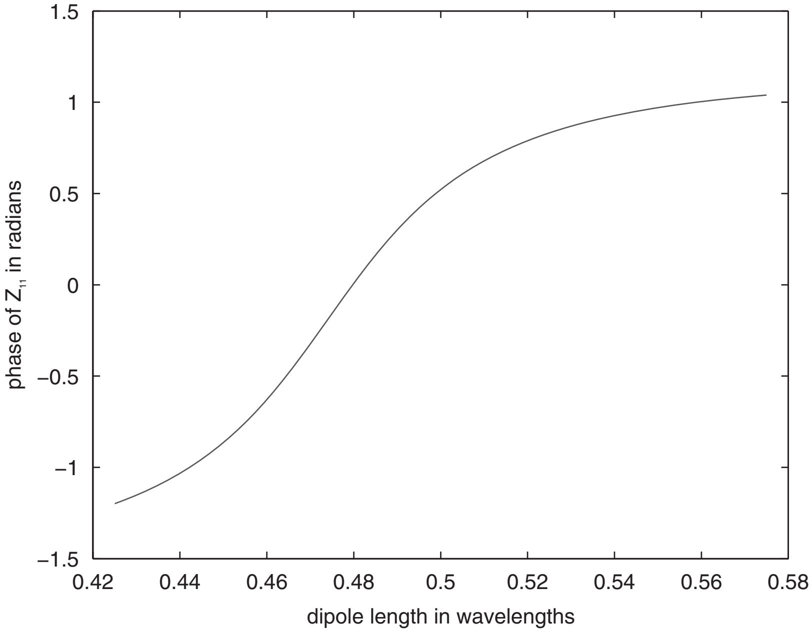


Fig. 8.29 Phase of dipole input impedance as a function of antenna length.

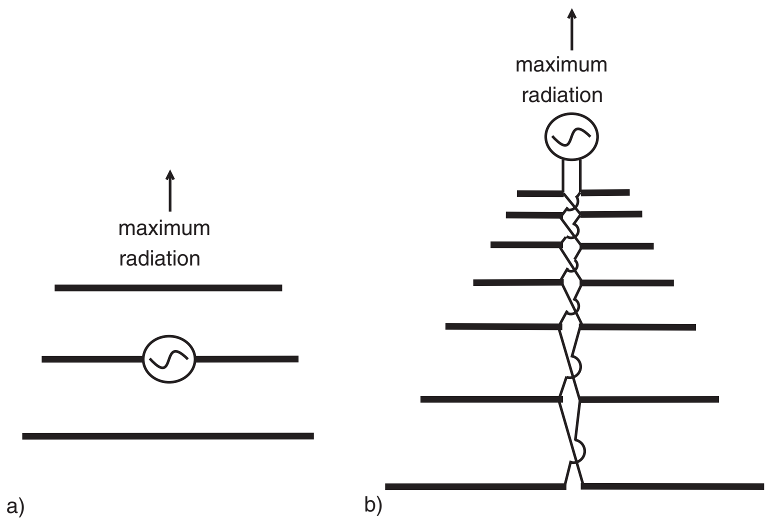


Fig. 8.30 Yagi-Uda and log-periodic arrays.

practical antenna based on this principle is shown in Figure 8.30b. This antenna consists of a series of interconnected dipoles for which the dipole lengths, and their spacings, are scaled by $1/\tau$ as we move between elements along the antenna axis from the longest element to shortest. This is an endfire array that is known as a *log-periodic dipole*

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

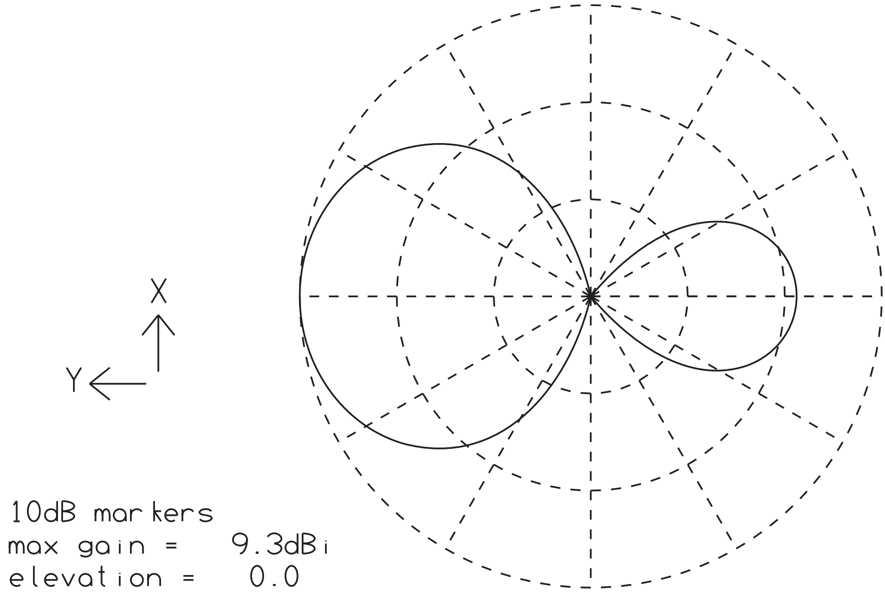


Fig. 8.31 Gain pattern of a three-element Yagi-Uda array.

array (LPDA). Obviously, the Rumsey result requires an infinite number of dipoles and the finite number in the practical implementation could have a serious impact upon performance. Fortunately, it turns out that when the LPDA is driven in the manner shown, only the dipoles that are close to resonance will be active and so the truncation can be chosen in line with a desired frequency range. The antenna can operate effectively between the resonant frequencies of the shortest and longest dipoles and will produce a gain pattern that is similar in nature to that of a three to four element Yagi-Uda antenna.

8.9 Aperture Antennas

A means of transferring energy between two points is through what is known as a *hollow waveguide*. In fact, transmission lines are themselves waveguides, but they always have two conductors and can therefore be analysed in terms of traditional circuit ideas. This is not the case for a hollow waveguide. We consider a waveguide consisting of a metal duct with rectangular cross section (see Figure 8.32a) and analyse wave propagation inside this duct. We first note that, on the surface of a perfect conductor (an effective model of a metal), the tangential electric field will be zero. We have seen earlier that, due to the skin effect, RF currents travel close to the surface of a conductor and are therefore tangential to it. Consequently, since the resistance of a perfect conductor is zero, Ohm's law implies that the voltage drop between any two points on the surface of the conductor must also be zero, i.e. the tangential electric field is zero. This has important implications for waves travelling down our duct.

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

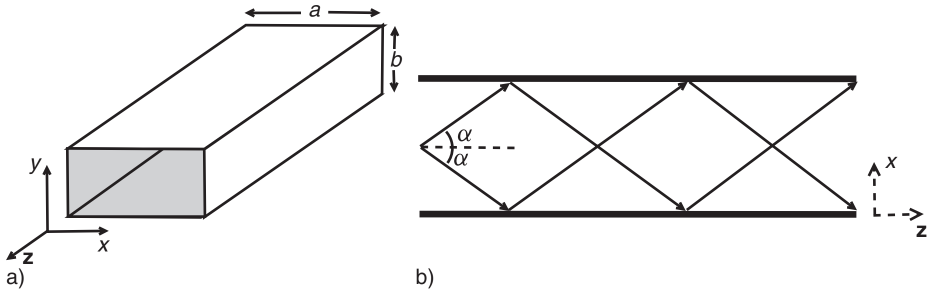


Fig. 8.32 A rectangular waveguide.

We first ask whether a plane EM wave can travel down the waveguide and consider the time-harmonic wave $\mathcal{E} = \Re\{\exp(j\omega t)\mathbf{E}\}$ where

$$\mathbf{E} = E_0 \exp(-j\beta z)\hat{\mathbf{y}} \tag{8.48}$$

where β is the propagation constant in free space. This field is orientated in the $\hat{\mathbf{y}}$ direction and therefore has no component tangential to the horizontal faces of the waveguide. There is, however, a non-zero tangential component at the vertical faces. As a consequence, a plane wave cannot travel down a rectangular waveguide. Consider now an electric field that consists of a combination of two plane waves that travel at angle α to the horizontal, i.e.

$$\mathbf{E} = (\exp(-j\beta(-x\sin\alpha + z\cos\alpha) - \exp(-j\beta(x\sin\alpha + z\cos\alpha))\, E_0\hat{\mathbf{y}} \tag{8.49}$$

These waves are reflected back and forth between the vertical sides of the waveguide (see Figure 8.32b). Field $\mathbf{E}$ will have zero tangential component on the horizontal sides of the waveguide and on the vertical side at $x = 0$. On the vertical side at $x = a$, the field will have zero tangential component if $\exp(j\beta a\sin\alpha) = \exp(-j\beta a\sin\alpha)$, i.e. if $\sin(\beta a\sin\alpha) = 0$. Consequently, $\mathbf{E}$ will have tangential components that are zero on all surfaces of the waveguide if $\beta a\sin\alpha = n\pi$ where n is an arbitrary integer. There is obviously an infinite set of wave modes that can travel down a rectangular waveguide and, from (8.49), the electric field will have the form

$$\mathbf{E} = 2jE_0 \sin\left(\frac{n\pi x}{a}\right)\exp(-j\beta' z)\hat{\mathbf{y}}, \tag{8.50}$$

where $\beta' = \beta\cos\alpha = \sqrt{\beta^2 - \pi^2 n^2/a^2}$ is the propagation constant in the waveguide. In fact, the above modes are a subset of a doubly infinite set of possible propagation modes with $\beta' = \sqrt{\beta^2 - \pi^2(n^2/a^2 + m^2/b^2)}$ the propagation constant for the nm mode. The nm mode is denoted by TE_{nm} where TE is used to indicate the fact that the electric field is transverse to the propagation direction (note that the magnetic field is not transverse). It will be noted that β' is frequency dependent and this dependence increases with mode number. Consequently, to reduce the problems of dispersion, it is usual to design a waveguide so that it only carries the TE_{10} mode and this can be achieved in the following fashion. Since $\beta = \omega/c$, it is clear that for mode TE_{nm} there is a frequency ω_{nm} below which the propagation constant becomes imaginary and no wave will propagate (in fact, the field decays exponentially and is said to be *evanescent*). Now consider frequencies ω such that

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

$\omega > \omega_{10}, \omega < \omega_{20}$ and $\omega < \omega_{01}$ (we have assumed that $a > b$). Then, for these frequencies, the TE_{01} mode is the only one that can propagate. A wave can be excited in a waveguide by placing a monopole antenna inside the guide (see Figure 8.33a). The monopole is placed a half wavelength from the end of the waveguide so that the reflected wave coincides with the direct wave (it should be noted that a wavelength in this case refers to that of the TE_{10} mode, i.e. $\lambda_{10} = 2\pi/\beta_{10}$ where β_{10} is the propagation constant of the TE_{10} mode). Although the monopole field will have a complex pattern (it will be a mixture of many modes), this will be reduced to the TE_{10} mode alone only a short distance down the waveguide due to the decay of evanescent modes. An open-ended waveguide can be used as an antenna (see Figure 8.33a) when the radiation is emitted into free space from the end of the guide. However, in order derive the radiation field we need to develop a little more theory.

Consider the electric field of an antenna at a point P, i.e.

$$E = \mathrm{j}\omega\mu_0 I h_{\mathrm{eff}} \frac{\exp(-\mathrm{j}\beta D)}{4\pi D}, \tag{8.51}$$

where D is the distance from the antenna to point P. We consider a surface S that is orthogonal to the line joining the antenna to P (i.e. the plane is tangent to the wavefront), located so that it is a large distance D_1 from the antenna (see Figure 8.34a). Consider

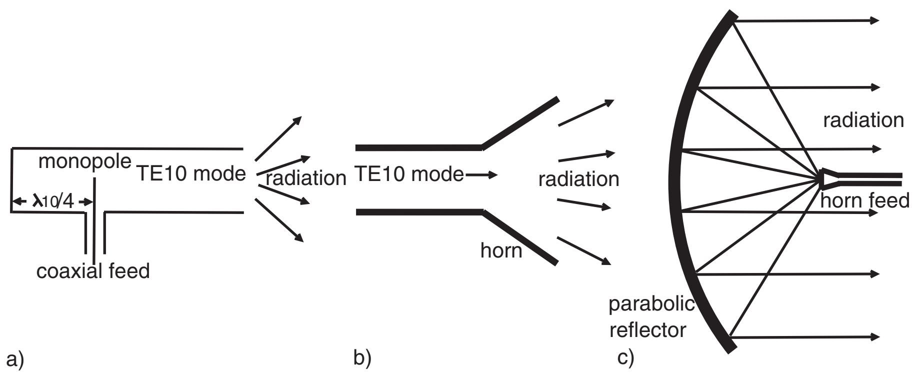


Fig. 8.33 Waveguide, horn and parabolic-reflector antennas.

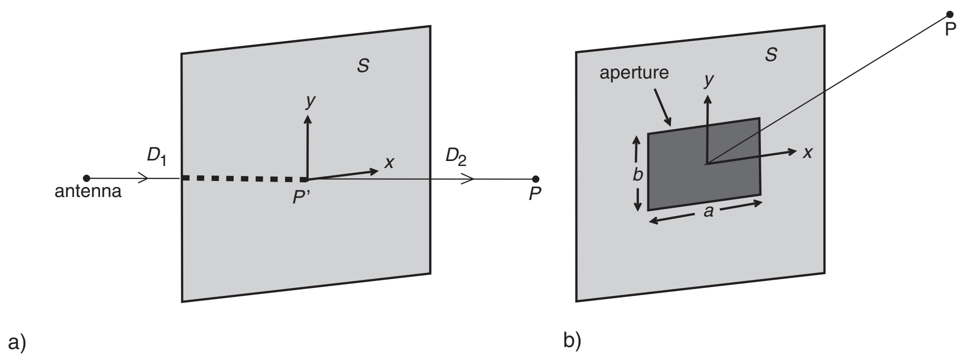


Fig. 8.34 Geometry for aperture theory.

local Cartesian coordinates x and y in the plane, centred on the point of intersection of the line and plane. Then, for a point in the plane, the distance from the antenna is $\sqrt{x^2+y^2+D_1^2} \approx R_1+(x^2+y^2)/2D_1$ and so the field at this point can be approximated as

$$E = \mathrm{j}\omega\mu_0 I h_{\mathrm{eff}} \frac{\exp\left(-\mathrm{j}\beta\left(D_1+\frac{x^2+y^2}{2D_1}\right)\right)}{4\pi D_1}. \tag{8.52}$$

Now define a potential

$$\psi = \mathrm{j}\beta \frac{\exp(-\mathrm{j}\beta\tilde{R})}{2\pi\tilde{R}} \tag{8.53}$$

for a source that is located at P ($\tilde{R}$ is the distance from the source to a general point). For points that lie in the surface S, $\tilde{R} = \sqrt{x^2+y^2+D_2^2} \approx D_2+(x^2+y^2)/2D_2$, where D_2 is the distance of P from the plane (note that $D = D_1+D_2$). Consequently, potential ψ can be approximated by

$$\psi = \mathrm{j}\beta \frac{\exp\left(-\mathrm{j}\beta\left(D_2+\frac{x^2+y^2}{2D_2}\right)\right)}{2\pi D_2}. \tag{8.54}$$

Now consider the integral $\int_S \psi E \mathrm{d}S$, then the integrand is given by

$$\psi E = -\beta\omega\mu_0 I h_{\mathrm{eff}} \frac{\exp(-\mathrm{j}\beta\left(D_1+D_2+(x^2+y^2)\left(\frac{1}{2D_1}+\frac{1}{2D_2}\right)\right)}{8\pi^2 D_1 D_2}. \tag{8.55}$$

Consequently,

$$\int_S \psi E \mathrm{d}S = -\beta\omega\mu_0 I h_{\mathrm{eff}} \frac{\exp(-\mathrm{j}\beta(D_1+D_2))}{8\pi^2 D_1 D_2} \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\mathrm{j}\beta(x^2+y^2)\left(\frac{1}{2D_1}+\frac{1}{2D_2}\right)\right)\mathrm{d}x\mathrm{d}y. \tag{8.56}$$

We note that

$$\int_{-\infty}^{\infty} \exp(-\mathrm{j}\alpha x^2)\mathrm{d}x = \sqrt{\frac{\pi}{\mathrm{j}\alpha}} \tag{8.57}$$

from which

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\mathrm{j}\beta(x^2+y^2)\left(\frac{1}{2D_1}+\frac{1}{2D_2}\right)\right)\mathrm{d}x\mathrm{d}y = \frac{\pi}{\mathrm{j}\beta}\frac{1}{\frac{1}{2D_1}+\frac{1}{2D_2}} = \frac{2\pi}{\mathrm{j}\beta}\frac{D_1 D_2}{D_1+D_2} \tag{8.58}$$

and, as a consequence,

$$\int_S \psi E = \mathrm{j}\omega\mu_0 I h_{\mathrm{eff}} \frac{\exp(-\mathrm{j}\beta(D_1+D_2))}{4\pi(D_1+D_2)}. \tag{8.59}$$

The right-hand side of (8.59) is the field at point P (note that $D = D_1+D_2$) and so

$$E = \int_S \psi E \mathrm{d}S. \tag{8.60}$$

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

Although we have only derived (8.60) for a point P located at $D_2\hat{\mathbf{z}}$, (8.60) is still valid for small deviations from this (i.e. points $\mathbf{R} = X\hat{\mathbf{x}} + Y\hat{\mathbf{y}} + D_2\hat{\mathbf{z}}$ for which $|X| \ll D_2$ and $|Y| \ll D_2$).

We now consider the situation where no radiation passes through plane S, except for an aperture A (this could be the situation where S is an opaque screen that contains a hole). E is now the field of the radiation emanating from this aperture and its value at point P will be given by (8.60) with the integral restricted to A, i.e.

$$E = \int_A \psi E \mathrm{d}S. \tag{8.61}$$

This expression is sometimes known as the *Kirchhoff integral* (Kirchhoff, 1883). Assuming that the aperture is small in comparison to the distance from point P, the distance from a point $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}}$ within the aperture to point P can be approximated as $\sqrt{(\mathbf{R}-\mathbf{r})\cdot(\mathbf{R}-\mathbf{r})} = R\sqrt{1-2\mathbf{r}\cdot\mathbf{R}/R^2 + r^2/R^2} \approx R - (xX + yY)/R$ where R is the distance of P from the origin and $r = |\mathbf{r}|$. Consequently, $\psi \approx \mathrm{j}\beta \exp(-\mathrm{j}\beta R)\exp(\mathrm{j}\beta(xX + yY)/R)/\pi R$ and equation (8.61) will now reduce to

$$E(\mathbf{R}) = \mathrm{j}\beta\frac{\exp(-\mathrm{j}\beta R)}{2\pi R}\int_A E(x,y,0)\exp\left(\mathrm{j}\beta\frac{xX+yY}{R}\right)\mathrm{d}x\mathrm{d}y. \tag{8.62}$$

Consider a rectangular aperture (see Figure 8.34b) with sides length a and b, then the above expression will imply that

$$E(\mathbf{R}) = \mathrm{j}\beta\frac{\exp(-\mathrm{j}\beta R)}{2\pi R}\int_{-b/2}^{b/2}\int_{-a/2}^{a/2} E(x,y,0)\exp(j\beta(x\hat{X} + y\hat{Y})\mathrm{d}x\mathrm{d}y, \tag{8.63}$$

where $\hat{\mathbf{R}} = \hat{X}\hat{\mathbf{x}} + \hat{Y}\hat{\mathbf{y}} + \hat{Z}\hat{\mathbf{z}}$ is a unit vector in the direction of P. The simplest example is a plane wave $E = E_0\hat{\mathbf{y}}$ that is incident upon the rectangular aperture, then the integral in the above expression is easily evaluated and yields

$$E(\mathbf{R}) = \mathrm{j}\beta\frac{\exp(-\mathrm{j}\beta R)}{2\pi R}E_0\frac{\sin(\frac{\beta a\hat{X}}{2})}{\frac{\beta\hat{X}}{2}}\frac{\sin(\frac{\beta b\hat{Y}}{2})}{\frac{\beta\hat{Y}}{2}}. \tag{8.64}$$

For a direction that is orthogonal to the aperture (i.e. $\hat{X} \to 0$ and $\hat{Y} \to 0$) the radiation field will be maximum in amplitude, with $E = \mathrm{j}\beta(\int_A E\mathrm{d}S)\exp(-\mathrm{j}\beta R)/2\pi R$, where A is the area of the aperture. Further, the directivity will also be maximum and its value can be calculated from

$$D_{\max} = \frac{4\pi}{\lambda^2}\frac{|\int_A E\mathrm{d}S|^2}{\int_A |E|^2\mathrm{d}S} \tag{8.65}$$

on noting that the total radiation will actually be that which emerges from the aperture (i.e. $P_{\mathrm{rad}} = (1/2\eta_0)\int_A |E|^2\mathrm{d}S$). For a plane-wave excitation of the aperture, we obtain that $D_{\max} = 4\pi A/\lambda^2$ since $D_{\max} = G_{\max}$. This brings us to the important concept of *effective aperture*, a quantity that is defined as $A_e = \lambda^2 G/4\pi$ ($A_e = A$ in the case of a lossless aperture excited by a plane wave). In the case of an aperture that consists of the end of a rectangular waveguide, the above expression for directivity yields that $D_{\max} = 32A/\pi\lambda^2$ and from which $A_e = 8A/\pi^2$ on assuming the aperture to be lossless.

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

In this case, the effective aperture is less than the physical area. This arises due to the tapering of the field towards the sides of the waveguide. However, whilst there is a slight reduction in the effective aperture, it turns out that there is a considerable reduction in the level of side lobes (this is a beneficial effect of tapering that we have already noted in the case of arrays). It is clear that increasing the aperture size will also cause an increase in gain and, as a result, a reduction in the width of the main lobe. Consequently, to more effectively direct the radiation, we need to increase the size of the aperture and could do this by opening out the waveguide into a horn structure, as shown in Figure 8.33b. We could further increase the aperture by introducing a parabolic reflector that is fed by such a horn (the effective area is now the area of the reflector). In the receive mode, the effective aperture is a measure of the area over which an antenna collects a radio signal and so the total power collected is $P_R = A_{\text{eff}}P$, where P is the power per unit area that is incident upon the antenna.

An antenna that has become increasingly useful in high-frequency applications, especially mobile phones, is known as the *patch antenna*. This combines aperture and array concepts to form a small antenna with high gain. The antenna consists of an insulating layer of high permittivity that is sandwiched between a metallic rectangle and a larger metallic ground plane (see Figure 8.35a). The antenna is usually constructed from printed circuit board and so the height h of the insulating layer is usually fairly small (a few millimetres). Further, it is usually sufficient that the ground plane spreads out a distance of only a few h beyond the edges of the rectangle. Consequently, the total size of the patch will not be much greater than that of the conducting rectangle. We can imagine the antenna as a length a of microstrip transmission line that is open at both ends (see Figure 8.35b). The distribution of voltage on the rectangle, in the direction of the side of length a, will be $\mathcal{V} = \Re\{V_0 \sin(\pi x/a)\exp(\mathrm{j}\omega t)\}$ when a is chosen to be a half wavelength for the substrate (i.e. $a = \pi/\omega\sqrt{\epsilon\mu_0}$ where ϵ is the permittivity of the substrate). As a consequence, the field under the rectangle will be given by $\mathcal{E} = \Re\{E_0 \sin(\pi x/a)\exp(\mathrm{j}\omega t)\}$ where $E_0 = V_0/h$. (Note that, to ensure that this mode dominates, we need to choose b such that $b < a$.) The simple transmission-line theory would imply that the impedance

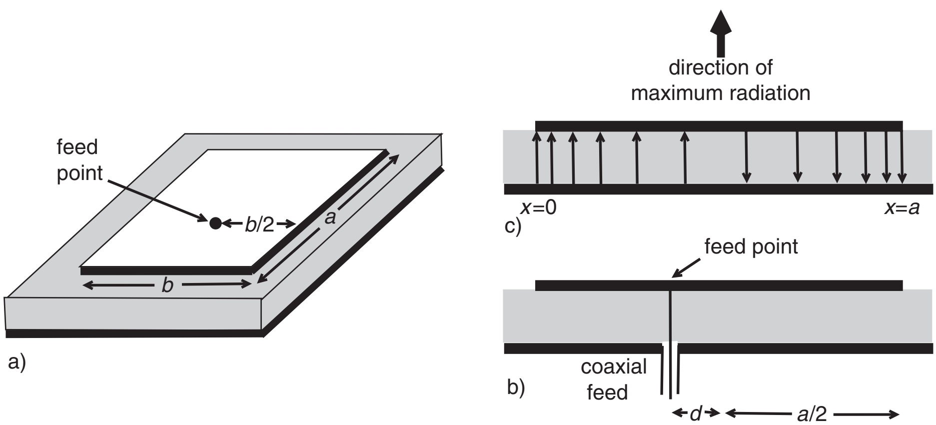


Fig. 8.35 The rectangular patch antenna.

https://doi.org/10.1017/9781108684514.009 Published online by Cambridge University Press

at the centre of the patch is zero and, at the ends, is infinite. However, a more detailed analysis shows that the impedance is finite at the ends and is given by $R_{\mathrm{edge}} = 60\lambda_0/b$ where λ_0 is the wavelength in free space (Balanis, 2016). To feed the patch from a transmitter, we could use the coaxial feed shown in Figure 8.35c. To make the antenna match the coaxial feed (usually a value of 50 ohms), we can use the fact that the input impedance varies across the patch according to

$$R_{\mathrm{in}} = R_{\mathrm{edge}} \sin^2\left(\frac{\pi d}{a}\right) \tag{8.66}$$

and choose the feed point appropriately (d is the distance from the centre of the patch). The radiation will take place from the apertures at the ends of the patch. In the horizontal direction the fields of the two apertures will cancel but, in the vertical direction, the fields will add constructively to produce a gain of up to 6 dB. The radiation will also be reflected in the ground plane and this will add a further 3 dB of gain. Consequently, if we ignore losses, the antenna can have up to 9 dB of total gain in a direction orthogonal to the patch.

8.10 Conclusion

In the present chapter we have discussed antennas, the means by which electronically generated RF energy is launched as a radio wave. We have discussed antennas with a variety of properties, broadband antennas that are appropriate when a range of frequencies are to be used, small antennas for personal communications and steerable antenna arrays that can be used for radar. Obviously, an antenna is of no use when no propagation path exists between the transmitter and receiver and so, in the next chapter, we study the subject of radio wave propagation in some detail.