Cambridge University Press
978-0-521-83481-0 - An Introduction to Radio Frequency Engineering
Christopher Coleman
Excerpt
More information

1 Basic concepts

Broadly speaking, radio frequency (RF) technology, or wireless as it is sometimes known, is the exploitation of electromagnetic wave phenomena in that part of the spectrum between 3 Hz and 300 GHz. It is arguably one of the most important technologies in modern society. The possibility of electromagnetic waves was first postulated by James Maxwell in 1864 and their existence was verified by Heinrich Hertz in 1887. By 1895, Guglielmo Marconi had demonstrated radio as an effective communications technology. With the development of the thermionic valve at the end of the nineteenth century, radio technology developed into a mass communication and entertainment medium. The first half of the twentieth century saw developments such as radar and television, which further extended the scope of this technology. In the second half of the twentieth century, major breakthroughs came with the development of semiconductor devices and integrated circuits. These advances made possible the extremely compact and portable communications devices that resulted in the mobile communications revolution. The size of the electronics continues to fall and, as a consequence, whole new areas have opened up. In particular, spread spectrum communications at gigahertz frequencies are increasingly used to replace cabling and other systems that provide local connectivity.

The purpose of this text is to introduce the important ideas and techniques of radio technology. It is assumed that the reader has a basic grounding in electromagnetic theory and electronics. This text is not intended to be a comprehensive description of radio, but rather to provide the reader with sufficient knowledge to be able to appreciate the more advanced literature on the subject. The current chapter introduces some basic ideas concerning radio frequency systems and subsequent chapters address various aspects of the technology.

1.1 Radio waves

For a static electric field, resulting from a bounded system of charge, the field strength E will fall away at least as fast as the inverse square of the distance R from the source ($E \propto R^{-2}$). If the system suddenly changes from one static configuration to another,

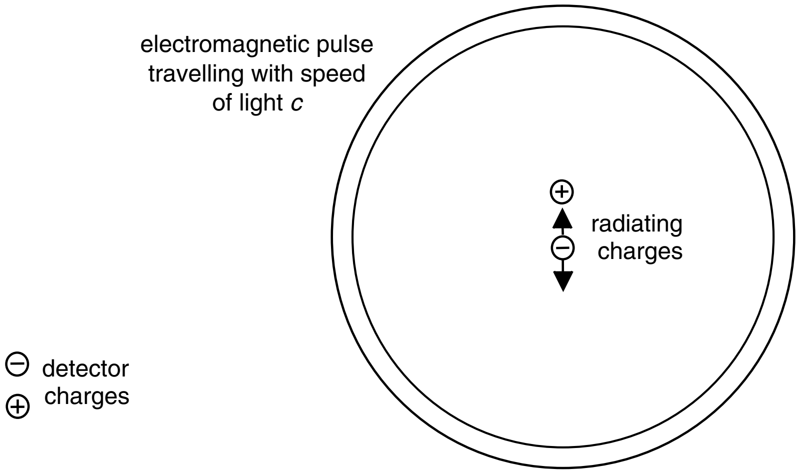


Figure 1.1 Radiation of a radial pulse.

the effect of this change will travel out from the system as a pulse. The pulse will travel with the speed of light ($c = 3 \times 10^8$ m/s) and hence, after time t, will be at distance $R = ct$ from the source. Remarkably, the amplitude of the pulse will only fall away as R^{-1} and this means that, at large distances, the effect of the pulse far outweighs that of the static fields with which it is associated. In principle, the system of charge can be *modulated* so that the fields carry away information as a series of suitably spaced pulses. As the pulses pass a second system of charge, the electric field will set these charges in motion and the pulses can be detected. The important point is that information can be carried over great distances due to the slow rate at which the pulse amplitudes decay. A dynamic electric field cannot exist in isolation and the electromagnetic theory of Maxwell predicts that there will be a magnetic field pulse associated with the electric field pulse. In fact, the electric field (E) of the pulse will be related to the magnetic field (H) through

$$E = \eta_0 H, \tag{1.1}$$

where η_0 is the impedance of the free space (377 Ω). Furthermore, the electric field, magnetic field and propagation direction will be mutually orthogonal (Figure 1.2). Far from the system of charge

$$E = \frac{K(t - R/c)}{4\pi R}, \tag{1.2}$$

where $K(t)$ is a function that depends upon the source modulation (possibly a series of pulses).

In general, any accelerating charge will give rise to a field that only falls away as R^{-1} and this leads to a great variety of ways in which fields, suitable for radio communication, can be generated. In particular, charges can be made to oscillate at a

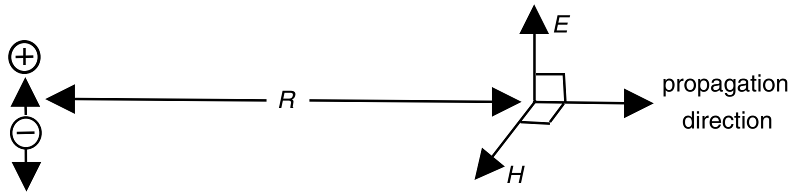


Figure 1.2 Electromagnetic field generated by a system of accelerating charge.

particular frequency ω (radians per second or $f = \omega/2\pi$ in terms of hertz) and this gives rise to a field for which function $K(t)$ is of the form $A\cos(\omega t + \phi)$ (A controls the amplitude of the field and ϕ its phase). The advantage of generating such a field is that it can allow many communications systems to coexist by operating them at different frequencies. Detection systems can be *tuned* to a particular frequency in order to receive signals from a particular source. In order to transmit information, the source will need to be *modulated* in some sense by this information. The modulating signal itself will occupy a range of frequencies known as the *baseband* and this will have a *bandwidth* that depends on the data rate. Modulation will cause the transmitted signal to spread in frequency around the purely sinusoidal *carrier* and, depending on the type of modulation, this spread can be wider than the bandwidth of the original baseband signal. Consequently, the type of modulation is an important consideration in allocating frequencies to users of the radio spectrum. Modulation is achieved by varying either A for amplitude modulation (AM), ω for frequency modulation (FM) or ϕ for phase modulation (PM).

The generation of electromagnetic waves can be illustrated through the evolution of the electric field of a simple oscillating dipole (a pair of vibrating charges with equal magnitude and opposite signs). The electric field lines will evolve as shown in Figure 1.3 (note that the field lines travel outwards from the source at the speed of light). When the charges pass, the field lines will join together and break away to make room for new field lines to develop (note that field lines cannot cross). After one period of oscillation, the field lines joining the charges will look identical to those at the start. The actual field lines, however, will have moved a *wavelength* λ out from the charges ($\lambda = c/f$ where f is the frequency of oscillation in hertz). As the oscillations continue, the field lines will continue to move out by a distance λ for each period of oscillation.

In a realistic communications system, the *radiation* fields will be produced by an electronic source that drives current into a metallic structure known as an antenna. At the atomic level, the antenna will consist of a complex combination of simple oscillating dipoles whose fields will combine to form a pattern of radiation that is dictated by the geometry of the antenna. If the metallic structure is a rod that is driven at its centre, we will have a *dipole antenna* (so called because the resulting radiation pattern is almost identical to that of a pair of oscillating charges). The field (and hence the radiated power) will be maximum in directions orthogonal to the axis of the antenna

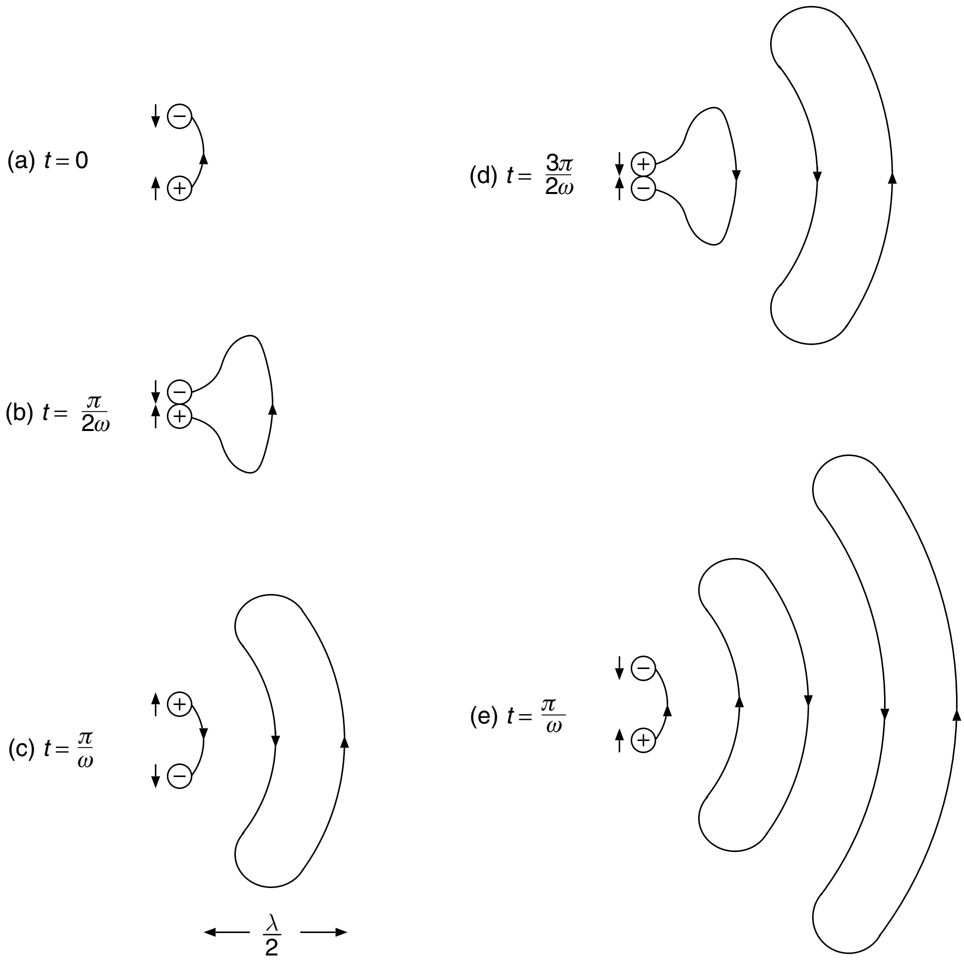


Figure 1.3 Development of oscillating dipole field.

and minimum (zero, in fact) in directions along the axis of the dipole (note that the radiation will be symmetric about the axis). A useful way of describing the radiation characteristics of an antenna is through its *directivity* (or the related concept of *gain*). The directivity D in a particular direction is defined by

$$\text{directivity} = \frac{\text{power radiated in a particular direction}}{\text{average of power radiated in all directions}}. \tag{1.3}$$

Unfortunately, not all the power supplied to an antenna will manifest itself as radiation and some will be lost as heat in its structure (and possibly in its surroundings). The efficiency η with which the power is converted to radiation is defined by

$$\text{efficiency} = \frac{\text{total power radiated}}{\text{total power supplied}}. \tag{1.4}$$

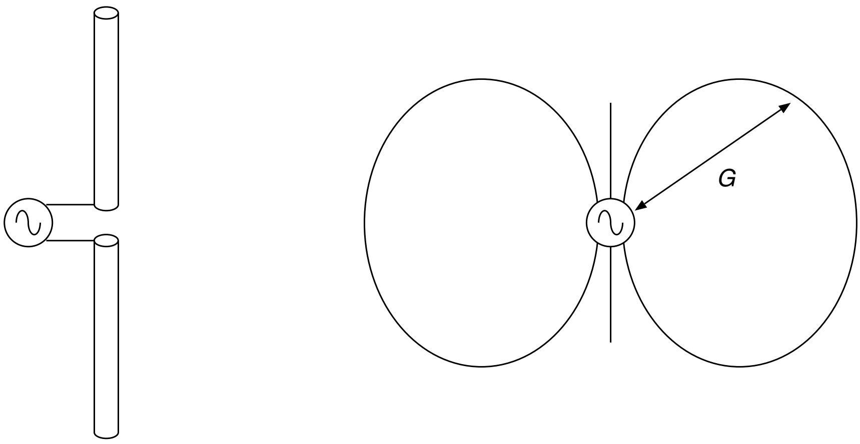


Figure 1.4 A dipole antenna and its gain pattern.

A more realistic measure of antenna effectiveness is the gain G, defined by

$$\text{gain} = 4\pi \frac{\text{power radiated into a unit solid angle}}{\text{total power supplied}} \tag{1.5}$$

and, from the above definitions, it will be noted that gain $= \eta \times$ directivity. Directivity describes the deviation of radiation properties (usually expressed in dB terms) away from those of an ideal isotropic antenna and is usually represented through its *directivity pattern*. For a general antenna, the directivity pattern is a three-dimensional surface that surrounds the origin (associated with the antenna). In a particular direction, the distance of the surface from the origin is the value of directivity in that direction. The *gain pattern* of an antenna is defined in a similar fashion and is more often quoted. The directivity pattern for a half wavelength dipole is shown in Figure 1.4 (a slice through the dipole axis) and for which it should be noted that the maximum directivity is approximately $5/3$ or 2.2 dBi (dB over isotropic).

From a circuit viewpoint, a transmit antenna can be represented by the circuit shown Figure 1.5. Note the $R_T + jX_T$ is the impedance of the RF source (the transmitter or Tx) and $R_r + R_L + jX_L$ is the impedance of the antenna. The antenna has two resistance contributions, R_L which represents the ohmic (heating) losses in the antenna and R_r which represents the losses due to power radiated away from the antenna (good loss). From the circuit model, it will be noted that the power P_r radiated by the antenna will be given by

$$P_r = \frac{V_T^2 R_r}{2(R_T + R_r + R_L)^2 + 2(X_T + X_A)^2}. \tag{1.6}$$

This power will take its maximum value $(V_T^2/8R_r)$ when $X_T = -X_A$ and $R_r = R_T + R_L$. A dipole is resonant (no reactance X_A in its impedance) when its length is about 0.47λ (λ is the wavelength c/f at the operating frequency f). At resonance, the radiation

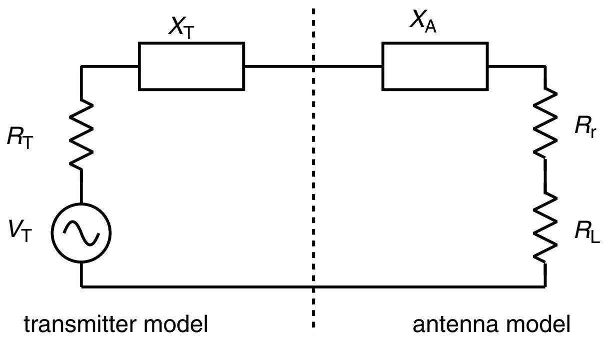


Figure 1.5 Circuit model of a transmit system.

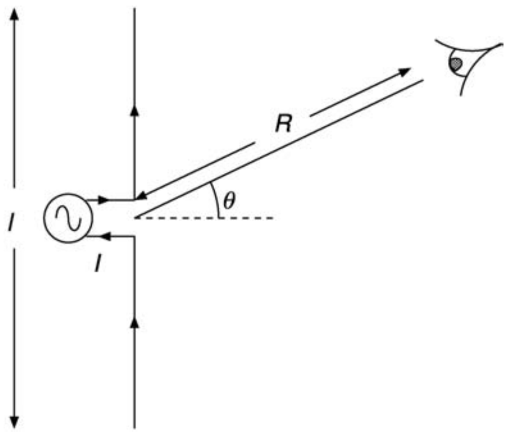


Figure 1.6 Conventions for effective length.

resistance is about 73 Ω and the loss resistance R_L is usually negligible for most practical dipoles (this means that the directivity and gain will be almost identical). Dipoles that are much shorter than a wavelength, however, have a very much smaller radiation resistance and a large capacitive reactance. For short dipoles, the radiation and ohmic resistances can often be comparable in magnitude and hence make such antennas inefficient as radiators (the directivity and gain will be significantly different in this case).

It is interesting to note that the amplitude of the electric field for any sinusoidally excited antenna has the form

$$E = \frac{\omega \mu I h_{\text{eff}}}{4\pi R} \tag{1.7}$$

at a distance R from the antenna. Quantity I is the current at the antenna feed and quantity h_{eff} has the dimensions of length and is known as the *effective length* of the antenna. For a dipole antenna, the effective length is approximately $0.64l \cos\theta$, where l is the geometric length and θ is the angle of observation when measured from the

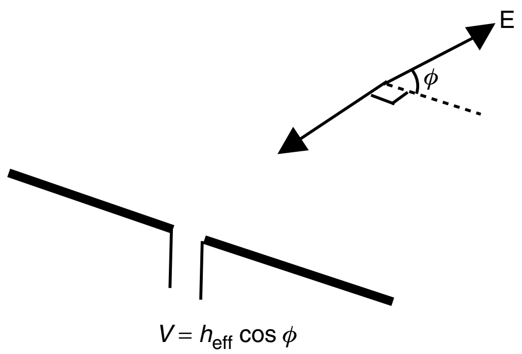


Figure 1.7 A dipole antenna used to collect energy from an electromagnetic wave.

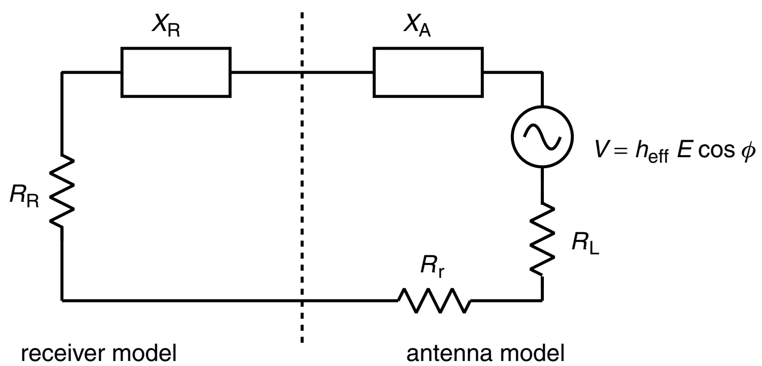


Figure 1.8 Circuit model of receive system.

plane of maximum radiation (the plane through the antenna feed that is orthogonal to the antenna axis). Now consider the case of a dipole that is used to collect energy from a time varying electric field E (the receive mode). It can be shown that an electric field will induce an open circuit voltage $h_{\mathrm{eff}} E \cos\phi$ in the dipole terminals, where ϕ is the angle between the field direction and the axis of the antenna (the θ used in calculating h_{eff} will be the angle between the source direction and the plane that is orthogonal to the antenna axis).

The above considerations bring us to the concept of polarisation. It will be noted that the electric field can point in any direction, just so long as it is orthogonal to the direction of propagation. That is, the wave can have many different *polarisations*. A receive antenna, however, will only extract the maximum power when it is *polarisation matched* to the incoming wave ($\phi = 0$). Consequently, in a communication system, we will normally design for polarisation match. In the case of a system that uses dipole antennas for both receive and transmit, this will require the dipoles to be parallel.

A circuit model for a receiving antenna is shown in Figure 1.8. From this model, it will be noted that an antenna exhibits the same impedance $R_r + R_L + jX_A$ in both

Cambridge University Press
978-0-521-83481-0 - An Introduction to Radio Frequency Engineering
Christopher Coleman
Excerpt
More information

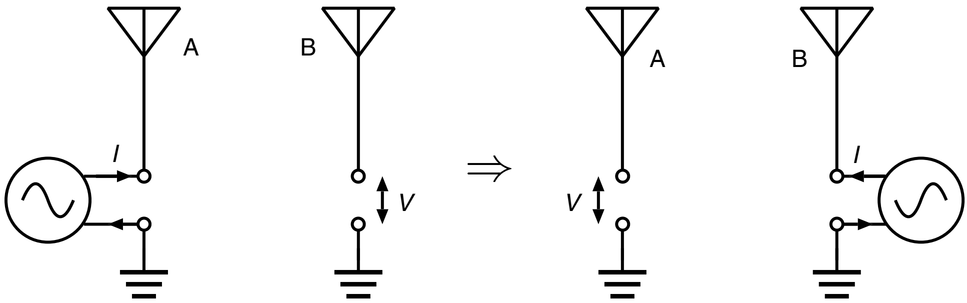


Figure 1.9 Reciprocity principle.

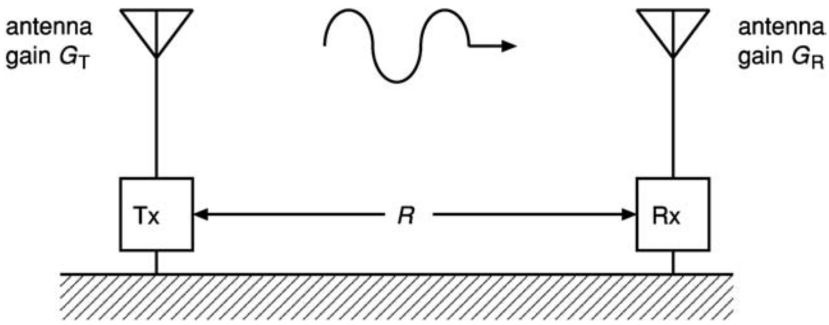


Figure 1.10 Transmit/receive system.

receive and transmit modes (maximum power will be received when $X_R = -X_A$ and $R_R = R_r + R_L$). A further result concerning the interchangeability of transmit and receive properties is known as *reciprocity*. Consider two antennas (A and B) with antenna A driven by current I and causing an open circuit voltage V in antenna B. If we now drive antenna B with the same current I and measure the open circuit voltage in A, we will find it to be the same voltage V. This is the case, even if A and B are completely different antennas. An important consequence of this result is the ability to infer two-way communication properties from one-way properties. To investigate the coverage of a mobile communications base station, for example, it is only necessary to investigate coverage of signals transmitted from the base station.

We have already noted that the field of a radio wave will fall away as the inverse of distance from the transmitter. As a consequence, the question arises as to the level of transmit power that is required in order to achieve a given level of power at the receiver (or Rx). This value can be calculated using what is known as the *Friis* equation. If we have communications between a transmitter and a receiver, distance R apart, the received power P_R and transmitted power P_T will be related through

$$P_R = P_T \left(\frac{\lambda}{4\pi R}\right)^2 G_R G_T, \tag{1.8}$$

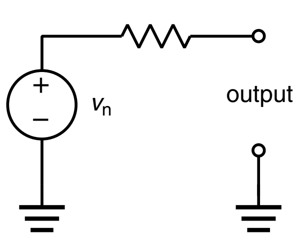
where G_R and G_T are the gains of the receive and transmit antennas, respectively. (As can be seen from this equation, gain also measures the effectiveness of an antenna in its receive role.) The Friis equation, and its variants, are extremely important tools in the design of a radio system.

1.2 Noise

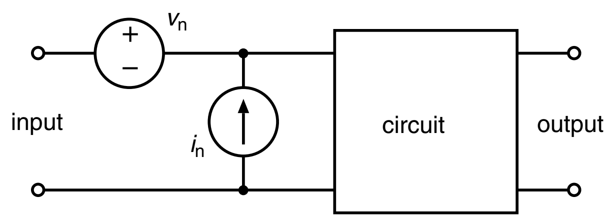
It might seem that we could transmit at any level of signal power and simply introduce a suitable amount of amplification at the receiver end. Unfortunately, this is not the case due to the fact that the signal will be competing with an ever present environment of random signals or *noise*. For example, a simple resistor will create a noise voltage v_n due to the random thermal motion of its electrons and this can be shown to have an rms voltage that satisfies

$$\overline{v_n^2} = 4kTBR, \tag{1.9}$$

where T (in kelvin) is the absolute temperature, B (in hertz) is the bandwidth of the measurement, R (in ohms) is the resistance and k is the Boltzmann constant (1.38×10^{-23} joules per kelvin). Equation 1.9 will still apply to a general impedance Z providing R is interpreted as the resistive part of the impedance (i.e., $R = \Re\{Z\}$). From a modelling viewpoint, the noise source can be regarded as an ideal voltage source of magnitude v_n in series with a noise free impedance. Alternatively, it can be regarded as an ideal current source of magnitude i_n in parallel with the impedance (note that $\overline{i_n^2} = 4kTBG$, where $G = \Re\{Z^{-1}\}$). In general, the noise in an electronic circuit can be modelled by removing the noise sources from within the circuit and replacing them by equivalent current and voltage sources at the input (Figure 1.11b). These equivalent sources can be quite complex since a general circuit can contain other forms of noise besides that due to the resistance (the shot and flicker noises of semiconductor devices, for example). In a radio receiver, the input signal will already be in competition with *external noise* from man-made sources (ignition interference, for example) and natural sources (lightning,



(a) Simple resistor noise source



(b) Equivalent noise sources for a complex circuit

Figure 1.11 Noise sources.

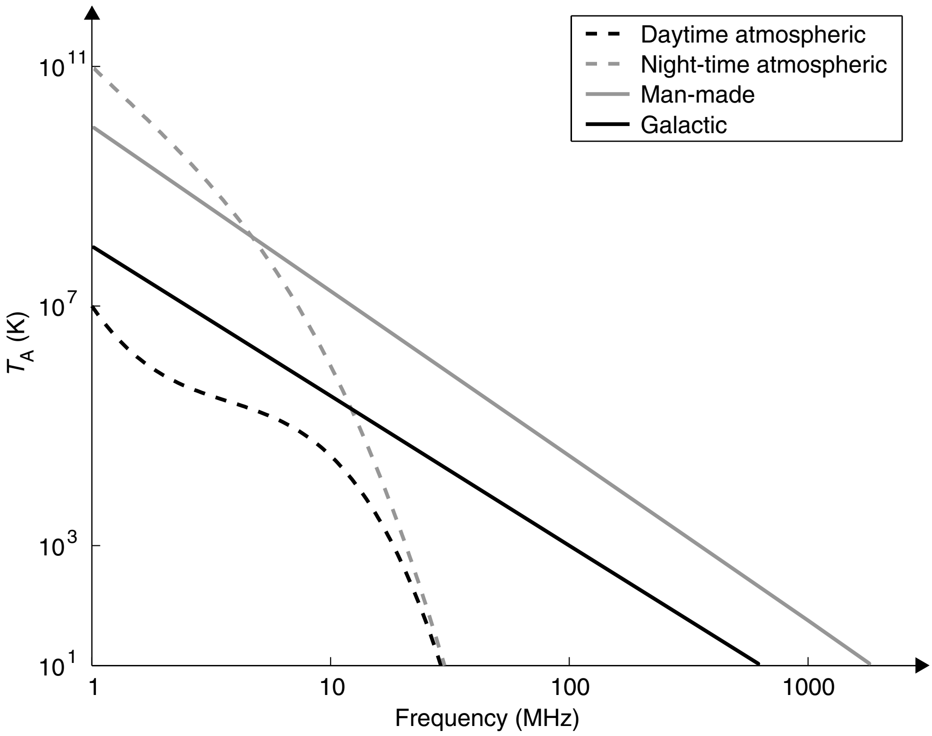


Figure 1.12 Typical antenna noise temperatures for a dipole and a variety of sources.

for example). Consequently, the input signal will need to be at a level well above that of the combined internal and external noise. It is possible to view the external noise as that arising from the antenna resistance $R_r + R_L$. To arrive at the correct level of noise, however, it is necessary to regard the system as located in an environment with an *antenna temperature* T_A that could be vastly different from the ambient temperature (around 290 K). Figure 1.12 shows some typical antenna temperatures resulting from a variety of external noise sources (note that the sources are uncorrelated and so we can simply add the temperatures to get the combined effect). In general, a noise source that is non-thermal can be treated as a thermal source with a suitably chosen *noise temperature*. External noise is the ultimate constraint since this is not under the control of the designer. For best performance, a radio receiver should be designed such that it is *externally noise limited* (i.e., the internal noise is below the expected level of external noise).

The maximum noise power that can be derived from a resistor R will be $N = kTB$ and this will be achieved when the load has impedance R. If an RF circuit is fed from a noisy source, it is clear that the amount of noise that reaches its output will depend on the circuit bandwidth B. The circuit itself will, however, add to this noise and it is important to ascertain whether the combined noise will swamp any desired signal that is present at the input. The crucial quantity in assessing circuit performance is the