15 Communication

In this chapter we look at the issue of sending messages over very large distances using electromagnetic radiation. An interstellar spacecraft is likely to have a small, low-power transmitter. Diffraction will reduce the intensity received at the Earth to tiny amounts, making the signal hard to detect even with large receivers. This limits the achievable data rates significantly. We will see how to compute the power of the received signal, how to modulate messages into electromagnetic radiation, and how to use signal redundancy – error-correction coding – to reduce information loss due to noise. The received power may be so low that reception occurs in the regime of individual photon counting; we will examine how many bits of information photons can carry. We will also examine the choice of wavelength – in particular optical vs radio – and the relative merits of placing receivers on the Earth compared to in space.

15.1 Requirements and issues

An interstellar spacecraft must return data back to Earth. Assuming the spacecraft itself does not return, this means it has to transmit the data electromagnetically. The very large distances, combined with the limited transmission power, makes this a major challenge.

One of the first questions we face is the choice of communication wavelength. As the technology used to generate, transmit, and receive electromagnetic radiation varies considerably over wavelength, this choice influences all aspects of the system design. Broadly speaking we have two choices: optical, meaning visible plus ultraviolet and infrared wavelengths, and radio. To date, interplanetary spacecraft have all used well-developed radio communication techniques, usually operating in the X-band at 7–8 GHz (4 cm) or Ka-band at 32–34 GHz (0.9 cm). Shorter optical wavelengths have the significant advantage of much less diffraction, and also larger bandwidths. Although optical communication is used extensively for short ranges over fibre optic cables, its use over long distances in free space has been limited. A notable exception is the Deep Space Optical Communications (DSOC) experiment on the Psyche spacecraft, which in 2023 sent signals to the Earth from a distance of 3 au.

Much shorter wavelength X-rays could in principle be used for communication, as they have even less diffraction than optical wavelengths. But X-rays require more energy to generate, are harder to collimate, and the technology to encode/decode signals with them is much less developed. The Earth's atmosphere is also opaque to X-rays. We'll examine the issue of wavelength choice in more detail in section 15.8.

As we will learn, the data rates from an interstellar spacecraft will be low, so transmission times will have to be long. This is already the case for some interplanetary missions. NASA's New Horizons spacecraft, launched in 2006, returned data to the Earth at a rate of 1–2 kb s^{-1} after its 2019 encounter with Pluto, when it was 32 au from the Sun. The 6.25 GB of data it

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

gathered took around 15 months to transmit. For comparison, a single 1024×1024 uncompressed image with $2^8 = 256$ grey scales requires $1024 \times 1024 \times 8\,\mathrm{b} = 8.4\,\mathrm{Mb} \simeq 1\,\mathrm{MB}$. It is likely that the data transfer from an interstellar spacecraft would take many years.

A defining characteristic of interstellar spacecraft is their very large distances from the Earth, with light travel times of years. Other than some potential uplink from the Earth early in the mission to provide software or data updates, the communication will be entirely one way. The spacecraft will therefore have to operate autonomously, which also means it cannot coordinate with the Earth when it will send data, and Earth cannot request that corrupt data be resent. This is a major challenge if data are to be received over a duration of years. Even if we arrange for continuous contact with the spacecraft, perhaps with multiple receivers, there will be inevitable outages. An appropriate degree and timescale of redundancy therefore needs to be built into the communications protocol to accommodate not only noise but also longer-term signal loss.

Finally, the spacecraft may gather more data than it can ever transmit. In that case, the spacecraft's software needs to be intelligent enough to prioritize the data it sends, perhaps based on the quality of the collected data, or based on what it finds. We will consider autonomous control briefly in the next chapter.

15.2 Power received: the Friis transmission equation

One of the major problems of sending electromagnetic waves over large distances is diffraction, as this reduces the received intensity in proportion to the distance squared. Consider a circular transmitting aperture, such as a telescope mirror, of diameter d_t transmitting waves of power P_t. The telescope is focused to infinity, but due to diffraction the waves will spread out. This issue is the same as illuminating a sail with a laser, and as discussed in section 11.2.1, the beam pattern is no longer a point but an extended region of characteristic angular size $\alpha = a\lambda/d_t$, where λ is the wavelength. (We assume for now that the transmitter and receiver are stationary.) For a circular aperture, $a = 1.22$ marks the first minimum of the Airy disk, the region that contains 84% of the power (see figure 11.10). Thus if a 1 m transmitter emits radiation at 1 μm from the distance of Proxima Cen (4.25 ly), the Airy disk at the Earth has a radius of $ra\lambda/d_t \simeq 49$ million kilometres. Any realistically sized receiver can therefore collect only a tiny proportion of the transmitted power.

To calculate how much power is received, let us first assume for simplicity that the power from the transmitter in the Airy disk is distributed uniformly. The physical radius of the Airy disk at distance r is $s = \alpha r$. Assuming there is no loss of signal along the path, then the received intensity – the power per unit area – in the Airy disk is

$$I \simeq \frac{0.84P_t}{\pi s^2} = \frac{0.84P_t}{\pi \alpha^2 r^2} = \frac{0.84P_t d_t^2}{\pi a^2 \lambda^2 r^2}\,. \tag{15.1}$$

For a receiver of area A_r that is much smaller than the Airy disk and contained entirely within it, the power received is $P_r = IA_r$. Hence

$$P_r \simeq 0.23\,\frac{P_t A_t A_r}{\lambda^2 r^2} \quad \text{(uniform disk)}\,. \tag{15.2}$$

Of course, the intensity is not uniform over the Airy disk, so in general we instead have to integrate the intensity variation of the diffraction pattern (equation 11.28) over the area

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

of the receiver. In practice, though, the Airy disk is much larger than the receiver (tens of millions of kilometres compared to perhaps a few kilometres at most). Provided the receiver is aligned on the centre of the Airy disk, the received intensity is more or less constant at $I_0 = P_t A_t/\lambda^2 r^2$, as given by equation 11.29. In that case the received power is[1]

$$P_r = \frac{P_t A_t A_r}{\lambda^2 r^2} . \tag{15.3}$$

Equation 15.3 is the *Friis transmission equation*. The equation holds for non-circular transmitters and receivers too. It sometimes appears with a numerical factor in front, the value of which depends on the geometry and nature of the receivers. Sometimes the numerical factor is absorbed into the areas of the transmitter and/or receiver to give effective areas. This is often done in the context of radio communication where the concept of antenna gains is used. Other factors may be included to account for signal losses along the path, finite efficiencies of receivers, and relativistic effects (see section 15.2.1). Although equation 15.3 has been derived assuming monochromatic radiation, it also holds for signals with a finite narrow bandwidth.

The Friis transmission equation shows how the received power depends on the transmitted power, the wavelength, the areas of the transmitter and receiver, and the distance between them. The $1/r^2$ dependence comes from the power being spread over a spherical cap. This term is very significant for interstellar distances. Due to mass limitations on the spacecraft, both the power and area of the transmitter are likely to be small. The Friis transmission equation shows that we can compensate for these to some extent by using a large receiver on the Earth (or in Earth's orbit). The $1/\lambda^2$ term also has a big impact: changing the wavelength from 1 cm to 1 μm increases the received power by a factor of 10^8, all other things being equal.

Table 15.1 shows some example values for the various terms in the Friis transmission equation. In addition to the received power, it also shows the received signal photon rate,

$$\dot{n}_s = \frac{P_r \lambda}{hc} . \tag{15.4}$$

The first line of the table corresponds to receiving a signal from a modestly powered satellite in low Earth orbit with a 10 cm diameter receiver in the X-band. The second line corresponds approximately to New Horizons at Pluto. The subsequent lines are for radio and optical communications from Proxima Cen for different transmitter and receiver sizes at different wavelengths. The number of receivers, N_r, is also varied. These are combined incoherently, so this is equivalent to, but more practical than, using a single larger receiver.

Using a 1 m diameter radio transmitter and a 70 m receiver, we see that the power received from the interstellar spacecraft is many orders of magnitude lower than what could be received from New Horizons. But by using a larger transmitter and 1000 receivers, we can attain a received power that is only about a thousand times lower than New Horizons. This is a lot of receivers, but perhaps is not the hardest or most expensive part of interstellar travel.

[1] This assumes our receiver measures all of the light over area A_r. The arriving wavefront is plane. Thus when the receiver is an imaging system (like a telescope), the transmitter can be considered as a point source, and this is imaged with some diffraction pattern on the detector (perhaps another Airy pattern). In the derivations, I am assuming that all of the light in this diffraction pattern is measured to obtain P_r. It may be the case, however, that in order to suppress light from the target star next to the transmitter, we only measure the signal in the central part of the diffraction pattern, as shown later in figure 15.3. In that case P_r would be reduced.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

Table 15.1 Received power P_r and photon counts $\dot{n}_s$ for different transmitter–receiver configurations, as computed from the Friis transmission equation (equation 15.3). This adopts a circular transmitter of diameter d_t and N_r circular receivers, each of diameter d_r and combined incoherently. In all cases the receiver is much smaller than the size of the transmitted beam's Airy disk at the receiver.

P_t [W]	d_t [m]	d_r [m]	N_r	λ	r	P_r [W]	$\dot{n}_s$ [s^{-1}]
10	1	0.1	1	3.6 cm	500 km	1.9×10^{-10}	3.4×10^{13}
24	2.1	70	1	3.6 cm	32 au	1.1×10^{-17}	1.9×10^{6}
10	1	70	1	3.6 cm	4.25 ly	1.4×10^{-26}	2.6×10^{-3}
10	1	70	1000	3.6 cm	4.25 ly	1.4×10^{-23}	2.6×10^{0}
10	10	70	1000	3.6 cm	4.25 ly	1.4×10^{-21}	2.6×10^{2}
10	10	70	1000	1 cm	4.25 ly	1.9×10^{-20}	9.4×10^{2}
10	1	10	1	1 μm	4.25 ly	3.8×10^{-19}	1.9×10^{0}
10	4	10	1	1 μm	4.25 ly	6.1×10^{-18}	3.1×10^{1}
10	1	30	1	1 μm	4.25 ly	3.4×10^{-18}	1.7×10^{1}
10	10	30	1	1 μm	4.25 ly	3.4×10^{-16}	1.7×10^{3}
10	1	30	1	0.25 μm	4.25 ly	5.5×10^{-17}	6.9×10^{1}
10	1	10	100	1 μm	4.25 ly	3.8×10^{-17}	1.9×10^{2}

Operating in the optical (1 μm) at the same power (10 W) with the same size transmitter (1 m) and using a 30 m telescope as a receiver, we receive much more power: just a third that received from New Horizons when it was 8000 times nearer to the Earth. Note, however, that the photon rate is much lower in the optical because, for a given power, each individual photon carries more energy. This is particularly relevant if reception is done in the photon-counting regime, which we discuss later.

In all of the interstellar examples listed in the table, the power at the receiver is reduced relative to the power at the transmitter by 17 to 27 orders of magnitude.[2] The received signal photon rates are only of order 1–1000 per second, unless much larger transmitters, receivers, or transmission powers can be used. How these photon rates corresponds to data rates (bits per second) depends on a number of other factors, in particular the amount of noise and how messages can be encoded into electromagnetic waves, which we will look at later in this chapter.

15.2.1 A relativistic correction

So far we have been assuming that the transmitter and receiver are stationary with respect to one another. This is not the case for a flyby mission, in which relativistic effects need to be considered on account of the large relative velocities.

[2] In physics we tend to think of increases or decreases in terms of factors, or orders of magnitude. In engineering one often uses decibels dB, which are tenths of orders of magnitude. A factor f corresponds to $10 \log_{10}(f)$ dB. Thus 20 dB corresponds to a factor of 100, −10 dB to a factor of 0.1, and so on. Use of dB is convenient when there are many loss and gain terms that need to be combined, as it replaces multiplicative terms with additive ones. The loss of the signal described by equation 15.3 is $f = P_r/P_t$, which for the bottom line of Table 15.1 is −174 dB.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

If we go back over the derivation of the Friis transmission equation, we see that relativistic effects must be taken into account in two places. These are the same considerations as for a moving laser sail in chapter 11, here for a receiver receding from the transmitter at velocity β. First, following equation 11.27, the power received in the receiver's rest frame is reduced relative to the transmitted power by a factor of $(1-\beta)/(1+\beta)$, the square of the 'Doppler factor'. Second, the diffraction angle in the receiver's rest frame is increased by the Doppler factor due to aberration, as given by equation 9.5. This angle enters as the square in equation 15.1. Taking both of these effects into account, the relativistic version of equation 15.3 is

$$P_r = \frac{P_t A_t A_r}{\lambda^2 r^2} \left(\frac{1-\beta}{1+\beta} \right)^2 . \tag{15.5}$$

Relativistic effects therefore reduce the power on the receiver moving at $\beta = 0.1$ to 67% of that for a stationary receiver. For $\beta = 0.2$, the power is reduced to 44%. These factors are significant, but not dramatic for our purposes, given that we are often ignoring numerical factors of similar size, such as those arising from aperture geometry or path losses.

The wavelength in equation 15.5 is the wavelength measured in the transmitter's rest frame (as is the distance r). (I considered the diffraction to occur in the transmitter's rest frame and then applied the Lorentz transformation to the received angle.) The wavelength of the signals in the receiver's rest frame is redshifted by a factor of $\sqrt{(1+\beta)/(1-\beta)}$ due to the Doppler effect (equation 9.12).

We derived a very similar result to the one above in equations 9.16 and 9.17, where we saw that the radiance of a moving source is modified by a factor of ψ^4. This equals $[(1-\beta)/(1+\beta)]^2$ for a radially moving source (ψ is defined in equation 9.7).

15.3 Spacecraft pointing accuracy

In deriving equation 15.3, we assumed that the receiver was pointed exactly at the spacecraft, so that the receiver is centred on the Airy disk. When transmitting at $\lambda = 1\mu$m with a $d_t = 1$ m aperture, the diffraction angle is $1.22\lambda/d_t = 0.25''$. The spacecraft needs to be able to point to the Earth with an accuracy much better than this. If it cannot, the receiver will only see the edges of the diffraction pattern, and the received power will be greatly reduced (see figure 11.10).

Numerous forces act on a spacecraft that could change its orientation, including stellar radiation pressure and incident particles, as well as the anisotropic radiation of onboard heat. The spacecraft must, therefore, actively track the receiver. Today's space telescopes can track bright sources much more accurately than this fiducial value of $0.25''$ for extended periods. The James Webb Space Telescope, for example, can track stationary targets to an accuracy of $0.001''$ and targets moving at $0.003'' \, s^{-1}$ to an accuracy better than $0.01''$. Thus an interstellar spacecraft could probably track the bright Sun to sufficient accuracy, provided it also has some means to manoeuvre.

Tracking the Sun is a start, but the Earth, and therefore the receiver, orbits the Sun. As seen from Proxima Cen, the Earth therefore changes its angular position relative to the Sun by $\pm 0.77''$ over the course of the year. This is larger than the example diffraction angle of $0.25''$ mentioned above. Thus to remain centred on the Airy disk, the spacecraft will need to know where the Earth is in its orbit at any time. As it cannot see the faint Earth – and even

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

if it could, that would provide its position 4.25 years before signals arrive – the spacecraft instead needs to predict its position. For this the spacecraft needs to know the time on Earth, more specifically the barycentric coordinate time (TCB). The maximum angular velocity of the Earth relative to the Sun as seen from Proxima Cen is $0.013''$ per day, so the TCB needs to be known to an accuracy of about a day or two. An accurate atomic clock on board the spacecraft, synchronized to the TCB at launch, can in principle maintain this time to one part in 10^{12} (see section 14.2.3). This corresponds to an error of 2 ms after a journey of 50 years. However, by travelling at relativistic speeds, this clock would suffer from time dilation and so no longer track the TCB. If the spacecraft can keep track of all its velocity changes, then it might be able to correct the clock to the TCB. Another option is to uplink the TCB to the spacecraft from the Earth just before it starts transmitting. This would have to accommodate the long light travel time, of course, for which an accurate position, and thus good navigation, is also required (chapter 14). As we saw in section 14.4, stellar astrometric navigation could instead provide the TCB to the required accuracy on the spacecraft without needing any contact to Earth.

If the spacecraft cannot achieve a high pointing accuracy for extended periods, there will be an unpredictable reduction in received power and thus a potential loss of transmitted information. One solution to this is to relax the pointing requirements by using a larger diffraction pattern, for example by using a longer wavelength or a smaller transmitter. This would also lower the received power (and data rate), but the variation in received power would be reduced and thus unexpected signal losses mitigated. We could then compensate for the larger diffraction angle by using a larger transmitted power or a larger receiver area, as per the Friis transmission equation.

15.4 Modulation

For an electromagnetic wave to carry information, it must vary in some way. This could be a variation in amplitude, frequency, phase, or polarization. How we vary these to imprint a message on the signal is the topic of modulation. For now we assume that we have enough power to form continuous waves, so we can ignore the quantization of radiation.

To illustrate the concept of modulation, we start with an electromagnetic wave at a single frequency, the so-called carrier wave. By varying the amplitude of the wave, with different amplitudes corresponding to different letters in an alphabet for example, we can send a message. This is the basis of amplitude modulation (AM). We can instead vary the frequency by small amounts compared to the carrier wave frequency. This is frequency modulation (FM). They are illustrated in figure 15.1. These analogue modulation schemes were used extensively in terrestrial radio broadcasts in the past, where AM carrier waves typically had frequencies between 0.5 MHz and 2 MHz (medium frequency bands), although carrier waves over a broader range of 0.1–20 MHz were also used. FM radio used higher frequencies, typically 90–110 MHz.

An FM signal of course requires a finite bandwidth for modulation (frequency changes) to occur. In principle an AM carrier wave could be monochromatic, but only if it had an infinite duration. From Fourier transforms we know that a signal of finite duration must occupy a finite frequency range. Moreover, in order to send a message, each amplitude level can only last a finite amount of time, and therefore also occupies a range of frequencies.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

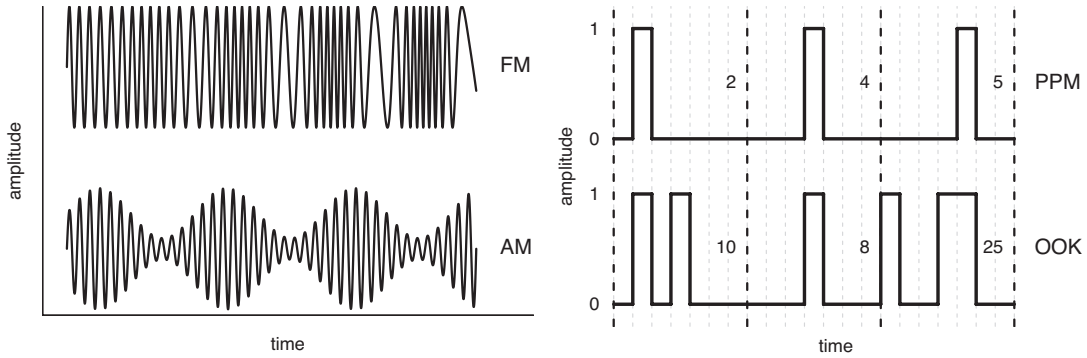


Fig. 15.1 Modulation schemes: analogue schemes AM and FM (left) and digital schemes OOK and PPM (right). In both digital schemes shown, $m = 7$ slots make up a frame. With PPM, only one slot is activated, so the three frames shown could represent the numbers 2, 4 and 5. With OOK, multiple slots can be activated and are interpreted as bits of a binary number. The three frames represent $2^1 + 2^3 = 10$, $2^3 = 8$ and $2^0 + 2^3 + 2^4 = 25$.

No matter what the modulation scheme, to send a signal we require a finite bandwidth. All other things being equal, the larger the bandwidth, the more variation of the carrier wave that can be distinguished, and so the larger the rate at which we can send data. An important result from signal processing quantifies this statement. Suppose we are trying to modulate a number of pulses in a signal that has bandwidth B (in Hz). The smaller the bandwidth, the broader each pulse in time. The largest number of independent pulses that can be resolved (per second) is $2B$, which is known as the *Nyquist rate*. If we send pulses at a higher rate, we can no longer distinguish between them. Another way of thinking about this is, if we want to send more data per second, the signal must vary more rapidly. To properly record this, we need to measure the signal more frequently, which requires a larger bandwidth.

AM and FM are examples of analogue modulation schemes, whereby a carrier wave is modulated by a continuously varying signal. In digital modulation, the carrier wave is modulated by a digital signal. The digital equivalent of AM is amplitude shift keying (ASK), in which the amplitude changes are confined to a number of discrete levels. Similarly, frequency shift keying (FSK) encodes a signal as discrete changes in the frequency. Digital modulation has now superseded analogue modulation for most purposes. Digital modulation is also suitable when intensities are so low that we count individual photons.

To compute the data rate for a particular modulation scheme, we need to know how many different levels m it uses. The smallest number of levels is a binary, $m = 2$, scheme. This means that we send 0s or 1s encoded, for example, as two different amplitudes or two different frequencies. Each pulse then conveys one bit of information. If we instead have $m = 4$ different levels, then we can send the numbers 0, 1, 2, 3, which we can interpret in binary terms as 00, 01, 10, and 11, which is two bits of information. In general, the number of bits we can send per pulse for m levels is $k = \log_2 m$. If pulses are sent at rate f_p, the data rate in bits per second is $R = f_p k = f_p \log_2 m$. The maximum pulse rate for a channel of bandwidth B is the Nyquist rate, $f_p = 2B$. Hence the maximum data rate, or *channel capacity*, is

$$C\,[\mathrm{b\,s^{-1}}] = 2B \log_2 m \quad \text{(noise-free)}. \tag{15.6}$$

This suggests that, even with a finite bandwidth, the channel capacity can be made arbitrarily large if m is made arbitrarily large. For example, when using FSK modulation, then in theory

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

we could separate the levels by vanishingly small frequencies and so fit arbitrarily many of them into a channel of any finite bandwidth. In practice, noise limits the number of levels we can distinguish, and this places a limit on the channel capacity if we are not to incur errors. We shall look at this in section 15.7.2.

There exist many digital modulation schemes. One of the simplest is on–off keying (OOK), show in figure 15.1. This transmits a binary message as a series of pulses in slots of fixed time width T_s. A pulse indicates a 1 and no pulse indicates a 0. A series of m slots makes up a frame, which can be interpreted as an m-bit binary number. This in turn can be interpreted as a codeword via a codebook. For example, to send ASCII (American Standard Code for Information Interchange) characters, a scheme of 128 different characters (codewords), we require $\log_2 128 = 7$ bits and thus $m = 7$ slots in a frame. A series of frames then makes up a message. In the absence of noise, OOK conveys m bits of information per frame. Each pulse, and thus OOK as a whole, requires a bandwidth of order $1/T_s$.

Another digital modulation scheme is pulse position modulation (PPM), also shown in figure 15.1). Like OOK, a message consists of a series of frames, and each frame is again divided up into m equal-duration time slots. But in PPM, only one slot in a given frame is activated (pulse sent). This allows us to convey one of m different codewords. For example, with $m = 128$ we can convey the ASCII characters. Assuming each slot has an equal probability of activation, this conveys $\log_2 m$ bits of information per frame on average.

When the duration of a single slot in either scheme is T_s, then in OOK a frame has a duration of $T_s \log_2 m$, whereas in PPM a frame has a duration of $T_s m$. Thus with a common slot (pulse) duration, OOK has a larger data rate by a factor of $m/\log_2 m$ (in the absence of noise). Alternatively, to achieve the same data rate in both schemes, PPM requires shorter pulses, and thus requires a larger bandwidth by a factor of $m/\log_2 m$.

One reason we may nonetheless choose PPM over OOK is that PPM usually requires less average power, and is usually more energy efficient. To see this, consider both schemes transmitting a signal at the same bit rate, so PPM has shorter slots. PPM requires energy to activate just one slot, whereas OOK must activate any number of slots. The number depends on the message sent, but on average it is more than one. Assuming we send the same number of photons per slot in the two schemes, PPM requires fewer photons, and thus energy, per frame, so lower average power. However, as its slots are shorter, it will have to send these photons more quickly, so needs more peak power to send a pulse. These considerations are very relevant for interstellar communications, where spacecraft are likely to be power limited.

15.5 Noise

When discussing the Friis transmission equation in section 15.2, the implication was that a larger received power – equivalently a larger received photon rate at a given wavelength – permits a higher data rate. This is only part of the story. The data rate also depends on the noise, because this affects our ability to distinguish between the levels in the modulation scheme.

The effect of noise is shown schematically figure 15.2. The signal is created by modulating the intensity of light between two levels that represent the two binary digits. But there are many things that could affect this signal by the time it arrives at the receiver. The transmitter's power may not be stable, there may be a contaminating light source, the atmosphere

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

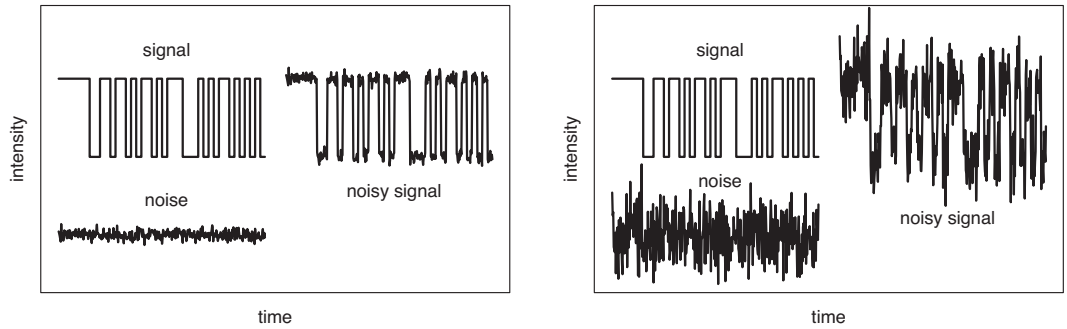


Fig. 15.2 A demonstration of noise. In the left panel, the message is encoded as a signal by using two intensity levels, show at the top. The additive/subtractive noise is below, and their sum – the received noisy signal – is shown to the right. The right panel shows the same signal, but with a higher noise level.

may scatter some of the signal, or electronics may emit radiation, to mention just a few examples. These phenomena can be considered as random additions to or subtractions from the signal. This is noise and, as we see in figure 15.2, it corrupts the signal, making it harder to distinguish the two levels. If the noise is large enough, we may not be able to reconstruct the original signal correctly.

Noise is characterized not by its overall intensity, but by its degree of fluctuation. This is because if a contaminating source is bright, but constant, we still see the same variations in our signal, just at a larger overall intensity. We could subtract that constant – by measuring it when sending no signal – and so retrieve the original signal. Thus when we refer to the degree of noise, we do not mean the average power, but rather its standard deviation.

Our ability to identify the original signal depends, therefore, not so much on the power of the signal – the difference between the levels in figure 15.2 – but on how large this power is compared to the noise. This is the signal-to-noise ratio (SNR).

For many noise sources, the average power of the noise and its standard deviation are not independent. This is the case for photon counting, because photons are emitted by a random quantum process that can be described by Poisson statistics. In our two-level signal example, suppose that a contaminating source (e.g. the star seen near the spacecraft) contributes on average n_b photons to each pulse. (Such contaminants are conventionally referred to as the 'background', motivating the 'b' subscript.) Due to the random nature of light emission, not every pulse contains exactly n_b photons. Instead, the number of photons per pulse follows a Poisson distribution with mean n_b. The variance of a Poisson distribution equals its mean. The standard deviation, which is the noise associated with the pulse, is therefore $\sqrt{n_b}$.

The source of the original signal is subject to the same statistics, so even in the absence of any contaminating sources there is noise in the signal itself. Denoting the average number of photons in the signal pulse with n_s, this 'self-noise' is $\sqrt{n_s}$. In the absence of any background ($n_b = 0$) the SNR is therefore $n_s/\sqrt{n_s} = \sqrt{n_s}$. This is an important result, because it shows that although the signal itself contains noise, the SNR is increased by using a larger intensity. This is essentially why we want to maximize the received power from our interstellar spacecraft.

When there is a background, we will measure $n_m = n_s + n_b$ photons (over a time slot). To extract the signal, we need to subtract the background. As we cannot distinguish background photons from signal photons, we cannot do this directly. What we typically do is measure

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

the background in the absence of the signal over many time slots in order to determine its average $\overline{n_b}$. Subtracting this from the measurement, our estimate of the signal is $n_s + n_b - \overline{n_b}$. This is the sum of three independent Poisson-distributed variables, which is itself a Poisson-distributed variable with the variance equal to the sum of the individual variances. When $\overline{n_b}$ is measured so accurately as to have a negligible variance, the variance in our estimate of the signal is $n_s + n_b$. The SNR of our signal estimate is

$$\text{SNR} = \frac{n_s}{\sqrt{n_s + n_b}} . \tag{15.7}$$

This tells is that if we have an intense background, then even if its average can be perfectly subtracted in a statistical sense, its fluctuations – which scale as $\sqrt{n_b}$ – still contribute to reducing the SNR of the estimated signal. However, because the signal in the numerator grows linearly with n_s, whereas the noise in the denominator grows only with the square root of the number of photons, gathering more photons still increases the SNR.

The actual calculation of noise and SNR in real situations is more complicated than has been described here. The main point is that we want to have as many signal photons and as few background photons as possible in order to maximize the SNR. The situation is also more complicated when it comes to single photon counting, something we will examine separately later.

15.6 Contaminating light from the target star

When a spacecraft passes Proxima Cen at a distance of 0.1 au, their angular separation as seen by an Earth-based receiver is just 75 mas. Both the star and signal transmitted by the spacecraft deliver plane wavefronts of uniform intensity to the distant receiver. The star is much brighter than the spacecraft's signal (quantified below), so following the discussion from the previous section, the noise contaminating the signal will be very large.

One way to reduce the contamination is to use a wavelength filter. A star emits light over a very broad range of wavelengths. If we can confine our signal to a narrow bandwidth, we can use a narrow filter centred on the signal to remove most of the starlight. If the transmitter is an optical or infrared laser, then a bandwidth of less than 1 nm is plausible.

In principle we could receive the signal using a bare detector, but this would receive light from the entire sky, which includes a lot of more contaminating sources beside the contaminating target star. We could use a baffle to narrow this down. But in practice we would use an imaging telescope as a receiver to spatially separate the spacecraft's signal from the star. Let us calculate what improvement this provides.

Both star and spacecraft can be considered as point sources. When the telescope is diffraction-limited, each point source will form an Airy diffraction pattern on the focal plane of the telescope. For a telescope with an aperture of 3 m diameter receiving at 1 μm, the diffraction angle is about 70 mas, which is similar to the geometric angular separation between the star and the spacecraft mentioned above. This configuration is shown in figure 15.3, where the star is only five times brighter than the spacecraft's signal (we'll see below how we may achieve this). We could only get close to the telescope's diffraction limit when observing above the atmosphere or when using adaptive optics (see section 11.2.3). Otherwise the telescope is seeing-limited, and the images on the focal plane will be broader, and the overlap larger.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

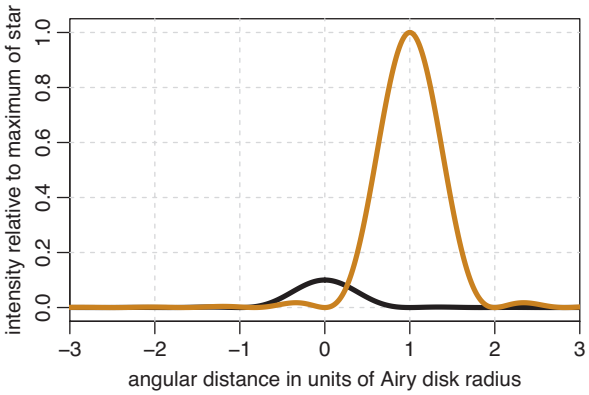


Fig. 15.3 Overlapping Airy diffraction patterns. When the receiver is an imaging telescope, the diffraction pattern of the on-axis faint signal from the spacecraft's signal will probably be significantly overlapped by the diffraction pattern of the off-axis brighter star. In the case shown the two sources are separated by one Airy disk radius and the star is five times brighter than than the signal.

We already saw in section 15.2 how to compute the power of the signal at the receiver. We can compute the power from the star at the receiver as follows. Let I_λ be the spectral intensity (units $\mathrm{W\,m^{-2}\,m^{-1}}$) at the star's surface. For an isotropically radiating star of radius $R_\star$ observed from distance r, the spectral intensity at the receiver is $I_\lambda (R_\star/r)^2$, assuming no losses along the path. Integrating this over the area A_r of the telescope's aperture gives the spectral flux (units $\mathrm{W\,m^{-1}}$) as $I_\lambda A_r (R_\star/r)^2$. To block out as much of this light as possible, the receiver only observes over the narrow bandwidth $\Delta\lambda$. The power (units W) received at the telescope from the star is therefore

$$P_{r,\star} = I_\lambda \Delta\lambda A_r \frac{R_\star^2}{r^2} \,. \tag{15.8}$$

This power is distributed according to the diffraction pattern of the intensity given by equation 11.28.

If we approximate the star as a black body of effective temperature $T_\star$, then the star's spectral radiance – power emitted per unit surface area per unit solid angle per unit wavelength – is given by the Planck function,

$$B(\lambda, T_\star) = \frac{2hc^2}{\lambda^5} \left[\exp\left(\frac{hc}{\lambda k T_\star} \right) - 1 \right]^{-1} \,. \tag{15.9}$$

Each surface element of the star emits isotropically over the half sphere, so integrating the Planck function over this gives its spectral intensity,

$$I_\lambda = \int B(\lambda, T_\star) \cos\theta \, d\Omega = B(\lambda, T_\star) \int_0^{2\pi} d\phi \int_0^{\pi/2} \cos\theta \sin\theta \, d\theta = \pi B(\lambda, T_\star) \,. \tag{15.10}$$

For a black body we can use the Stefan–Boltzmann law to express the stellar radius in terms of $T_\star$ and the stellar luminosity $L_\star$:

$$L_\star = 4\pi \sigma_{\mathrm{SB}} R_\star^2 T_\star^4 \,, \tag{15.11}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

where σ_{SB} is the Stefan–Boltzmann constant. Thus for a black body we can write equation 15.8 as

$$P_{r,\star} = \frac{B(\lambda, T_\star)\Delta\lambda A_r L_\star}{4\sigma_{SB} T_\star^4 r^2} \quad \text{(black body)}. \tag{15.12}$$

Let us compute the received power per unit area of receiver for our nearest star. Proxima Cen has $L_\star = 0.0015\,L_\odot$, $T_\star = 3000\,K$, and $r = 4.25\,ly$ (section 2.5.2). At $\lambda = 1\,\mu m$ we then have $B(\lambda, T_\star) = 990\,W\,m^{-2}\,nm^{-1}\,str^{-1}$. Adopting a bandwidth of $\Delta\lambda = 1\,nm$, we get $P_{r,\star}/A_r = 1.9 \times 10^{-14}\,W\,m^{-2}$, which is 97 000 photons $s^{-1}\,m^{-2}$.

When our spacecraft at Proxima Cen has a 1 m diameter telescope transmitting at 10 W, then using the Friis transmission equation (equation 15.3), this delivers an intensity at the receiver of $P_r/A_r = 4.9 \times 10^{-21}\,W\,m^{-2}$, or 0.024 photons $s^{-1}m^{-2}$. The power from the star is four million times larger than the power from the signal. Without further mitigation, it would be very hard to extract the signal from the (noisy) glare of the star.

This is where the spatial separation provided by the imaging system comes in. The ratio of four million assumes that, to extract the signal, we have to integrate the light over the entire diffraction pattern of both signal and star in the image plane. But if they are spatially separated, as in figure 15.3, we can integrate instead in just a narrow region centred on the Airy disk of the signal. Of course, if the Airy pattern of the star is four million times higher than the one of the signal, and not five times as shown in the figure, the small spatial separation shown won't help much on its own. But we can now introduce an aperture mask, or coronagraph, centred on the star to suppress its light before it gets onto the focal plane. This is routinely done in astronomy to look for the faint light of exoplanets orbiting stars. Suppression factors of 10^6 have been achieved for separations of a few tenths of an arcsecond, for example using the SPHERE instrument on the VLT. With this suppression factor, the light from Proxima Cen would be reduced to a level similar to that of the signal in the above example. This increases the SNR of the extracted signal significantly and, therefore, the channel capacity (data rate), as we shall see in the next section.

Higher suppression factors can probably be achieved for telescope receivers in space, which will improve things further. A factor of 10^8 to 10^9 at an angular distance of several λ/d_r is predicted for the upcoming Roman Space Telescope. Furthermore, in the calculation above I assumed a low-power transmitter, 10 W, which is what we might expect to have available on a low- or intermediate-mass sailcraft. If our spacecraft has instead used nuclear propulsion to get to the target star, there is likely orders of magnitude more power available, some of which could be used for the communication system, boosting the SNR further.

Bringing everything together, the ratio of the power received from the star to that received in the signal is

$$\frac{P_{r,\star}}{P_r} = \frac{L_\star B(\lambda, T_\star)\lambda^2\Delta\lambda}{4\sigma_{SB} T_\star^4 P_t A_t C} \quad \text{(black body)}, \tag{15.13}$$

where C is the suppression factor of the coronagraph. The dependence of this power ratio on wavelength is complicated by the presence of the Planck function. But for $T_\star = 3000$ K, the ratio decreases with decreasing wavelength for wavelengths below about 1 μm. A narrower

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

bandwidth reduces the power ratio, because the star emits over a broad spectrum, whereas the signal only emits over the specified bandwidth.[3]

The dependence of the power ratio on the properties of the star – $T_\star$ and $L_\star$ – is not obvious, because both the Planck function and $L_\star$ are strong functions of temperature. But on the main sequence, cooler stars have much lower luminosities than hotter stars so, all other things being equal, we achieve a higher data rate from a spacecraft at Proxima Cen than from a spacecraft at Alpha Cen.

Finally, note that both the receiver area A_r and distance r cancel in the power ratio. This does not imply we do not want a large receiver or a nearby target. Collecting more photons from the signal, even if that increases the number of photons from the star by the same factor, still increases the SNR, as discussed in section 15.5.

15.7 Communication in the presence of noise

Interstellar communication with photons is an example of sending signals over a noisy channel. Whatever the message, and however it is encoded into a signal, this needs to be transmitted, received, and then decoded, as outlined in figure 15.4. The channel is everything between the transmitter and receiver. This can include the transmitters and receivers themselves – both their electronics and optics – as well as the medium between them, in our case interstellar space, but in terrestrial communications this could be cables or optical fibres. Due to noise in the channel, the received signal is not identical to the transmitted signal, as discussed in section 15.5.

In its simplest form, *coding* is the technique of converting a message into a transmittable signal, or code.[4] Suppose we want to send text over a binary channel, which is one that can only transmit two different signals, for example 0 and 1. We need to define a codebook that

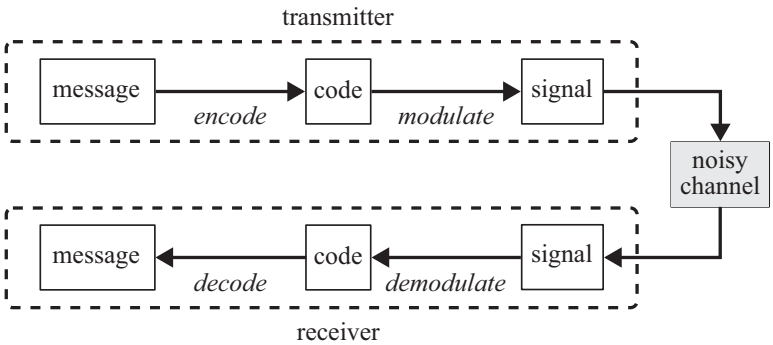


Fig. 15.4 The concept of a noisy channel. A message is encoded and modulated before being sent over a noisy channel that may corrupt it. The received signal is then demodulated and decoded in an attempt to reconstruct the original message. The coding/decoding is designed to make the message robust to noise.

[3] This makes the situation somewhat different to the otherwise similar problem of imaging a faint planet next to a bright star. A planet emits or reflects light over a broad wavelength range, so using a narrow filter would reduce the light from the planet too, leading to a net decrease in the SNR.

[4] 'Encoding' refers to converting the message into the code and 'decoding' as converting it back again. Modulation, discussed in section 15.4, describes mapping the code to an electromagnetic signal that can be transmitted.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

maps our text to binary digits. This may map individual letters, as in ASCII or Morse code, or it may map entire words. Coding can also involve compressing a message, for example by removing redundant data, such as areas of an image with the same brightness, or it can involve encrypting the message to prevent it being read by eavesdroppers. Here we are interested specifically in channel coding, the goal of which is to make the message robust, to some level, to errors induced by noise. A channel code, or *error-correcting code*, is one that can both detect the presence of errors and correct them to some degree. This is achieved by introducing redundancy into the signal. We are further concerned here only with forward error-correcting codes, which are codes that enable the receiver to detect and correct errors without having to ask the transmitter to re-transmit parts of the message. This is essential for the multi-year light travel times of interstellar communication.

15.7.1 Example error-correcting code: repetition coding

Consider a signal digitally encoded into a string of L binary characters, 0 and 1. Due to noise in the channel, each character has a probability p that it can be flipped. This occurs independently for each character. Table 15.2 shows an example with $L = 10$ and $p = 0.1$. The top row is the original message; beneath this are various possible received messages. As we would expect with these parameters, most received messages contain at least one error. How can we reliably send a message in such a situation?

A simple but inefficient approach is to resend the message multiple times. For each character the receiver takes a majority vote to decide which value it should have. Suppose we send the message R times ('repetitions'). For any one character in the message, the probability that r of these repetitions are errors is given by the binomial probability distribution,

$$P(r|p,R) = \frac{R!}{r!(R-r)!}\, p^r (1-p)^{R-r}\,. \tag{15.14}$$

Majority voting means that we choose the character that appears in at least $r_{\min} = \lceil R/2 \rceil$ of the repetitions. The ceiling function $\lceil\ \rceil$ gives the smallest integer larger than or equal to its

Table 15.2 Repetition coding. The original message is in the top line. The following lines show examples of randomly generated received messages when the bit flip probability is $p = 0.1$.

1	1	1	0	1	0	1	1	0	1
1	1	1	0	1	0	1	1	0	0
1	1	1	0	1	0	1	1	0	1
1	1	0	0	1	0	1	1	0	0
1	1	1	0	1	0	1	1	0	1
1	0	1	0	0	1	0	1	0	0
1	1	0	0	1	0	1	1	0	1
1	1	1	1	1	0	1	1	0	1
1	1	1	0	1	0	1	0	0	1

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

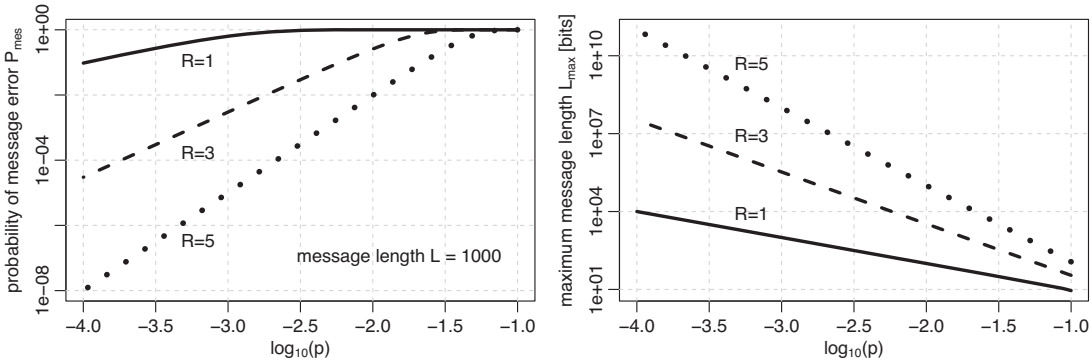


Fig. 15.5 Error suppression by repetition coding. Left: Variation of the probability P_{mes} (equation 15.16) that there is at least one error in a message of length $L = 1000$, as a function of the probability p that a single bit can be flipped, for different numbers of message repetitions R. Right: Variation of the maximum message length L_{max} that can be sent before we expect one bit to be an error.

argument (for example when $R = 5$, $\lceil R/2 \rceil = 3$). To ensure we have no tie-breaks, we select R to be odd.

The probability that our majority vote for this one character is wrong is the probability that r_{min} or more of the repetitions were flipped, which is

$$P_{\text{char}} = \sum_{r_{\text{min}}}^{R} \frac{R!}{r!(R-r)!} p^r (1-p)^{R-r} . \tag{15.15}$$

As the character flipping takes place independently for each character, the probability that a message of length L contains at least one error is

$$P_{\text{mes}} = 1 - (1 - P_{\text{char}})^L . \tag{15.16}$$

The left panel of figure 15.5 shows how this probability varies with the noise level p, for three values of R, for a message length of $L = 1000$. As we would expect, the probability decreases for smaller noise or more repetitions. But note how nonlinear the variation is: with $p = 10^{-3}$, the probability of a message error for $R = 3$ is 3.0×10^{-3}, whereas for $R = 5$ it is only 1.0×10^{-5}. That is, we increase the message length by a factor of $5/3 = 1.7$ but lower the probability of getting an error by a factor of 300. The plot also shows that for a given noise level p, we can decrease the probability of an error in the message to arbitrarily low values by increasing the number of repetitions.

Another way to look at this is to ask what is the expected number of characters in error. This is LP_{char}. Hence the longest message we can send before we expect to get a single error somewhere in the message is $L_{\text{max}} = 1/P_{\text{char}}$. The right panel of figure 15.5 shows how L_{max} varies with p for different values of R. For $p = 10^{-3}$, when sending the message just once ($R = 1$), $L_{\text{max}} = 999$, but if we increase the number of repetitions to $R = 3$ we get $L_{\text{max}} = 3.3 \times 10^5$, and with $R = 5$ we get $L_{\text{max}} = 1.0 \times 10^8$. Thus at the cost of reducing the information rate by a factor of five (for $R = 5$), we can increase the reliable message length by a factor of 10^5.

This simple method of repetition coding is inefficient. Yet it illustrates the fundamental concept of adding redundancy to a message in order to increase the probability of recovering the original message in the presence of noise. Moreover, it shows that it is possible to achieve

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

an arbitrarily low message error probability for a given message length by increasing the total amount of data sent. This is a key result that we will explore formally in the next section.

Given the huge amounts of data sent over noisy channels every day (think of the internet), error-correcting codes have become critical to human society. There are many sophisticated error-correcting codes available, the general goal of which is to minimize the amount of data that has to be sent to recover a message at some specified error rate. Examples include block codes (such as Reed–Solomon), convolution codes, and low-density parity-check codes. They work in different ways, but the general idea is to spread out the redundant signal over time so that the effects of noise can be averaged out.

One particular challenge of interstellar communication is that in addition to noise, which corrupts signals on short timescales, there are inevitably longer periods of receiver outage. The transmitting satellite will not know these have occurred. So the coding scheme needs to spread out signals over a sufficiently long timescale in order to reduce message loss to an acceptable level.

15.7.2 Channel capacity from the Shannon–Hartley theorem

For communication, we are ultimately interested in how many bits of information per second we can receive. In section 15.4 we saw that for continuous waves the maximum data rate, or channel capacity, is given by $C = 2B \log_2 m$ (units $\mathrm{b\,s^{-1}}$), where B is the bandwidth of the channel, and m is the number of distinguishable levels we can achieve with the waves. This no longer holds in the presence of noise, however, because the levels will overlap and become indistinguishable, thus causing the received message to differ from the transmitted message. It can be proven that for an analogue channel with additive white[5] Gaussian noise, the maximum achievable data rate, the channel capacity, is given by

$$C = B \log_2(1 + \mathrm{SNR}), \tag{15.17}$$

where SNR is the signal-to-noise ratio. This is known as the *Shannon–Hartley theorem*. When the bandwidth is 1 MHz and the SNR is 10, the channel capacity is $3.5\,\mathrm{Mb\,s^{-1}}$ With an SNR of 1, $C = B$. By comparing this to the no-noise case of equation 15.6, we see that the SNR can be related to an effective number of distinguishable levels as

$$m = \sqrt{1 + \mathrm{SNR}}\,. \tag{15.18}$$

The Shannon–Hartley theorem is a special case of the more general *noisy channel coding theorem*, which says it is theoretically possible to communicate over a noisy channel up to some maximum rate with an arbitrarily low probability of error. This is quite remarkable: no matter how large the noise, we can, in principle, still pass information reliably at a non-zero rate. How low the error rate will be depends on the actual error-correcting code used to encode and decode the message. If we can define a good enough algorithm, we can achieve an arbitrarily low error rate at the channel capacity. In practice, the actual data rate achievable at some error rate is lower than the channel capacity. However, error-correcting codes have been invented that get quite close to the theoretical limit.

It is possible to transmit at a rate above the channel capacity, but only at a non-zero error rate. In that case, the larger the data rate, the larger the minimum error rate.

[5] White noise is noise that has a uniform spectral power density across the bandwidth.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

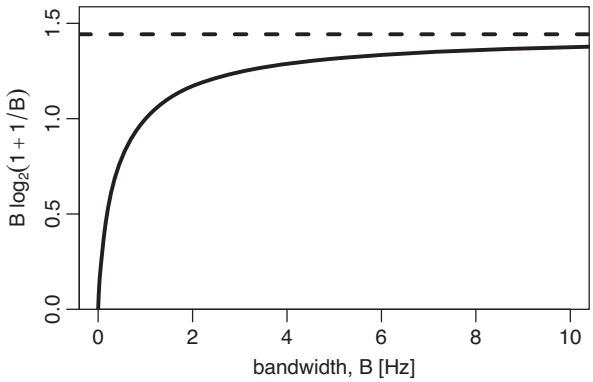


Fig. 15.6 The variation of the channel capacity with bandwidth for a channel with Gaussian white noise and the ratio of the signal power to the noise power per bandwidth of $P/N_B = 1$ (equation 15.19). For all P/N_B, the capacity increases with B and asymptotes for infinite bandwidth to a value of $1.44(P/N_B)$ (dashed line).

The linear dependence of the channel capacity on the bandwidth of the signal in equation 15.17 suggests that to raise the channel capacity, we should set the bandwidth as high as possible. In contrast, given the logarithmic dependence of channel capacity on the SNR, once the SNR is of order 1, increasing it further only increases the channel capacity slowly. Thus to achieve a high channel capacity, it appears we should focus on achieving a high bandwidth. However, a larger signal bandwidth will often include more noise too, because many noise sources have a broad power spectrum. For white Gaussian noise, the noise per unit bandwidth N_B (units $\mathrm{W\,Hz^{-1}}$) is constant, so the total noise is BN_B. If P is the received power in the signal, then the SNR is P/BN_B and the Shannon–Hartley theorem becomes

$$C = B\log_2\left(1 + \frac{P}{BN_B}\right) . \tag{15.19}$$

Figure 15.6 shows that in this situation the channel capacity increases monotonically with increasing B (the curve has a similar shape for all P/N_B). Even though the SNR decreases with increasing bandwidth, which implies we want a smaller bandwidth, the extra carrying capacity of the larger bandwidth more than compensates for this. Thus the larger the bandwidth, the higher the maximum data rate (for given P/N_B). There are diminishing returns for ever larger bandwidth, however: at large B the channel capacity asymptotes to a value of $(1/\ln 2)(P/N_B) = 1.44(P/N_B)$. This is nonetheless relevant to interstellar travel. While constraints onboard the spacecraft may significantly limit how large P can be, and there is a limit to how small we can make N_B, we can compensate for this by using a large bandwidth.

Another way to see this dependence on bandwidth is to consider the energy E required to send a bit of information (units $\mathrm{J\,b^{-1}}$). If R is the actual data rate ($R < C$, in $\mathrm{b\,s^{-1}}$) then $E = P/R$. Writing the bandwidth per unit data rate as $\gamma = B/R$ – a dimensionless quantity with units $\mathrm{b^{-1}}$ – then equation 15.19 can be rewritten as

$$\frac{E}{N_B} > \gamma(2^{1/\gamma} - 1) . \tag{15.20}$$

The quantity E/N_B, with units $\mathrm{b^{-1}}$, can be considered as the energy required to send a bit for a given noise per unit bandwidth. It depends only on γ, which measures how many times larger

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

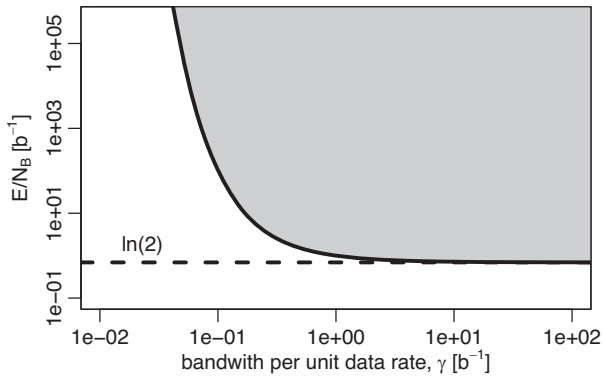


Fig. 15.7 The variation of the minimum energy per bit per noise level, E/N_B from equation 15.20, as a function of the bandwidth per unit data rate, γ. The grey region is that accessible according to the Shannon–Hartley theorem.

the bandwidth is than the data rate. This function is shown in figure 15.7. Consider sending a signal at some fixed data rate R. When we use a relatively narrow bandwidth ($\gamma \ll 1$) this plot shows that we require a large amount of energy for a given noise per unit bandwidth (large E/N_B). This situation is typical of communication on Earth. Many people want to send different messages simultaneously, and so that these don't interfere (overlap), we have to use narrow bandwidths. To achieve this for a given noise level we must use a large amount of energy. To increase the data rate, γ must become smaller, and the figure shows that the energy received must increase. On Earth we can build stationary transmitters fed with large amounts of power, and as the distances over which signals must be sent is relatively small, this is achievable.

On a spacecraft, in contrast, the amount of power available is often limited, so for a given noise per unit bandwidth the term E/N_B is small. The only way we can send a signal – to be in the accessible region in figure 15.7 – is to use a larger bandwidth. Even then we see that there is a minimum energy per bit (for each unit power of noise per unit bandwidth) of $\ln(2) = 0.693\,\mathrm{b}^{-1}$ required to receive any signal, even with an arbitrarily large bandwidth.

Hence communication in the terrestrial and in the interstellar domains in the presence of noise are quite different: terrestrial is low bandwidth, high power; interstellar is high bandwidth, low power. In theory we can use arbitrarily small bandwidths if we have enough power. In contrast, even with an arbitrarily large bandwidth we still need a minimum power.

The above discussion assumes that both signal and noise occupy the full bandwidth B. In practice, the signal is bandwidth-limited. A laser, for example, might have a bandwidth of 1 nm. There is then no point in using a receiver with a broader bandwidth, as this simply adds noise but no signal. So in practice we would use a filter to limit the bandwidth at the receiver, as we did in section 15.6, but then spread the transmitted signal over the entire bandwidth to maximize the energy efficiency.

To now compute the channel capacity we would need to compute the SNR at the receiver. This involves specific details of the modulation method, which we won't delve into here. We would then need to specify an error-correcting code in order to determine the error rate and the actual data rate, which may be well below the theoretical limit. We will look at some examples in section 15.9.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

15.7.3 Channel capacity at low photon rates

When the received photon rate is very low, we can no longer consider communication being done with continuous waves like those shown in the left of Figure 15.1, but must instead consider the information carried by individual photons. In this limit, the Shannon–Hartley theorem of section 15.7.2 no longer applies. The upper limit on the channel capacity is instead given by *Holevo's bound*. This is an upper bound to the amount of classical information in a quantum state that is accessible to a receiver.[6]

An important concept in this approach is the amount of information that a single photon can carry, or more specifically, the number of modes a photon can take. A mode can be the photon's frequency, polarization, or time of arrival. In the context of pulse position modulation, discussed in section 15.4, the time of arrival is the position of a pulse in the frame. Each slot corresponds to one mode. If we can count individual photons and place them in one of seven slots (enough to encode ASCII for example), there are $m = 7$ modes. We can also combine types of modes. If we additionally had two polarizations (e.g. left and right circular) and 10 different frequencies, we would have $7 \times 2 \times 10 = 140$ modes per photon. The inverse of this number we write as M, the number of photons per mode.

Without diving into the rather complicated theory, let's just take a brief look at some of the concepts involved. In the absence of noise, the capacity of the channel in bits per mode is, from the Holevo bound,

$$C_m = g(\eta M) \quad \text{(noise-free)} \tag{15.21}$$

where η is the efficiency of the detector (fraction of photons detected), and g is the function

$$g(x) = (1+x)\log_2(1+x) - x\log_2 x\,. \tag{15.22}$$

When there are N noise photons per mode on average, the number of bits per mode is instead

$$C_m = g(\eta M + [1-\eta]N) - g([1-\eta]N) \quad \text{(noisy)}. \tag{15.23}$$

This expression is for a specific type of noise, known as the thermal channel. The channel capacity in bits per photon, in both the noise-free and noisy cases, is

$$C_p = \frac{C_m}{\eta M}, \tag{15.24}$$

the factor η being here because this is the number of bits per *detected* photon.

Figure 15.8 shows how the channel capacity C_p varies with the number of photons per mode M. The dashed line is for the noise-free case. As we would expect, the channel capacity increases as M decreases, because there are more information-carrying modes for each photon.

When there is noise, we see from figure 15.8 that the higher the noise per mode N, that is, the lower the SNR in a mode, the lower the channel capacity C_p for any M. In the limit of sufficiently large M, the capacity of a noisy channel tends toward that of a noise-free channel (the lines converging on the right of the figure). This is intuitive, because with a fixed number of noise photons per mode, the more signal photons there are, the closer this is to the noise-free case. As with the noise-free case, increasing the modes per photon (decreasing M) at

[6] Although this considers quantum effects, it is not the same as quantum communication, which is a different topic that will not be considered here.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

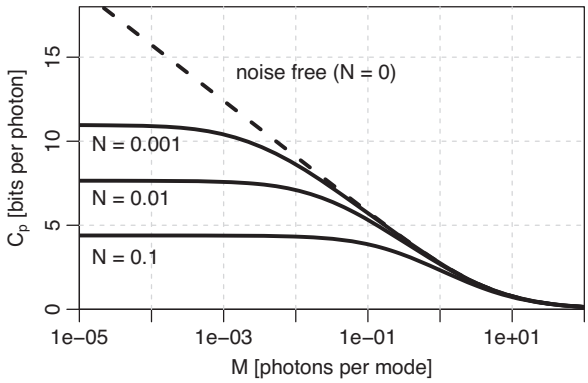


Fig. 15.8

The variation of the channel capacity in bits per photon (equation 15.24), as a function of the number of photons per mode in a photon-counting communication channel, for a detection efficiency $\eta = 0.5$. The dashed line shows the noise-free relation. The solid lines show three examples of noisy channels, where N is the average number of noise photons per mode. As the variation of C_m with M is fairly linear for most of the range shown, the plots of C_p and C_m with M look qualitatively similar.

constant N increases the channel capacity. That is, even though the SNR in a given mode is decreasing to the left in the figure (at fixed N), the corresponding increase in the number of modes for each photon to carry the information more than offsets this. The result is an overall increase in the channel capacity. This has a limit, however, as we see that for a given N, the channel capacity stops increasing once M drops below some level. For example, if we have $M = 0.01$ signal photons in each mode, yet $N = 0.1$ noise photons in each mode, then at $\eta = 0.5$ we get $C_p = 4.3$. Decreasing M further – packing more modes into each photon – barely increases the channel capacity.

The dependency of the channel capacity on the noise is more intuitive when we plot these quantities for a given value of M. This is shown in figure 15.9 for several values of M, again at $\eta = 0.5$. As the noise per node decreases, the SNR for a given M (and thus per mode) increases, and the channel capacity increases. Again, this increases continues only up to some asymptotic limit.

Note that the channel capacity in bits per mode, equation 15.23, depends on η, the detector efficiency. We find that as this efficiency increases to 1.0, all of the noisy lines in figure 15.8 converge on the noise-free case (see exercises). The variation of C_p with η is counter-intuitive, however, because C_p is the channel capacity per *detected* photon. For the noise-free case, C_p actually increases with decreasing η (at given M). But the rate at which photons are detected has decreased, so the bit rate has also increased. If we plot the noise-free channel capacity per *incident* photon (prior to inefficient detection), C_m/M, then this is indeed lower for lower efficiency.

The channel capacity in bits per second is

$$C = \dot{n}_s \eta C_p, \tag{15.25}$$

where $\dot{n}_s$ is the incident signal photon rate (prior to detection), and thus $\dot{n}_s\eta$ is the detected signal photon rate. Note that ηC_p is the number of bits per incident photon, so the channel

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

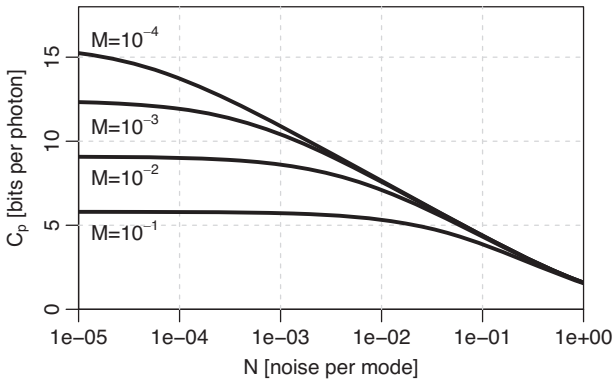


Fig. 15.9 The variation of the channel capacity in bits per photon (equation 15.24), as a function of the number of noise photons per mode in a photon-counting communication channel, for four different values of M (photons per mode), with $\eta = 0.5$.

capacity can also be seen as the product of the number of bits per incident photon and the incident photon rate.

Comparing equation 15.23 to the classical expression in equation 15.17, we see that in the former, the bandwidth and SNR have been replaced by the number of photons per mode and the noise per mode (as well as an efficiency factor). Photons per mode can be thought of as a generalization of the concept of bandwidth.

It is important to appreciate that the ideas sketched in this section give a fundamental theoretical limit for the maximum data rate. Achieving this will be difficult in practice. The value of N, the number of noise photons per mode, as well as the actual noise model, depends on the physical implementation, which will need low-noise single photon-counting detectors. We also need an efficient error-correcting code (section 15.7.1) that can approach the theoretical limit. Studies that assume this method can be used for interstellar communication often compute the incident photon rate $\dot{n}_s$ using the Friis transmission equation, but then adopt a value for ηC_p, the number of bits per incident photon, without specifying the implementation. From these two numbers we can compute an expected data rate using equation 15.25.

To get some idea of what the limit is in the ideal, noise-free case with $\eta = 1$, consider the case of communicating from Proxima Cen in the bottom row of table 15.1, which has $\dot{n}_s = 190$. Adopting $M = 10^{-2}$, equations 15.21 and 15.24 give $C_m = 0.081$ bits per mode and $C_p = 8.1$ bits per detected photon. This gives a theoretical maximum data rate of $C = 1540\,\mathrm{b\,s^{-1}}$.

Theoretical studies aside, there have been practical implementations of communication channels using individual photons. Farr et al. (2013) demonstrate an optical communication channel using PPM that achieves $C_p = 10$–15 bits per photon. In one example, using $m = 2^{19}$ slots (19 bits per slot, which ideally is $M = 2^{-19} = 1.6 \times 10^{-6}$ modes per photon), they achieved 15.4 bits per detected photon, or 13.2 bits per incident photon ($\eta = 0.86$). This compares to a noise-free theoretical limit of 20.7 bits per detected photon, as computed from equations 15.21 and 15.24. The data rate was $3.2\,\mathrm{kb\,s^{-1}}$. Although the goal of Farr et al. (2013) was to demonstrate a large number of bits per photon rather than per second, they could achieve a higher data rate of $142\,\mathrm{kb\,s^{-1}}$ at the only slightly lower value of 10.4 bits per detected photon.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

The experiments of Farr et al. (2013) were done in the lab, not in space, let alone at interstellar distances. However, this approach has been employed in space communication. Villoresi et al. (2008) demonstrated transmission and receipt of single photons between the Earth's surface and a satellite 1500 km away, a distance over which the signal was attenuated by nearly 16 orders of magnitude. Interplanetary-scale optical communication has also been done with large intensities. The DSOC experiment on the Psyche spacecraft has demonstrated optical communications with the Earth from a distance of up to 3 au. This uses a 4 W laser operating at 1.55 μm with a 22 cm transmitting telescope together with a 5 m receiving telescope on the Earth, to achieve data rates of up to 270 Mb s^{-1}. When at 2.7 au, it measured a data rate of 8.3 Mb s^{-1}.

15.8 Choice of wavelength and location of receiver

Earlier in this chapter we saw how the wavelength of a signal influences its diffraction and bandwidth and thus the maximum data rate achievable. Another important factor affecting both data rate and choice of wavelength is where we place the receiver. It could be located on the surface of the Earth, in space, or on the surface of another solar system body such as the Moon. The interdependent choice of receiver location and communication wavelength involves several factors, including: the maturity of technology for building transmitters and receivers; sources of absorption and emission, and thus noise; and the cost of installing and maintaining the receivers. We must also remember that the received wavelength is not the same as the transmitted wavelength due to the Doppler effect. To receive signals from the spacecraft when it is travelling at different velocities, we need to be able to tune to the received wavelength (and variable bandwidth) over the course of the mission.

In this section we look at the impact of the Earth's atmosphere on communication, the merits of putting receivers in space, and some technological issues concerning the choice of wavelength.

15.8.1 Absorption and emission from the Earth's atmosphere

When receivers are placed on the Earth, the spacecraft's signals must pass through the atmosphere. This generates two problems. The first is that the atmosphere is opaque to electromagnetic radiation at certain wavelengths, as seen in figure 15.10. Blueward of about 0.3 μm the atmosphere is almost entirely opaque due to absorption by oxygen and ozone (including the ozone layer that protects life from solar ultraviolet radiation). Between 0.3 μm and 1.0 μm there is little absorption. It is of course no coincidence that animals can see and that plants photosynthesize using light in this region. In the range 1–8 μm there is again a lot of absorption, mostly from water and carbon dioxide, although there are a few gaps (exploited by astronomers to make infrared observations from the ground). From 8 μm to 14 μm there is a low-absorption window (at least in low-humidity regions), apart from an absorption band of ozone in the middle. Redward of 14 μm out to radio wavelengths of a few millimetres, the atmosphere is opaque due initially to carbon dioxide and then to water. Astronomy at these wavelengths must be done from space. The opacity at the long wavelength end of this region is dependent on the amount of water vapour, which is why (sub)millimetre telescopes tend to

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

be placed on very high altitude dry sites, like the Atacama Large Millimeter/submillimeter Array (ALMA) on the Chajnantor plateau in northern Chile at an altitude of 5000 m. Between wavelengths of a few millimetres and a few tens of metres, the atmosphere remains largely transparent. At longer wavelengths the atmosphere becomes opaque again due to scattering by free electrons in the ionosphere.

Which of these atmospheric windows could we use for interstellar communication from the ground? Most communication with satellites is done in the range 0.7–7 cm (4–40 GHz). Shorter wavelengths down to 3 mm (frequencies up to 100 GHz) can be used, but absorption by oxygen and water in the troposphere become increasingly problematic. Shorter wavelengths are preferable on account of the lower diffraction at the transmitting spacecraft (section 15.2) and the larger bandwidths obtainable. This leaves either the infrared windows between 1 μm and 14 μm or the visible window at 0.3–1 μm.

This brings us to the second problem with the Earth's atmosphere: emission, and more specifically thermal emission. Black bodies at room temperature have their peak spectral radiance at about 10 μm. This can be seen from Wien's displacement law parametrized by wavelength, which is $\lambda_{\max}\,[\mu\mathrm{m}] = 2898/T$. This means that telescopes, instruments, and detectors may all contribute significant background noise. Detectors and parts of instruments can be cooled, and this is routinely done in astronomy when observing at these wavelengths. But neither the telescope nor the sky can be cooled, so the background remains significant.

The black body distribution is broad, so to avoid thermal emission we should use signals at much shorter wavelengths than the thermal peak. This points to the visible window. The problem here is that the sky is much brighter during the day than the night, which either

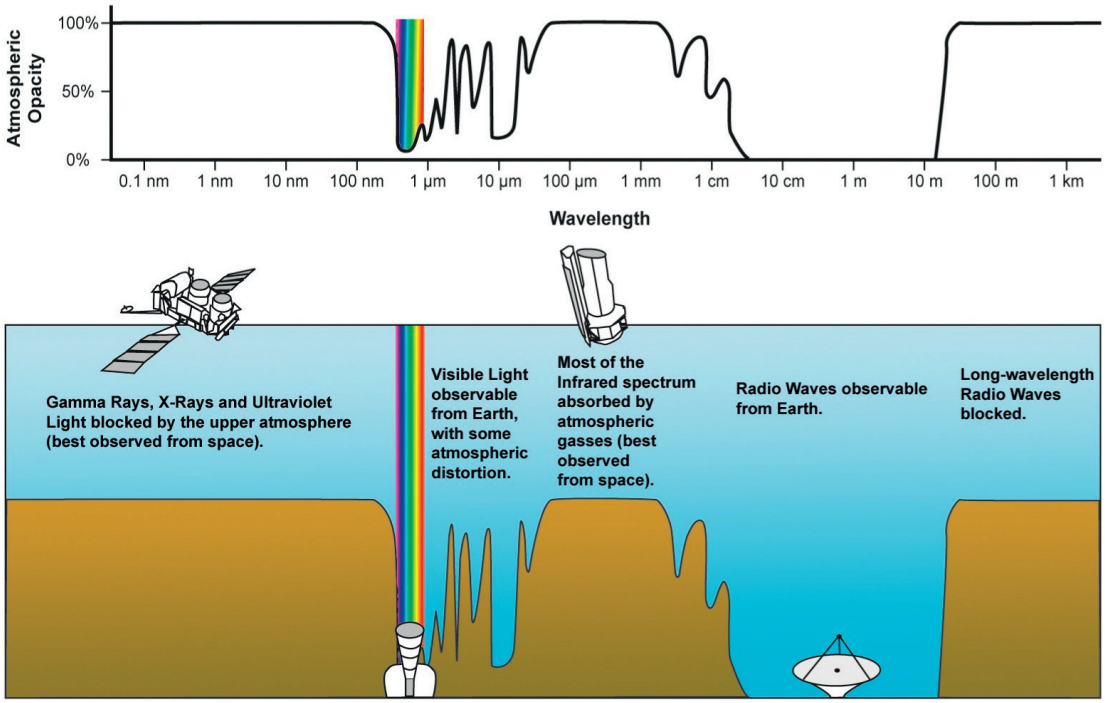


Fig. 15.10 Absorption by the Earth's atmosphere as a function of wavelength. Credit: NASA, public domain, from Wikimedia Commons.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

significantly lowers the SNR or inhibits detection entirely, thus reducing the time available for reception by a single receiver. This is not such a limitation for much of the radio spectrum.

Even at night there are additional emission sources that can contaminate an optical or infrared signal. The most obvious is the Moon. Just as it does for the Sun, the Earth's atmosphere scatters light from the Moon, so the night sky brightness depends on the angular distance of the Moon from the direction we observe. The Moon is bright enough that observations of faint astronomical sources are only done around the time of the new Moon. Yet even on moonless nights there is emission from the Earth's atmosphere itself, known collectively as airglow. One source of airglow is the recombination of ions in the upper atmosphere that were ionized by sunlight during the day. The decay of molecules excited into metastable states by sunlight also emits light. One metastable state of molecular oxygen has a radiative lifetime of about 75 minutes, for example, and so this emission is strongest at dusk. Another source of airglow is photochemical luminescence (chemiluminescence), the process whereby light is released following a chemical reaction that has left a molecule in an excited state. One example of this is the combination of hydrogen and ozone that produces oxygen and excited hydroxyl radicals (OH) near the mesopause at an altitude of 85 km. Another is the reaction of sodium and ozone that results in sodium D-line emission. Contributions of different natural emission sources seen from the surface of the Earth are shown in figure 15.11.

15.8.2 Receivers in space

In order to receive signals from our spacecraft, the receiver must have a direct line of sight to it. Depending on the declination of the target star, the spacecraft may not be visible for all of the day/night on account of the Earth's rotation. Alpha Cen has a high declination, just 27° from the south pole, so is continuously visible from a site sufficiently far south. But daytime reception is anyway not possible for signals at visible/infrared wavelengths, so multiple ground stations at different longitudes would be necessary to ensure continuous contact.

One way around these problems is to place the receiver in space or on the surface of the Moon. This is expensive due to the extra costs of launching and maintaining (or replacing) structures in space. If the receiver is in Earth's orbit, the orbit has to be designed to avoid the target star being eclipsed by the Earth or Sun.

Observing from space also avoids the problems with the Earth's atmosphere outlined in the previous section, but it does not remove all sources of absorption and emission. If a spacecraft travelled to a star-forming region or far into the Galactic plane, optical signals in particular would be attenuated by absorption from gas and dust in the ISM. However, as we saw in section 12.3, the Sun is located in a low-density region of the ISM, so this absorption is barely relevant for nearby stars.

Electrons in the ISM significantly disperse electromagnetic waves with frequencies below several kilohertz, so we could not use these very low frequencies. But as such low frequencies suffer from significant diffraction and only offer limited bandwidth, they are not really of interest for communication.

Celestial emission might be a problem no matter where the receiver is placed. Unresolved stars generate a diffuse background, which is strongest at low Galactic latitudes. Zodiacal light, the scattering of sunlight by dust in the solar system, is strongest in the ecliptic plane, and is easily visible with the naked eye from the Earth's surface on a dark night. These two

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

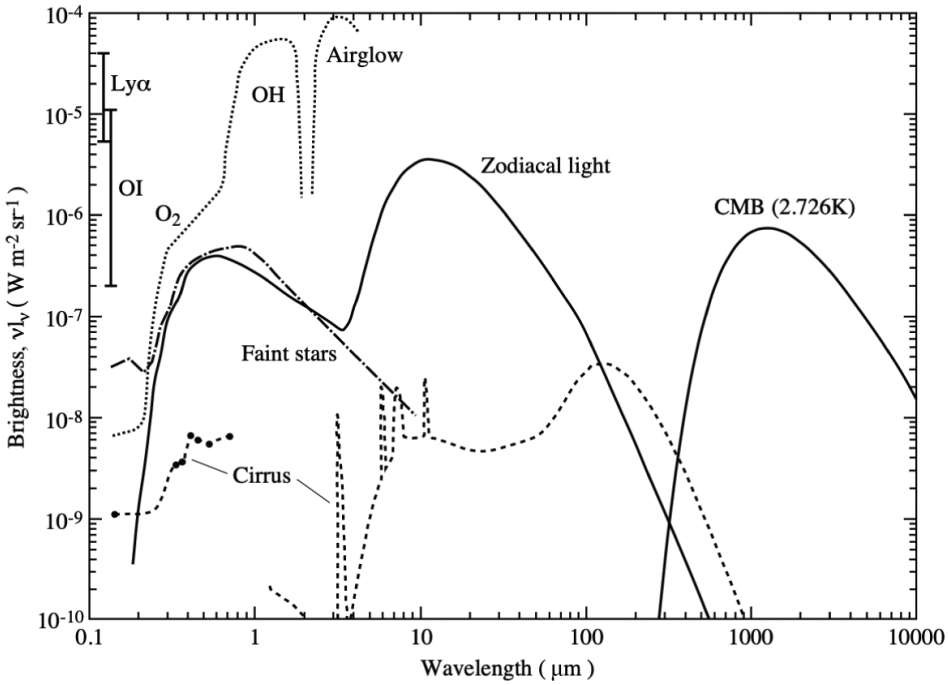


Fig. 15.11 Sources of sky emission from a high-altitude site on the Earth's surface looking away from the Galactic and ecliptic planes. 'Cirrus' refers to emission from interstellar dust, not Earth clouds. OI is neutral oxygen. Credit: Reproduced with permission of EDP, from Leinert et al. (1998).

sources mostly affect the optical and near-infrared. At longer wavelengths there is emission from warm interstellar dust particles. At the longest wavelengths we see emission from the cosmic microwave background (CMB) with a peak flux density at around 1 mm. The relative contributions of these different sources of emission are shown in figure 15.11.

A receiver placed on the surface of the Moon, like one in space, would also benefit from the lack of an atmosphere. It may also be easier to construct or maintain than a receiver in an Earth or solar orbit. There may still be a problem of maintaining a line of sight to the target, however, as the Moon rotates once on its axis every 27 days, so multiple receiving locations may again be required.

15.8.3 Technological issues affecting the choice of wavelength

As already discussed in section 15.2, minimizing the diffraction angle of the transmitted signal is a major factor in maximizing the signal intensity received at the Earth. The diffraction angle is of order λ/d_t. Whereas the ratio of radio to optical wavelengths is of the order of 1000, the ratio of radio to optical telescope diameters that could plausibly be installed on interstellar spacecraft is likely to be much lower: an optical telescope diameter of 1 m is plausible, a radio telescope diameter of 1 km less so. This suggests that, all other things being equal, the limited mass and size of an interstellar spacecraft favours the use of optical rather than radio wavelengths for communication.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

In order to achieve a diffraction-limited beam, the transmitting telescope's surface needs to have the correct curvature to within a fraction of a wavelength. In the near-infrared this means of the order of 0.1 μm. This has usually been achieved by casting and polishing glass mirrors and then covering them with a thin coating of reflective aluminium. At radio wavelengths the surface accuracy requirements are significantly more relaxed, more like 1 mm. As high-energy particles will damage the leading surface of a spacecraft (see section 12.4), we may want to deploy the telescope only near the destination (although it would be pointing against the direction of travel). Deploying structures to achieve a very specific and exact shape is challenging.

One idea suited for laser sailcraft in particular is to repurpose the sail as a transmitting telescope. This would save mass. However, as we saw in section 11.4.1, the optimal material for a laser-powered sail may not be highly reflective (e.g. graphene), making it unsuitable as a telescope. The sail may also not have the right shape, as the optimal shape from a stability perspective is probably not ideal for a telescope (see section 11.4.2). The sail could be reshaped, and perhaps even recoated, in situ, but both of these add considerable complexity and risk to the mission.

The properties of the transmitter are only part of the equation, literally in the Friis transmission equation (section 15.2). To increase the data rate we can also increase the total area of the receiver. There are practical limits to the size of individual telescopes. Currently, the largest steerable optical and radio telescopes have diameters of order 40 m and 100 m, respectively. Larger areas need much more massive structures to support them on the Earth or Moon (area $\sim d^2$, but supporting mass $\sim d^3$). So to achieve a larger area, we would probably use arrays of small telescopes.

So that the receiver can spatially separate the spacecraft signal from the bright target star, the receiver would probably be an imaging system (see section 15.6). To achieve the required angular separation, each telescope in the receiver array needs a minimum diameter due to diffraction, and must be equipped with a coronagraph. If situated on the Earth's surface, each telescope must also be fitted with adaptive optics (see section 11.2.3). If the spacecraft and star are physically separated by 0.1 au, then to separate them in the focal plane by at least the diffraction angle at 1 μm, we need individual telescopes at least 3 m in diameter (and in practice larger, as it's better to separate the star and spacecraft by more than one diffraction angle). In that case the individual telescopes can be combined incoherently, meaning each just acts as a separate light bucket. If the individual telescopes are smaller, then to achieve the required angular separation, we need to combine at least subsets of the telescopes coherently to synthesize a larger aperture (see section 11.2.2).

The idea of coherent light combination could be used to produce the transmitted beam on the spacecraft too. That is, instead of using a single large mirror, we could retain just parts of the telescope's surface (see figure 11.11). This changes the diffraction pattern at the receiver, in particular by putting more power into the side lobes (see figure 11.12), but it would make it easier to synthesize a large transmitter on the spacecraft.

Transmitting a coherent signal using a synthesized aperture requires that the individual mirror segments be positioned precisely (sub-wavelength) on the equivalent aperture. If this cannot be achieved, then instead of mirror segments at the correct locations, we can instead use a set of small individual transmitters at arbitrary locations. Each transmits the same signal, but at different phases that account for them not being in the correct locations. That is, we emulate a single telescope (with large gaps) by introducing the appropriate phase shifts

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

into individual small transmitters. This is known as a phased array antenna, which we also mentioned in the context of laser sails in section 11.3.2. This gives us the freedom to place the individual transmitters essentially anywhere, which gives more flexibility onboard a spacecraft. The direction in which the signal is sent is then determined by the phase offsets. This is also referred to as an electronically steerable antenna. Such transmitters are already used on astronomical satellites where the physical movement of an antenna would be too disruptive, for example on Gaia.

Current electronics technology only allows the measurement of phases for frequencies below about 1 THz, that is, wavelengths above 300 μm, so phased array antennae are mostly used at radio wavelengths. However, optical phased arrays have also been made, in which optical path differences are introduced by changing the refractive index of materials electronically. This avoids the need for cumbersome delay lines that are otherwise used in optical aperture synthesis (e.g. figure 11.11).

In summary, there are many factors that influence the optimal wavelength for communication. The choice for a specific mission will be a trade-off between all factors, and will be strongly influenced by the spacecraft mass budget, the required data rate, risk, and cost. However, it seems likely that interstellar spacecraft will use visible or near-infrared wavelengths, and a receiver that is located in orbit or on the Moon that has continuous contact with the spacecraft.

15.9 Case studies

We now look at some specific concepts for sending signals back from interstellar spacecraft. To put this in some context, the New Horizons mission to Pluto (and later Arrokoth) transmitted its data from over 30 au back to Earth at radio frequencies (X-band, 3.6 cm) using a 2.1 m diameter antenna. The signal was received by a 70 m antenna. The data rate was around 1–2 kb s^{-1}, and it took around 15 months to transmit the 6.25 GB dataset obtained at Pluto.

Milne et al. (2016) outline a very large, high-power antenna to achieve a very high data rate from Alpha Cen as part of the Icarus project (see section 8.6). The radio transmitter is a parabolic reflector of 1 km diameter and 40 t mass that is fabricated and deployed after arrival. It transmits 1 MW of power at 32 GHz (1 cm) with a 10% (3 GHz) bandwidth. With this they aim for a data rate of 10 Gb s^{-1}. For comparison, high-definition television is 1.5 Gb s^{-1}, which is already much higher than is achieved from current space telescopes. Using equation 15.17, the data rate and bandwidth imply an SNR in the received signal of 9.1. Heat dissipation is an issue at these high powers, especially at high frequencies because the transmitting electronics are smaller. To achieve this high data rate, a 15 km diameter receiver is required which, given its size, is placed in space at the Earth/Sun L4 or L5 Lagrange point.

Milne et al. (2016) also consider an optical downlink with a 0.532 μm laser with an average power of 1 kW transmitted by a 10 m diameter mirror. A 200 m diameter telescope receives the signal on the Earth. Using the Friis transmission equation (equation 15.3) we can compute that this would deliver an average power at the receiver of 5×10^{-12} W, or about 14 million photons per second. They actually report values a factor of 10 lower, because they more realistically take into account optical losses in the transmitting and receiving telescopes, absorption by the Earth's atmosphere, and imperfect detectors. The downlink has a

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

very narrow bandwidth of 3 GHz, or 0.003 nm. They use a PPM modulation scheme with 1024 slots in the frame, which with optimal encoding corresponds to 10 bits per frame (see section 15.4). On average they expect this optical link to achieve a data rate of 1 kb s^{-1}.

Parkin (2020) describes an optical communication channel for a laser sailcraft at Alpha Cen as part of the Breakthrough Starshot initiative. It uses a laser of 100 W power operating at 1.02 μm with a bandwidth of 0.1 nm. This is transmitted via the sail, which has been reconfigured as a reflecting telescope of 4.1 m diameter. The receiver is a 30 m diameter telescope. The spacecraft is moving at a velocity of $0.2\,c$ away from the Earth, so the received signal is redshifted to 1.25 μm. The 'relativistically moving' Friis transmission equation (equation 15.5) predicts that this would deliver photons at a rate of 1210 s^{-1} at the receiver. This assumes perfect efficiency of the optics and electronics and no signal losses. Overall the communication system is stated to deliver 288 signal photons per second at the receiver, indicating a system efficiency of about 25%.

The study of Parkin (2020) uses a telescope to collect and focus the optical signal from the spacecraft. As we saw in section 15.6, there can be considerable contamination from the light of the target star, which here is Alpha Cen, and this is taken to be the dominant source of noise. Adopting a PPM modulation of the signal, the study concludes that the received 288 signal photons per second can carry data at a rate of 260 b s^{-1} with an error rate of 10^{-3}. This is quite a large error rate but the expectation is that it could be reduced to 10^{-6} with a better coding scheme.

A larger data rate and/or lower error rate is possible if we have larger optics. Lubin (2016) describes a system with a 30 m diameter telescope transmitting a 10 W signal at 1.06 μm from Alpha Cen. Employing an enormous array of telescopes on or near the Earth, with an equivalent single-telescope diameter of 10 km, signal photons are received at a rate of 1.6×10^{10} s^{-1} (from equation 15.1). The author then assumes that each photon can carry 0.025 bits on average, a value that reflects some SNR (channel capacity calculations are not provided). Combined with the quoted photon rate, this delivers a data rate of 390 Mb s^{-1}. This is enough to stream a 640 × 480 video at 30 frames per second, which requires 221 Mb s^{-1} when adopting 24 bits per pixel and no compression or channel coding.

Hippke (2019) considers an optical system operating at visual wavelengths (0.43 μm; bandwidth 1 nm) that uses a relatively large power (1000 W) but a small transmitter (1 m diameter). The receiver on the Earth's surface has a diameter of 39 m, the diameter of the ELT in Chile. The photon rate from the spacecraft at Alpha Cen on the receiver (prior to detection) is given to be 1370 s^{-1}. This compares to 6500 s^{-1} from equation 15.3, which is prior to accounting for losses of any relativistic velocity. They assume the light from Alpha Cen can be perfectly suppressed so contributes no noise. They instead assume that the dominant noise source is emission from the Earth's atmosphere, estimated to be 0.6 photons s^{-1} nm^{-1} m^{-2} $arcsec^{-1}$. Taking the collecting area of the telescope as 1170 m^2, this corresponds to 702 sky photons per second. They then use the approach from section 15.7.3 to compute the data rate. Specifically, they adopt a very low value of the number of photons per mode of $M = 10^{-5}$ and give the number of bits per incident photon to be 1.81, which for the stated detector efficiency of $\eta = 0.5$ corresponds to $C_p = 1.81/0.5 = 3.62$ bits per detected photon. Using equation 15.23 together with equation 15.24, we see that this has assumed about $N = 0.18$ noise photons per mode. As the incident signal photon rate is 1370 s^{-1}, we get a data rate of $1.81 \times 1370 = 2480$ b s^{-1}.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

15.10 Summary

Getting data back to the Earth from a distant spacecraft is a major challenge. As the transmitter will inevitably be limited in size, diffraction of the transmitted beam results in an extremely low intensity at the Earth. Limitations of the transmitter can be compensated to some degree by using a large area receiver, as shown by the Friis transmission equation (equation 15.3). Shorter wavelengths reduce diffraction, and for this reason in particular we probably want to use optical rather than radio communications.

To give some example numbers, a 10 W transmission with a 1 μm laser from a 4 m diameter telescope at Proxima Cen delivers an intensity at the Earth of 6×10^{-18} W m^{-2} at the centre of an Airy diffraction pattern, the central disc of which is 24 million kilometres in diameter. To maximally block out the far brighter target star, which typically has a small angular separation from the spacecraft, the receiver should be an imaging telescope equipped with a coronagraph and narrow band filter. This telescope needs a minimum diameter of a few metres to separate the signal and star on its focal plane, as well as adaptive optics correction if the telescope is observing through the Earth's atmosphere. With a single 10 m receiving telescope, the above signal intensity corresponds to just 31 photons per second, less when considering losses in the telescope and detector. With an array of 100 such telescopes, combined incoherently, this increases to 3100 photons per second.

The amount of information individual photons can carry depends on how a message is encoded and modulated. At low photon counts, pulse position modulation (PPM) is convenient. By modulating in time, an individual photon can convey more than one bit of information. Up to 15 bits has been demonstrated experimentally with PPM.

The received data rate depends also on the noise, of which there are many sources, including the target star, absorption and emission in the Earth's atmosphere, and detector electronics. Noise introduces errors in the messages. To overcome this, redundancy must be built into the signal by means of an error-correcting code. Remarkably, even in the presence of large noise, it is theoretically possible to communicate at non-zero data rates with an arbitrarily low probability of message error. Codes have been developed that in some situations can almost reach the theoretical data rate limit, or channel capacity.

Assuming we can extract 10 bits per photon from the signal, then 3100 received photons per second gives a data rate of 31 kbs^{-1}, or nearly 1000 Gb per year. This corresponds to about 30 000 uncompressed 2048 × 2048 images of 8 bits per pixel. Compression would reduce the raw data size, but the error-correcting code required to overcome noise would increase it. Losses and inefficiencies would also lower the final data rate, perhaps by several factors.

A spacecraft would probably transmit its preciously won data over a period of several years (whether a flyby or a rendezvous). As the transmitter and receiver cannot coordinate, the spacecraft cannot retransmit data in the case of inevitable signal loss. To cope with this, the coding algorithm needs to provide some long timescale redundancy too.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016

15.11 Exercises

1. I outlined in section 15.2.1 how the Friis transmission equation needs to be modified for relative motion between transmitter and receiver. Verify the resulting equation by re-deriving it from first principles.
2. Compute the power and photon rate on the Earth-based receiving telescope for the DSOC experiment mentioned at the end of section 15.7.3, assuming an end-to-end efficiency of 0.1. How many bits per photon does this correspond to?
3. Show that for the Holevo bound, the channel capacity in bits per mode from equation 15.23 tends towards the noise-free case of equation 15.21 as the detector efficiency $\eta \to 1$.
4. List the relative merits of doing interstellar communication in the optical and in the radio. Under what circumstances might one be clearly preferred over the other?

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 15 Nov 2026 at 02:45:23, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009700429.016