10 Modern Radio Systems

In the previous chapters, we have studied the basic elements of radio. This has included the generation and detection of RF signals, their launch and capture as radio waves and the propagation of these waves. Thus far, we have concentrated on radio signals that consist of a carrier with simple analogue modulation. However, modern communication is increasingly accomplished by means of digital modulation, even when conveying analogue information. This has allowed coding that has decreased the bandwidth requirement of analogue signals such as television and audio. However, there is now considerable pressure on the available spectrum and this has necessitated the development of novel approaches that can optimise spectral usage. In the current chapter we consider some of these approaches, including techniques such as spread spectrum, cellular radio and MIMO. Additionally, we consider some of the non-communication uses of radio including surveillance, navigation and astronomy.

10.1 Digital Communication Systems

In digital communications, information is transferred as a stream of distinct modulation states and such communications, in the form of Morse code, were in frequent use well before the age of radio. Morse code was developed as a way of sending complex messages down telegraph wires in an efficient manner, the modulation states consisting of a series of long and short pulses (dashes and dots) that conveyed the information. In the early days of the telegraph, when clicks were the information sent, the dots and dashes were differentiated by the time between them. Table 10.1 shows the Morse code for some basic characters (the letters and numbers), but there is a Morse code for most of the important characters (e.g. · — · — ·— for a period). The basic idea of the code was developed by Samuel Morse in 1834 and further refined by Alfred Vail. In later usage, tones were transmitted with the dots and dashes represented by short and long tones with the dash being 3 times the length of the dot. The dots and dashes were separated by the length of a dot, the letters by the length of 3 dots and words by the length of 7 dots. The complexity of the code for a particular character was chosen on the basis of its frequency of usage. Consequently, the most frequently used character 'e' was designated a single dot. Modern analysis of the code shows it to be fairly close to the theoretical ideal in terms of efficiency. For reasons explained in Chapter 2, Morse code became a convenient means of conveying complex messages in radio and continued to be used for long-distance communications well into the 1960s.

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

Table 10.1 Some basic Morse code.

A	·—	M	——	Y	—·——
B	—···	N	—·	Z	——··
C	—·—·	O	———	1	·————
D	—··	P	·——·	2	··———
E	·	Q	——·—	3	···——
F	··—·	R	·—·	4	····—
G	——·	S	···	5	·····
H	····	T	—	6	—····
I	··	U	··—	7	——···
J	·———	V	···—	8	———··
K	—·—	W	·——	9	————·
L	·—··	X	—··—	0	—————

Whilst digital radio communications have always been with us, the development of efficient means of digitising data has resulted in nearly all radio communications becoming digital. Most broadcasting has now switched over to digital forms and mobile phones have been digital for quite some time. A major issue, however, is the way in which this digital data is modulated onto the radio carrier. Binary communications use two modulation states and these are represented by phase states in the case of *phase-shift keying* (PSK), by amplitude states in the case of *amplitude-shift keying* (ASK) and two frequency states in the case of *frequency-shift keying* (FSK). The modulating sequence $g_0g_1g_2g_3g_4\ldots$ for binary communications consists of single-bit symbols (these can take the values 0 or 1). Consider a modulated signal of the form

$$E(t) = A(t) * \cos(\omega(t)t + \phi(t)), \tag{10.1}$$

then, in the case of ASK,

$$A(t) = A_0 g_{[\frac{t}{T}]}, \quad \omega(t) = \omega_c \quad \text{and} \quad \phi(t) = 0, \tag{10.2}$$

where A_0 is a constant amplitude, $[\frac{t}{T}]$ is the integer part of $\frac{t}{T}$, T is the duration of the pulse representing a bit and ω_c is the carrier frequency. In the case of PSK,

$$A(t) = A_0, \quad \omega(t) = \omega_c \quad \text{and} \quad \phi(t) = \pi g_{[\frac{t}{T}]} \tag{10.3}$$

and, in the case of FSK,

$$A(t) = A_0, \quad \omega(t) = \omega_c + \Delta f g_{[\frac{t}{T}]} \quad \text{and} \quad \phi(t) = 0, \tag{10.4}$$

where Δf is the frequency deviation. PSK exhibits the smallest bit-error rate against noise (10^{-5} for 10 dB of SNR) with FSK and ASK far behind (about 10^{-3} and 10^{-2} of bit-error rate for 10 dB of SNR, respectively). It is possible to shorten the modulating sequence ($g_0g_1g_2g_3g_4\ldots$) by using a quaternary base (symbols can take the values 0, 1, 2 or 3), but this requires a modulation system with four states. The most important example of this is quadrature phase-shift keying (QPSK) for which

$$A(t) = A_0, \quad \omega(t) = \omega_c \quad \text{and} \quad \phi(t) = \frac{\pi}{4}(2g_{[\frac{t}{T}]} + 1). \tag{10.5}$$

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

QSPK has the same symbol error rate as PSK, but it has the advantage that it conveys more information in a symbol and hence is more efficient. The key to effective digital communications is the efficient representation of the data to be communicated in terms of the modulation states, i.e. the *coding*. The high efficiency of modern codes has meant that most forms of data (voice, television, etc.) can now be transferred digitally with a much lower bandwidth than could be achieved with analogue modulation. However, the continual pressure on radio spectrum has meant that even greater efficiency is required. Consequently, more imaginative ways of utilising the spectrum have needed to be developed and some of these are the subject of the next few sections.

10.2 Spread-Spectrum Systems

The Shannon–Hartley theorem

$$C = B\log_2(1 + \mathrm{SNR}) \tag{10.6}$$

provides a relationship between channel capacity C (bits per second), the channel bandwidth B (Hz) and the SNR. The relationship suggests that an increase in bandwidth will improve capacity. Indeed, it even suggests that there can even be transmission of data when the SNR is less than 1 ($C \approx 1.433 \times B \times \mathrm{SNR}$). Systems that increase capacity by increasing bandwidth are known as *spread-spectrum* systems. Surprisingly, the idea of spread-spectrum systems was first suggested in 1941 by the Hollywood actress Hedy Lamarr and the pianist George Antheil. They donated their idea to the US government to help the war effort, but the idea was not taken up until the 1980s. Since then, it has become the basis of many important technologies, 3G mobile communications being an important example. Two major examples of spread-spectrum systems are the frequency-hopping (FH) and the direct-sequence spread-spectrum (DSSS) varieties.

In FH spread-spectrum systems, the signal will hop around a set of frequencies in a pseudo-random fashion and this will be simultaneously tracked by a receiver that has knowledge of this sequence. In theory, the signal will only occupy a channel containing interference for a limited time and this will result in an overall increase in SNR. In addition, many users can use the same frequencies by each using a different pseudo-random sequence of frequencies since, with different random sequences, there will be a low probability of collision amongst these users. An important example of this approach is the Bluetooth system for interconnecting electronic devices.

In the DSSS variety of spread spectrum, a digital baseband sequence $a(t)$ is modulated by a much higher bit rate pseudo-random sequence $p(t)$ to form a new sequence $a(t)p(t)$ that is now *spread* in frequency. The new sequence is then used to modulate a carrier and this results in a wide bandwidth RF signal. At the receiver, the process is then reversed. The RF signal is first demodulated and then the spread baseband sequence is multiplied by a replica of the pseudo-random sequence in order to reconstruct the original baseband sequence.

There are two major advantages to the DSSS approach. Firstly, if we were to reduce the overall power level we could then maintain the same channel capacity by increasing the bandwidth. Secondly, although an increase in bandwidth might seem wasteful of spectrum, this is compensated for by the fact that many users can occupy the same channel by using uncorrelated pseudo-random sequences (multiplication by the wrong sequence will simply produce noise).

A realisation of DSSS is illustrated by Figures 10.2 and 10.3 with the delay τ in the receiver system used to compensate for the delay caused by propagation. In mobile communications one of the major problems is that of multiple propagation paths with the differential delays causing *inter-symbol interference*. For communication channels that suffer from multi-path propagation, DSSS offers a solution by adding together copies of a signal that have been demodulated using a range of delay parameter τ. Due to the pseudo-random nature of the code, signals that do not match the delay will only produce noise. However, signals that match the delay will coherently add. Importantly, this approach does not require synchronisation of the receiver and transmitter.

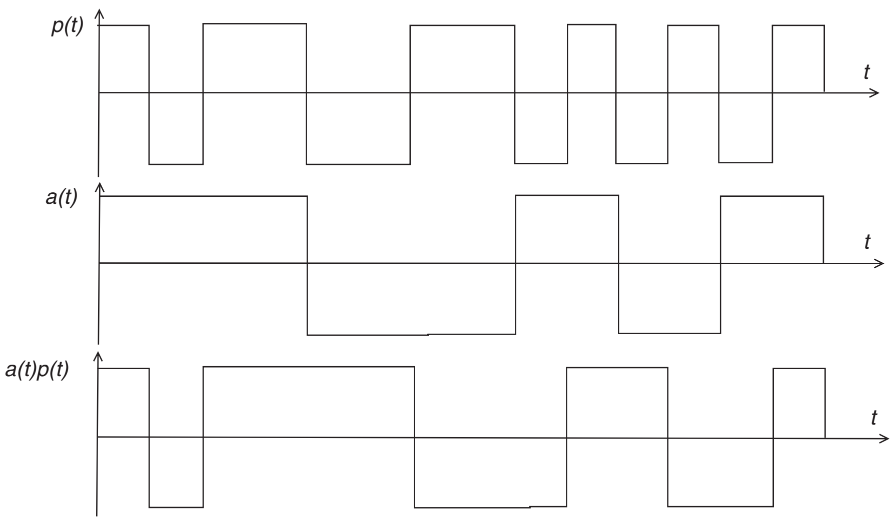


Fig. 10.1 Formation of a spread baseband signal.

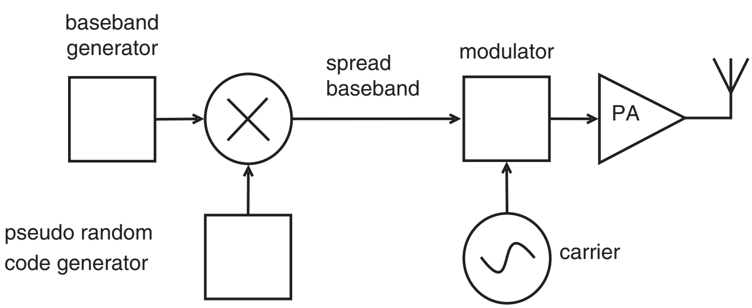


Fig. 10.2 A simple DSSS transmitter.

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

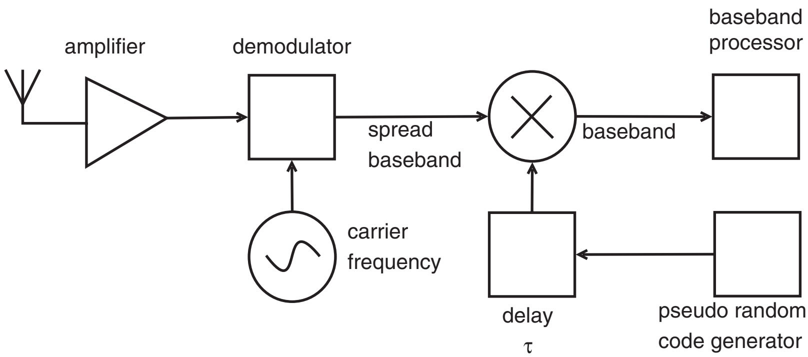


Fig. 10.3 A simple DSSS receiver.

For broadcast radio, an important spread-spectrum system is known as *digital audio broadcasting* or DAB for short. This system provides multiple broadcast services, together with text information. The data is split up into a total of 1,536 parallel data streams (also known as multiplexes) that are transmitted using DQPSK (differential quadrature phase-shift keying) on 1,536 sub-carriers that are separated by 1 kHz. (In DQPSK the changes in phase carry the information rather than absolute phase states.) The multiplexes are orthogonal since their carriers are harmonically related (i.e. they are easily separated from each other using fast DFT techniques) and the data rate of each multiplex is kept low enough so that the multiplexes do not interfere with each other. Such a system is known as OFDM or *orthogonal frequency-division multiplexing*. As with users of mobile communications, mobile users of broadcast services will also suffer from multi-path problems. With DAB, however, we essentially have many low-data-rate channels for which the delays of multi-path are insufficient to cause such significant overlap of the now very much longer symbols, i.e. inter-symbol interference will be reduced. As mentioned in Chapter 9, fading is another problem caused by multi-path. However, the phase shift that causes the fading is frequency-dependent (see Eq. (9.15) of Chapter 9) and so, at any one time, only a limited amount of the multiplexes will be affected by the fading (i.e. we have frequency-selective fading). As a consequence, providing that each service is suitably spread across the carriers, the effect of fading will be minimised.

10.3 Cellular Radio

Although digital and spread spectrum techniques have led to a far more efficient usage of the spectrum, the explosion in personal communications (including video and internet) has forced even more dramatic solutions. The most important of these is *cellular radio*, a system that limits the coverage of a channel so that its frequency can be reused at other locations. The total network area is divided into small cells, each of which contains a radio base station (RBF). The users within a cell will then have their transmit power limited so that they can only communicate with the RBF of that cell. As a consequence, the same set

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

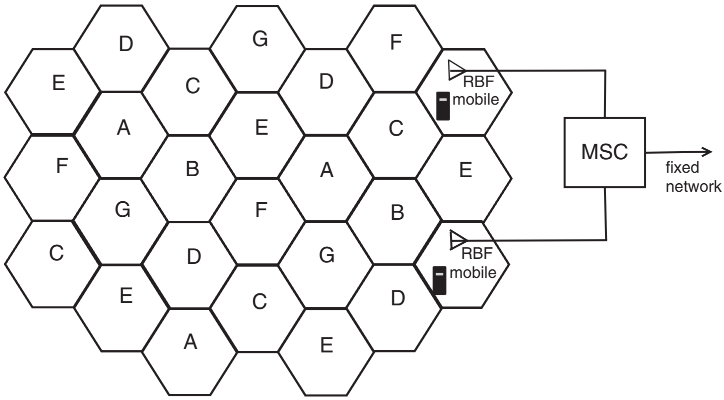


Fig. 10.4 A cellular radio system.

of channels can be used in other cells that are sufficiently isolated. The RBFs within the network are all interconnected through a mobile switching centre (MSC) which passes control from one RBF to another as a user passes from one cell to another, appropriately changing the channels in the process.

The design of a cellular system can be affected by many factors and the cells can vary in size according to the topography and location of the RBF. To ensure that adjacent cells do not share frequencies, the system needs to be designed around a *cluster* of cells that ensures this. A typical cluster consists of seven cells and Figure 10.4 shows a cellular system that is based on such a cluster (frequency sets are labelled A to G). It will be noted this cluster topology allows reuse of the frequency sets in adjacent clusters. However, it is inevitable that there will be some interference between the cells and this will be the dominant source of interference. From Figure 10.4, it will be noted that the minimum distance between cells with the same frequency set is approximately $4.583R$ where R is the cell radius. Then, assuming that all transmitters have the same power level, and that power decays as $(1/\text{distance})^n$ ($n \approx 4$ in most cases), the *signal-to-interference ratio (SIR)* is given by

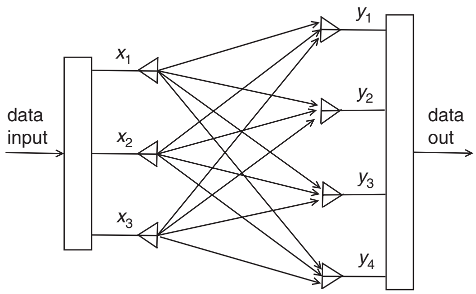
$$\text{SIR} = \frac{\text{minimum power within a cell}}{6 \times \text{maximum power between cells}} = \frac{R^{-n}}{6(4.583R)^{-n}} = \frac{4.583^n}{6}. \tag{10.7}$$

For the configuration of Figure 10.4 this would suggest an SIR of greater than 10 dB.

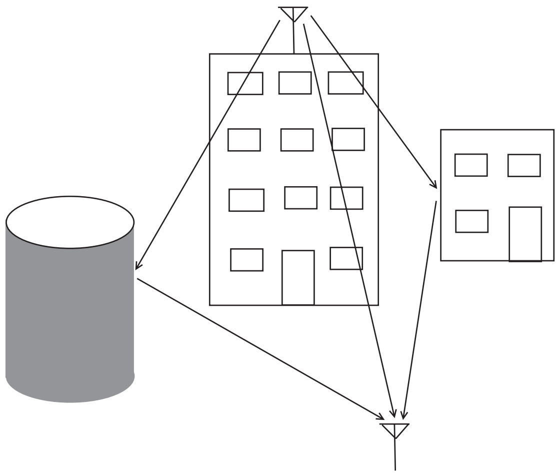
10.4 MIMO

The problem of communications capacity has pushed us into ever-more exotic techniques, one of the more recent being MIMO (multiple input multiple output). The systems that we have so far considered are SISO (single input single output) systems for which, as we have seen, there are several techniques for increasing capacity. MIMO, however, opens up further possibilities. The architecture for a typical MIMO system is

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press



a) A basic MIMO configuration



b) A multi-path environment

Fig. 10.5 MIMO and multi-path.

shown in Figure 10.5a and it will be noted that there are several transmit antennas (N_T say) and several receive antennas (N_R say). Our study of array antennas has already shown us that we can use multiple antennas on transmit, and receive, to increase the strength of the received signal (typically by a factor $N_T N_R$) and hence the SNR. By the Shannon theorem, this means that we have an increase in the channel capacity, i.e. we have what is known as *array gain*. As we have mentioned with respect to DSSS, it is possible to distinguish signals in the same channel through modulation of the baseband by different pseudo-random codes. If we modulate the baseband on separate transmit antennas by different pseudo-random codes, it is now possible to distinguish the contributions from different transmitters at the receive end. Further, if the baseband signals at each transmit antenna are the same, it would then be possible to beamform the transmitter at the receive end. In a multi-receiver system, such as a mobile telephone system, this would allow transmit beamforming at the receivers. Consequently, there are obvious gains to be made with MIMO systems. It will be noted that all of the above advantages also apply to MISO (multiple input single output) systems, but we will now see is that there are tremendous gains to be made when there are multiple receive antennas.

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

We have previously noted that multi-path can be a problem for communications systems by causing both inter-symbol interference and fading. These problems can be overcome by means of diversity techniques in which copies of the signal are transmitted at different frequencies (*frequency diversity*) or at different times (*time diversity*). This can be expensive in resource and so, as we have seen in the previous section, the alternative is to spread the signal over many channels so that, at any one time, only a small part of the signal is in a fading channel. Through multiple antennas, MIMO systems can allow what is commonly called *space diversity*. The simplest form of spatial diversity is for the receiver to have two, or more, antennas that are suitably spaced (half a wavelength or more). For sufficient spacing, it is likely that at least one antenna will receive signals that do not meet the conditions for fading. MIMO systems, however, have the potential to increase data capacity by turning multi-path from a problem into an advantage. Figure 10.5b shows a typical multi-path situation. It will be noted that each path effectively represents a separate communication channel and so has the potential to carry different data from the other paths, provided the system can separate out these paths. In the simplest approach, the signal to be transmitted is split into N_T data streams and these are transmitted through the propagation medium by the N_T separate antennas. At the receive end there will now be N_R data streams, but each of these will contain contributions from all of the transmitted streams. The trick is to separate out the individual transmitted streams. If x_1 to x_{N_T} are the transmitted signal, and y_1 to y_{N_R} are the received signals, then

$$\begin{aligned}
y_1 &= h_{11}x_1 + h_{12}x_2 + \cdots + h_{1N_T}x_{N_T} \\
y_2 &= h_{21}x_1 + h_{22}x_2 + \cdots + h_{2N_T}x_{N_T} \\
&\vdots \\
y_{N_R} &= h_{N_R1}x_1 + h_{N_R2}x_2 + \cdots + h_{N_RN_T}x_{N_T}.
\end{aligned} \tag{10.8}$$

The coefficient h_{ij} describes the propagation between the jth transmitter and the ith receiver. These coefficients are found by transmitting pilot signals from the transmitter, a process that is repeated at intervals short enough to account for time variations in the channel. In order extract the transmitted data streams from the received data streams, we need to solve (10.8). Whether this can be done depends on the number of effective communication channels. The number of transmitted data streams will need to be limited by this number and the minimum of N_T and N_R. For an environment rich in multi-path, however, MIMO has the potential to greatly increase capacity without additional power or bandwidth. MIMO can be combined with other techniques for increasing capacity and the combination MIMO with OFDM is one promising avenue.

10.5 Radar Systems

One of the major non-communications applications of radio waves is *radar* (radio detection and ranging). In a classical radar, the transmitted signal is interrupted by a *target* from which a small amount of energy is reradiated back to a receiver. The receiver will normally ascertain the direction of the target using a steerable array (mechanical or electronic steering) and the time of flight of the signal will then provide the target range. As far back as 1904, the German inventor Christian Hulsmeyer invented a system that detected ships by the reflection of radio waves. However, it was not until the mid-1930s that radar, in its currently recognisable form, came into existence. In 1935, a team led by the British scientist Robert Watson-Watt developed what is often credited as being the first radar system. This development was crucial to the defence of Britain in the second world war and big strides were made in its further development during this war. Radar systems can be regarded as radio systems in which the environment modulates the signal, hence allowing an operator to glean information about the environment. Whilst their primary use has been for detecting ships and aircraft, they are now increasingly used for gleaning information about the natural environment (wind profiling radars for example).

A typical radar configuration is shown in Figure 10.6. As mentioned earlier, the antennas will normally be steerable arrays whose steer direction will provide the direction of the target, with accuracy depending on the size of array. The range, however, is obtained from the time of flight and the accuracy with which this can be measured depends on the nature of the radar signal. For a signal consisting of pulses separated by time T, ie. *pulse-repetition frequency* $1/T$, the accuracy will be $cT/2$ (c is the propagation speed). The power P_R returned from the target is related to the transmitted power P_T through the *radar equation*,

$$P_R = P_T G_R G_T \left(\frac{\lambda}{4\pi R_T}\right)^2 \left(\frac{\lambda}{4\pi R_R}\right)^2 \frac{4\pi\sigma}{\lambda^2}, \tag{10.9}$$

where R_T and R_R are the ranges of the target from the transmitter and receiver respectively and G_T and G_R are the gains of the transmit and receive antennas respectively. σ is the *radar cross section* of the target and represents the amount of power reflected when a field with unit power per unit area is incident. Typically, a light aircraft has a cross section of about 2 m^2 and a jumbo jet a cross section of about 100 m^2. In essence the radar equation is the double application of the Friss equation with the target acting as both a receiver and transmitter. Radar cross sections can be quite complex, often depending on both the direction of the illumination and the direction of the reception.

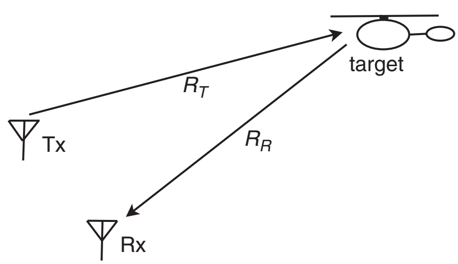


Fig. 10.6 A general radar configuration.

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

Besides signal returns from a desired target (an aircraft for example) there will often be returns from unwanted targets such as rough sea and rough ground. Sea in particular can return a considerable amount of power and has the potential to mask the desired target. Such interference, termed *clutter*, will be in addition to the noise that is experienced by communication systems. In radar, the signal-to-clutter ratio (SCR) can be just as important as SNR in determining radar performance. Fortunately, the motion of the target will itself cause a frequency shift in the radar return and this will normally allow the radar return to be distinguished from the clutter. The frequency shift is known as the *Doppler shift* and is related to the target dynamics through

$$\Delta f = -\frac{f}{c}\left(\frac{\mathrm{d}R_T}{\mathrm{d}t} + \frac{\mathrm{d}R_R}{\mathrm{d}t}\right). \tag{10.10}$$

To see this, consider the time δt between the crests of the radar carrier signal at the transmitter. When the first crest arrives at the receiver it will have travelled distance $R_T(t + R_T/c) + R_R(t + R_T/c)$ and when the second crest arrives it will have travelled distance $R_T(t + R_T/c + \delta t) + R_r(t + R_T/c + \delta t)$ i.e. the second crest will have travelled the extra distance $(\mathrm{d}R_T/\mathrm{d}t + \mathrm{d}R_R/\mathrm{d}t)\delta t$. The time between crests will have increased by $(\mathrm{d}R_T/\mathrm{d}t + \mathrm{d}R_R/\mathrm{d}t)\delta t/c$ and so the frequency of the wave will have changed according to (10.10). The Doppler of a target will not always ensure that it can be distinguished from clutter since targets travelling transverse to the look direction will generate a low Doppler. Further, the dynamic nature of the sea will mean that its radar returns will also exhibit Doppler and could serve to mask slow-moving targets.

For many radars the transmit and receive antennas are closely located and we have what is known as a *monostatic radar* ($R_T = R_R$). This is obviously necessary for compact platforms such as ships and aircraft. Consequently, in such radars, care must be taken to ensure that the transmit and receive phases do not overlap, i.e. the receive phase will need to occur between the transmit pulses. However, a radar signal consisting of pulses is not always convenient and, in the case of continuous radar signal, the transmitter and receiver will need to be well separated in order to avoid what is known as direct signal interference or DSI for short.

A radar for which the transmitter and receiver are well separated is known as *bistatic radar* (see Figure 10.7a) and a particularly important example is the passive bistatic

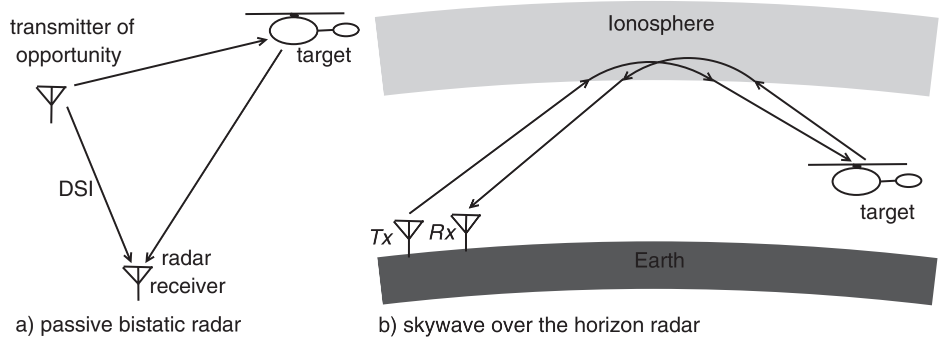


Fig. 10.7 Bistatic radar.

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

radar. In this case, the radar illuminator is a transmitter of opportunity, a broadcast transmitter for example. The signal from such a transmitter is usually continuous and so there are no quiet periods for reception, necessitating good isolation of the receiver from the transmitter. An important example of such an illuminator is that provided by the DAB system. In the UK, each transmitter has a power of the order of 10 kW, an operating frequency of around 222 MHz and a bandwidth of 1.536 MHz. For a continuous radar signal, the range resolution ΔR is related to the bandwidth B of the illuminator by $\Delta R = c/2B$ and so this example of a passive radar will have a range resolution of approximately 98 metres. However, as mentioned above, DSI is a major issue for such radars and the receive antennas will need good nulling in the direction of the transmitting antenna. However, a small amount of DSI is necessary for cross-correlation with the target returns for the purpose of target detection. Given the pressure on radio spectrum, passive radar offers a way of optimising the usage of spectrum by reusing radio signals created for other purposes.

Another form of bistatic radar is known as *over-the-horizon radar* (OTHR). These radars operate on frequencies in the HF band ($3 - 30$ MHz) and use sky-wave, or surface-wave, propagation in order to achieve over-the-horizon reach. Such radars usually transmit a *frequency-modulated continuous wave* (FMCW) signal in order to avoid the interference that pulse modulation can produce and typically have bandwidths between 10 and 50 kHz. To reduce the DSI that occurs due to ground-wave propagation, the transmitter and receiver will normally need to be separated by 100 km or more and the radar will need to operate at frequencies for which the receiver is inside the skip zone of the transmitter. Further, due to the dispersion caused by the ionosphere, sky-wave radars need to operate with relatively narrow bandwidths and consequently have limited range resolution. Nevertheless, skywave OTHR radar has been found to be an invaluable tool for long-range large surveillance.

For airborne radar, there is a further option known as *synthetic aperture radar* or SAR for short. This radar uses the flight path to simulate a long array antenna. A radar with relatively low resolution is located on the aircraft which, during flight, records radar returns (phase and amplitude) at regular intervals (see Figure 10.8). After a suitably large number of samples, these are combined in the fashion of an array. Returns from

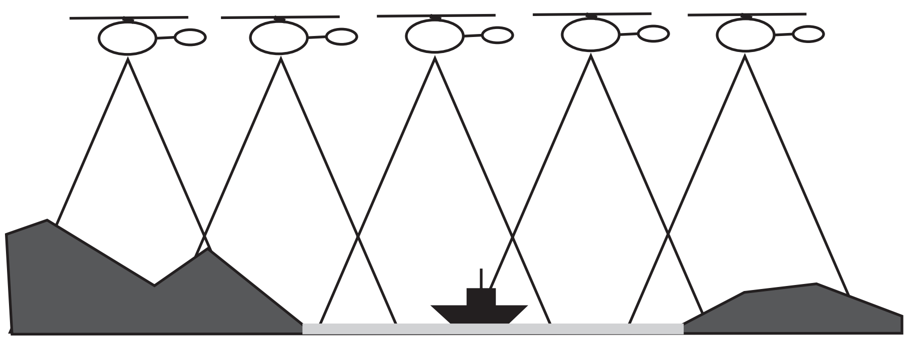


Fig. 10.8 A typical SAR scenario.

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

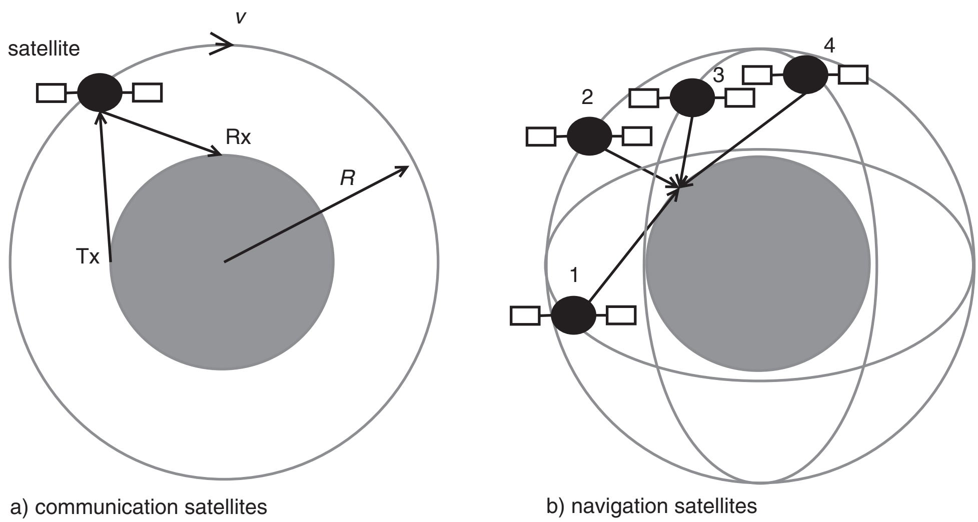


Fig. 10.9 Communication and navigation satellites.

a particular point on the ground are combined, with phase and amplitude corrections to account for the change in aircraft position during the observation period, to form a high-resolution image. It is clear that such radars work best with static, or at least slowly moving targets. However, moving targets can be handled with more sophisticated processing.

10.6 Satellite Systems

The first artificial satellite to be put into orbit around the Earth was Sputnik 1, launched by the Soviet Union in 1957. This satellite merely carried radio beacons (at frequencies of 20,005 MHz and 40.002 Hz) which only lasted for 21 days, but it started the era of radio in space. Satellites offered the possibility of very long-range over-the-horizon communication that did not depend on the ionosphere. Consequently, in 1962, the USA launched the first orbiting communications satellite known as *Telstar* and a new era of global communications started. Unfortunately, because of their orbital motion, satellites such as Telstar could only communicate for a limited time. Further, as we have seen with radar targets, a Doppler shift is imposed upon the signals and this needs to be compensated for. For a circular orbit, the speed v of the satellite is related to the radius of orbit R through $v = \sqrt{GM/R}$ where G is the gravitational constant ($G = 6.67 \times 10^{-11}\,\mathrm{N{\cdot}m^2/kg^2}$) and M is the mass of the Earth (5.98×10^{24} kg). If a satellite is placed in an orbit at an altitude of 35,785 km above the Earth, the angular speeds of the Earth and satellite will match and the satellite will be fixed relative to Earth's surface. This is known as a geostationary orbit. Geostationary satellites are able to provide continuous communication without any Doppler effect. However, since the satellites are at a great distance from Earth, they require powerful transmitters and suffer a significant time delay in communication.

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

Another major use of satellites is for navigation purposes, GPS (Global Positioning System) being one of the most well-known examples. The GPS system consists of a constellation of 32 satellites in circular orbits around the Earth at an altitude of 20,200 km. Each satellite carries a stable atomic clock which is synchronised with those on other satellites and ground stations. Further, each satellite continually transmits information about its position at the time on its clock (transmitted on frequencies 1575.42 MHz and 1227.60 MHZ). Providing a user has at least four satellites in view, he will be able to obtain an estimate of his current position and time. Consider a Cartesian coordinate system based at the centre of the Earth and let there be N satellites in view. Let a user at position (x, y, z) receive a message at time τ_i, according to his clock, about the position of satellite i, i.e. (x_i, y_i, z_i) at time t_i according to the satellite's clock. Then,

$$(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 = (\tau_i - \Delta - t_i)^2 c^2, \tag{10.11}$$

where Δ is the bias in the user's less-accurate clock. As there are then N equations for the unknowns (x, y, z) and Δ, the user will need four satellites in view ($N = 4$) for a solution. If there are more than four satellites in view ($N > 4$), the best fit solution to all N equations will usually give the most accurate estimate of position and time bias Δ. It is obvious that the ionosphere can cause some inaccuracy due to its effect upon the propagation of the satellite signal. However, the fact that the ionosphere is a dispersive medium, together with the dual frequency of GPS, allows a correction to be calculated.

Radio communication between the ground and a satellite suffer from an effect known as *Faraday rotation*. It turns out that the interaction of a radio waves with the ionosphere is quite subtle and is altered by the magnetic field of the Earth. This causes the electric field (and likewise the magnetic field) to rotate about the propagation direction as the wave propagates. Along a path between points A and B, the electric field will rotate through the angle

$$\phi = \frac{\pi}{c_0} \int_A^B \frac{f_p^2 f_H}{f^2} \cos\theta \, \mathrm{d}s, \tag{10.12}$$

where ω_p is the plasma frequency, ω_H is the gyro frequency and θ is the angle between Earth's magnetic field and the direction of propagation. (The gyro frequency is proportional to the magnitude of Earth's magnetic field and has a value around 1.4 MHz in the ionosphere.) This would be fine were it not for the fact that ionosphere is a quite dynamic medium that varies considerably with time. In particular, the plasma suffers from wavelike perturbations that are known as travelling ionospheric disturbances (TIDs). These disturbances are caused by waves in the neutral atmosphere that force plasma up and down Earth's magnetic field lines. Furthermore, they have wavelengths from tens of kilometres to many thousand kilometres and periods between minutes and hours. Through Faraday rotation, the passage of a TID will cause a sympathetic variation in the polarisation. Unfortunately, TIDs can be quite unpredictable and, as a consequence, so is the variation in polarisation. If the polarisation of the receive antenna is fixed, the received signal will suffer *polarisation fading* due to polarisation mismatch at the antenna. (It should be noted that the phenomenon of fading is also a significant problem for HF communications.)

In order to see how we can overcome the effects of Faraday rotation, we need to look at the phenomenon in more detail. Up until now we have tended to talk about fixed polarisations (horizontal and vertical referred to a surface such as that of the Earth), but this need not be the case. Maxwell's equations only require that the electric field be perpendicular to the propagation direction. If we consider a vertically polarised wave, we could regard this as the combination of two waves of equal magnitude, but with their electric fields rotating (angular speed ω) in opposite directions about the propagation direction (see Figure 10.10a). In mathematical terms, if we have propagation in the z direction and the linear polarisation is in the x direction,

$$\begin{aligned}
\mathcal{E} &= \Re\{E_0 \exp(\mathrm{j}(\omega t - \beta z))\hat{\mathbf{x}}\} \\
&= \Re\left\{\frac{E_0}{2}(\hat{\mathbf{x}} - \mathrm{j}\hat{\mathbf{y}}) \exp(j(\omega t - \beta z))\right\} + \Re\left\{\frac{E_0}{2}(\hat{\mathbf{x}} + \mathrm{j}\hat{\mathbf{y}}) \exp(\mathrm{j}(\omega t - \beta z))\right\} \\
&= \frac{E_0}{2}(\cos(\omega t - \beta z)\hat{\mathbf{x}} + \sin(\omega t - \beta z)\hat{\mathbf{y}}) + \frac{E_0}{2}(\cos(\omega t - \beta z)\hat{\mathbf{x}} - \sin(\omega t - \beta z)\hat{\mathbf{y}}) \\
&= \mathcal{E}_R + \mathcal{E}_L. \qquad (10.13)
\end{aligned}$$

The waves $\mathcal{E}_R$ and $\mathcal{E}_L$ are said to be circularly polarised (right or left according to the direction of rotation). When they are combined, the horizontal components cancel to leave only a vertical component. This component is said to be linearly polarised since its field is always parallel to a straight line (the vertical direction in this case). In general, any linearly polarised wave can be represented as the combination of circularly polarised waves. What is important is that the two circularly polarised waves travel at different speeds within the ionosphere. Consequently, if we split a linearly polarised wave into its components on entry to the ionosphere and combine them on exit, the polarisation of the combination will have changed on exit. If this happens in an unpredictable fashion we have a problem, as mentioned above. The solution, however, is to use a single circular polarisation. It is clear from (10.13) that a suitable antenna for circular polarisation could be made from the combination of two identical antennas that are 90° out of phase, one polarised in the $\hat{\mathbf{x}}$ direction and the other in the $\hat{\mathbf{y}}$ direction. Such an antenna could be made from a pair of orthogonal dipoles as shown in Figure 10.10b. To achieve the 90°

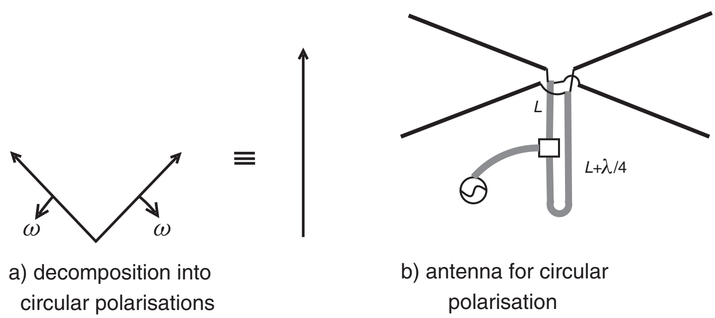


Fig. 10.10 Decomposition of linear polarisation into circular polarisations and a circularly polarised antenna.

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

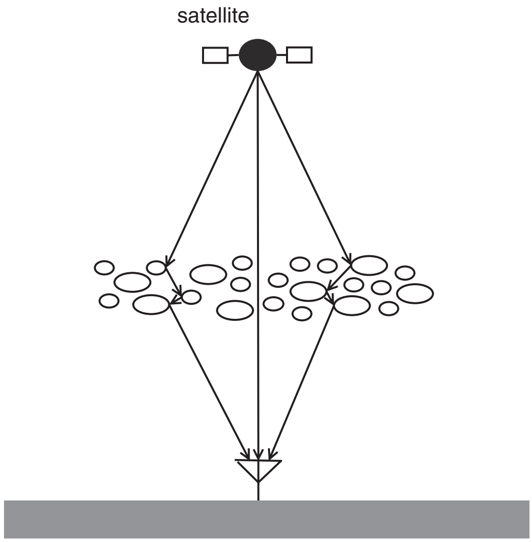


Fig. 10.11 The effect of turbulence upon satellite signals.

phase shift, the dipoles are fed from a power splitter, one by a transmission line of length L and the other by a line of length $L + \lambda/4$.

As we have mentioned in the previous chapter, both the neutral atmosphere, and the ionosphere, can be subject to turbulence that can cause irregular fluctuations in the refractive index. Because of this, there can be multiple propagation paths between the transmitter and the receiver (see Figure 10.11). This will cause there to be a spread of ranges between the receiver and transmitter, rather than one single range. Furthermore, if the random fluctuations are in motion, there will be a spread of Doppler shifts. In the equatorial, and auroral, regions the ionosphere is prone to turbulence that can, at times, be severe enough to cause appreciable range and Doppler spread. This phenomenon is known as *scintillation* and can severely compromise the operation of satellite systems. In the case of communication systems it can lower the capacity of a channel and in the case of navigation systems it can reduce the attainable accuracy. HF propagation can also suffer from the effects of scintillation and this can compromise the operation of HF communication systems and sky-wave OTHR.

10.7 Noise and Radio Astronomy

Although the last century has seen the technology of radio progress by leaps and bounds, factors such as propagation and noise still largely remain outside the control of the engineer. Our understanding of propagation has allowed us to choose the frequencies most appropriate to a radio system and modern electronics has allowed us to keep equipment noise down to an almost negligible level. However, external noise still remains a severe constraint. Figure 10.12 illustrates the various contributions to this

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

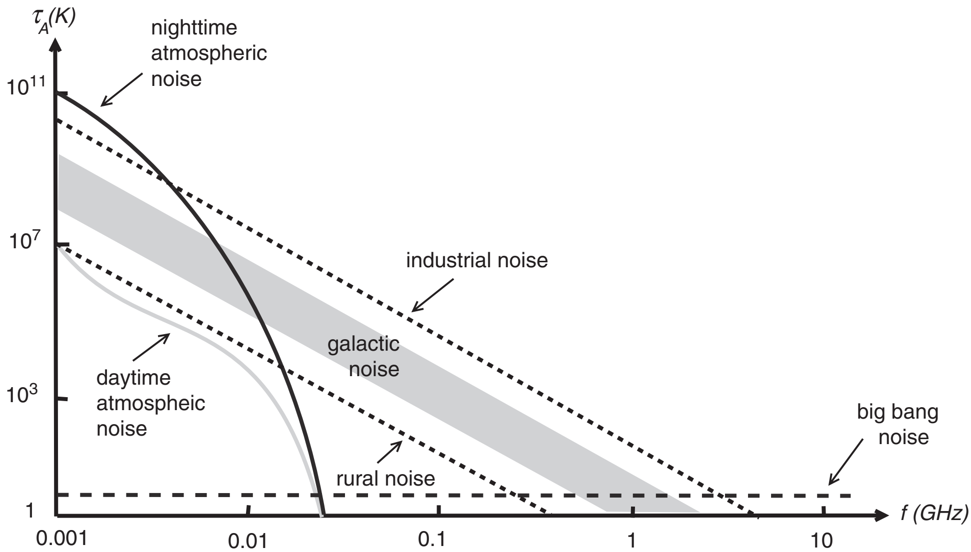


Fig. 10.12 Antenna temperature for external noise sources.

noise. Firstly, we live in an environment that is rich in man-made noise (car ignitions, plasma televisions and motors, to name but a few sources), and Figure 10.12 shows the typical variation of industrial noise with frequency. We might think we could do a lot better in the countryside, but Figure 10.12 also shows some typical man-made rural noise and it will be noted that its contribution is still appreciable. Besides man-made noise, there is also considerable noise from natural phenomena such as thunderstorms, known as atmospheric noise. These storms occur frequently and worldwide and their lightning flashes will generate considerable radio frequency energy across a broad spectrum. Such interference will propagate locally through the mechanism of surface waves and globally through refraction by the ionosphere. Consequently, at any location, there will be noise contributions from points across the planet. Furthermore, there is a considerable difference between the day and night thunderstorm contributions due to the difference between the daytime and nighttime ionospheric propagation. Due to its means of propagation, thunderstorm noise is highly directional and, as a consequence, its effect on a communications system can be highly dependent upon the type of receive antenna that is employed. Figure 10.12 also shows what is known as cosmic noise, i.e. noise coming from outside the Earth and its atmosphere. Here there are two major contributions: noise originating from the myriad of radio sources in the galaxy and the background cosmic radiation left over from the Big Bang at the start of the universe. The galactic noise is shown as a band since its strength depends on the view of the galaxy from Earth and this will change with time.

Obviously, the ionosphere blocks out cosmic radiation at lower frequencies, but it is the dominant source of naturally occurring noise at higher frequencies. At frequencies above 10 GHz, there are other problems that arise. Firstly, radiation from the sun can cause ionisation that causes further atmospheric noise due to discharge. Secondly, as

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

frequency rises there is increased absorption of radio waves. This has the same effect as a resistor in a circuit and adds noise to a radio signal.

It is clear that cosmic noise places a fundamental limit on the performance of radio systems at GHz frequencies and so the study of cosmic radio sources, known as *radio astronomy*, is of importance to radio. Indeed, radio astronomy had its beginning as a study of radio noise. In the early 1930s, Karl Jansky of Bell Laboratories in the USA made a study of radio noise with a directional antenna and found that the Milky Way was a significant source. Hence was born radio astronomy, but major developments in this area had to wait until after the second world war. However, it should be noted that, in the early 1940s, a radio amateur named Grote Reber built his own equipment and carried out detailed observations of the Milky Way on a frequency of 400 MHz. Using his equipment, Reber was able to make detailed radio maps of the sky. During the war, due to the development of radar, great strides were made in technology for GHz frequencies. Consequently, after the war, radio astronomy came into its own due to the ready availability of high-quality equipment. There are many good reasons for radio astronomy, but foremost is the limit placed by the atmosphere upon optical observations. As frequency rises the molecules in the atmosphere cause increased absorption of electromagnetic waves and this severely affects light (light is electromagnetic energy at frequencies of around 10^{15} Hz). Further, pollution from industrial emissions, and from man-made light emissions, has further exacerbated the problem. All of this has forced astronomers to site their optical telescopes on high mountains and even satellites. Consequently, astronomers have looked to the lower frequencies of radio waves in order to reduce these effects. This has been made all the more desirable by the discovery that the power of emissions by stellar sources varies with frequency f as $f^{-0.7}$.

The problem with radio astronomy is one of precision. To identify astronomical objects, we need to measure something that has an angular size that is a fraction of a second. Unfortunately, as we have seen in Chapter 8, the angular width of the main lobe of a radio antenna is of the order of $2\lambda/L$ where L is the overall size of the antenna. For practical antennas, however, there is a limit to L and hence to the angular precision. Since the signal strength is proportional to the antenna gain, we might swing the antenna around and look for the peak in the signal. The direction of the source is then the direction of the peak in the main lobe. However, around the peak the antenna gain is very insensitive to angular changes and so the direction of the source is difficult to ascertain with any precision. An alternative is to look for the directions where the signal has dropped to half power and then take the average. The half-power points in the antenna gain are very much more sensitive to changes in angle and so can be ascertained with much greater precision. All of this is fine when there is only one source within the main lobe, but in radio astronomy there is likely to be many sources within this lobe and the above procedure will not distinguish between them. The solution is to use a technique known as *interferometry*.

Consider two antennas that are separated by a distance D and pointed towards an astronomical source at elevation θ (see Figure 10.13). If a monochromatic wave arrives at antenna 2 with field strength $E = E_0 \cos(\omega t)$, then the output at antenna 2 will be $V_2 = h_{\text{eff}}(\theta) E_0 \cos(\omega t)$ and the output at antenna 1 will be $V_1 = h_{\text{eff}}(\theta) E_0 \cos(\omega(t - \tau))$

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

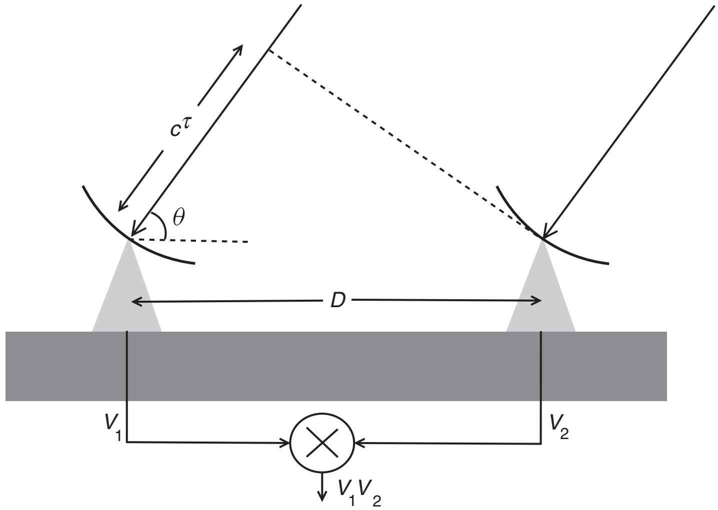


Fig. 10.13 Interferometric radio telescope.

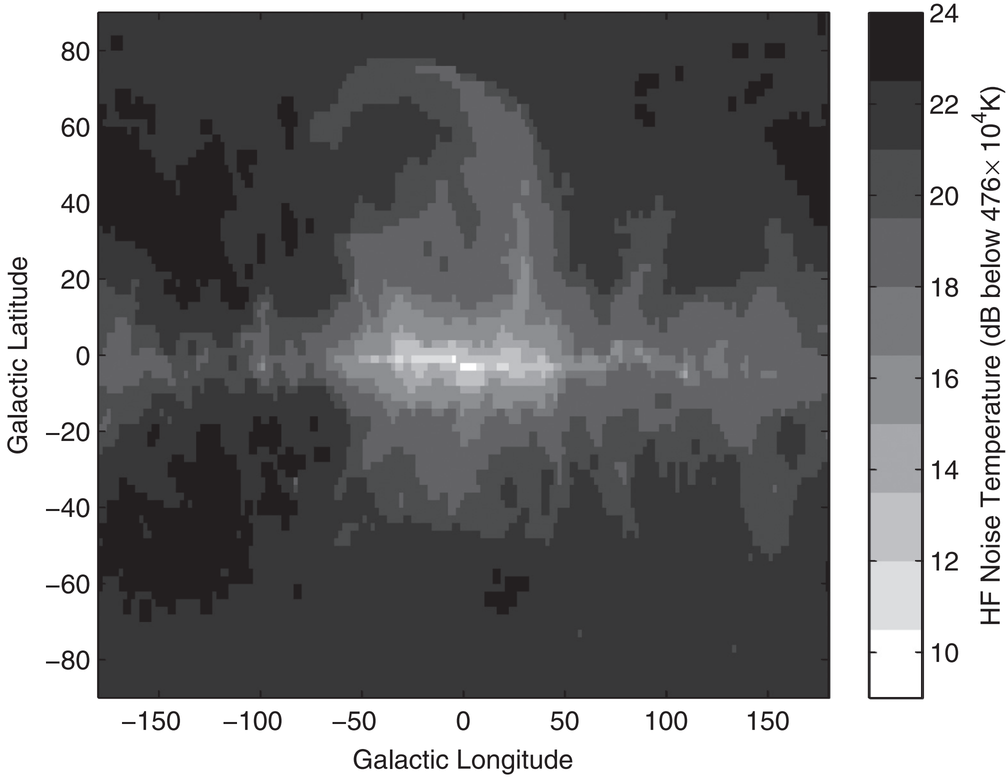


Fig. 10.14 The distribution of noise temperature across the Galaxy.

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press

where h_{eff} is the effective length of the antenna (note that this can be direction-dependent). The delay $\tau = D\cos\theta/c$ at antenna 2 is caused by the extra distance the wave must travel to this antenna. Consider the product of the two outputs, i.e.

$$V_1 V_2 = h_{\mathrm{eff}}^2 E_0^2 \left(\cos\left(2\omega t - \omega\tau\right) + \cos\left(\omega\tau\right)\right). \tag{10.14}$$

If we measure V_1 and V_2 over a number of periods, and average $V_1 V_2$ over this time, we obtain $h_{\mathrm{eff}}^2 E_0^2 \cos(\omega\tau)$. Then, if we divide this by the product of the amplitudes of V_1 and V_2 (i.e. $h_{\mathrm{eff}}^2 E_0^2$), we obtain a quantity $\Phi = \cos(\phi)$ where $\phi = \omega\tau$. Unfortunately, from a given value of Φ, we will only be able to ascertain ϕ up to a multiple of 2π. However, if the source has at least two frequencies (ω_1 and ω_2), and we find ϕ for these two frequencies (i.e. ϕ_1 and ϕ_2), we can obtain an unambiguous value for τ through $\tau = (\phi_2 - \phi_1)/(\omega_2 - \omega_1)$. In the situation that the interferometer is moving with respect to the source (this will happen due to the rotation of Earth), an alternative option is to measure the rate of change $\mathrm{d}\phi/\mathrm{d}t$ since this quantity will not exhibit an ambiguity. Quantity $\mathrm{d}\theta/\mathrm{d}t$ will be known from the motion of the interferometer with respect to the source and so θ can then be calculated from the time derivative of the relation $\phi = \omega D\cos\theta/c$. The major thing to be noted is that both ϕ and $\mathrm{d}\phi/\mathrm{d}t$ are proportional to the length of the baseline D and so the greater the baseline the more sensitive the measurement to changes in θ. This in turn leads to a more accurate estimation of θ. Figure 10.14 shows the kind of Galactic map that can be obtained with such observations. (This shows the distribution of noise temperature across the Galaxy at a frequency of 20 MHz.)

10.8 Conclusion

In the current chapter we have considered modern radio techniques for communications, surveillance and astronomy. The pressure on the radio spectrum has forced an emphasis on the optimisation of spectral usage and we have discussed techniques for achieving this end. For communications this has led to technologies such as spread spectrum, cellular radio and MIMO. In the sphere of radar, technologies such as passive bistatic radar allow the use of existing broadcasts as radar illuminators of opportunity (digital radio and digital television for example). Radio remains a dynamic technology and major developments in hardware have made possible techniques that could not have been contemplated several decades ago. OTHR is an example of a technology that is only possible because of the development of high-speed computing. Radio processing in the digital domain is now changing the face of radio and we are at the stage where both receiver and transmitter are digital almost to the antenna. This has heralded in a new era of software radio in which the function of hardware can be changed by simply changing the software.

https://doi.org/10.1017/9781108684514.011 Published online by Cambridge University Press