30. Communicating Information

(a) Candidates should be able to understand the term modulation and be able to distinguish between amplitude modulation (AM) and frequency modulation (FM).

All communication systems require a source and a receiver. Three such systems are illustrated in Fig. 3.1.

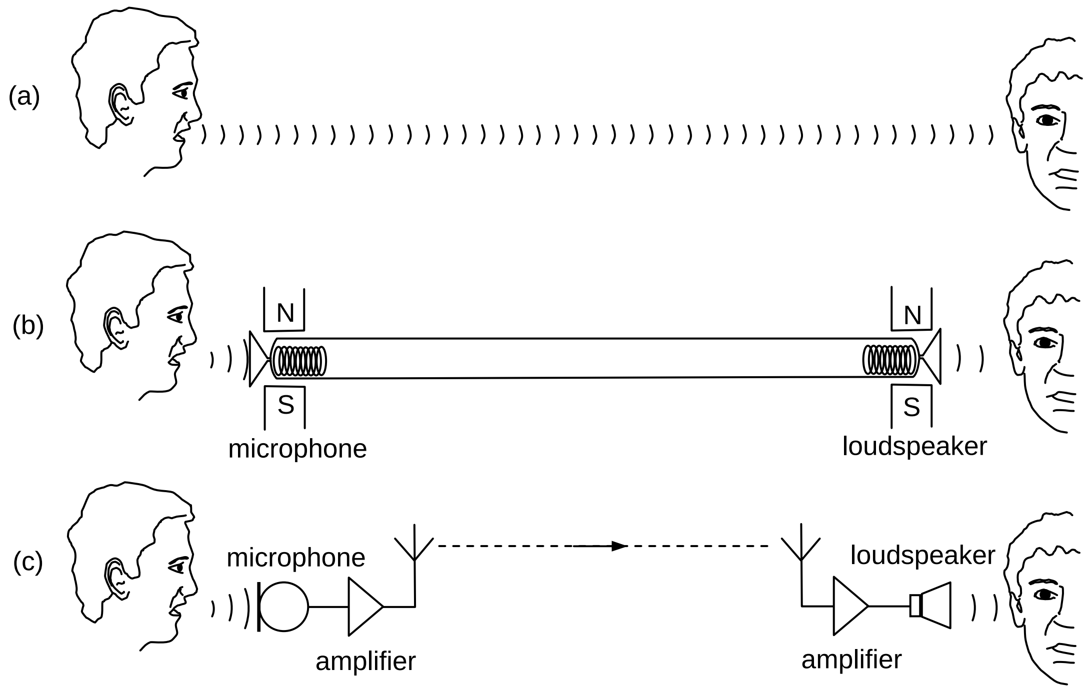

Fig. 3.1

Sound can be transmitted either directly as in (a) or via the alternating currents induced in a moving-coil microphone and received by a moving coil speaker as illustrated in (b).

It is also possible to communicate using radio waves by simply amplifying the audio signal and applying it to a suitable aerial as illustrated in (c). However, there are two fundamental problems with this system.

1. Only one radio station can operate in the region because the wave from a second operating station would interfere with the first.
2. The aerial required to transmit frequencies in the audio range (20 Hz to 20 kHz) would be both very long and inefficient (the radio waves would not travel very far unless huge powers were used).

Both of these problems are solved by the process of *modulation*, the principle of which is illustrated in Fig. 3.2. In modulation, a high frequency wave known as the *carrier wave* has either its amplitude or its frequency altered by the information signal in order to carry the information.

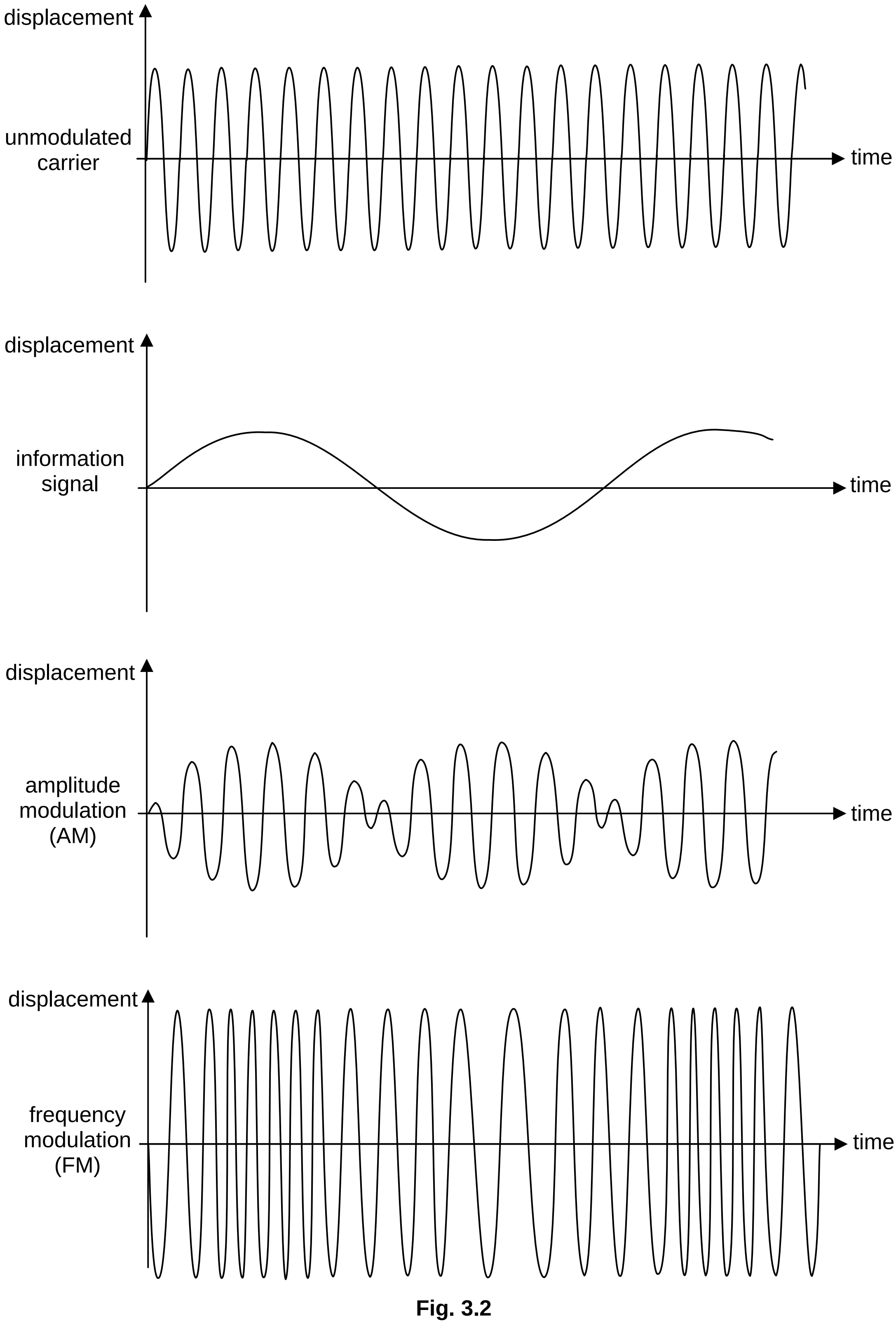
displacement
unmodulated carrier
time
displacement
information signal
time
displacement
amplitude modulation (AM)
time
displacement
frequency modulation (FM)
time

Fig. 3.2

For *amplitude modulation* (AM), the amplitude of the carrier wave is made to vary in synchrony with the displacement of the information signal. The variation in the amplitude of the carrier wave is a measure of the displacement of the information signal and the rate at which the carrier amplitude varies is equal to the frequency of the information signal.

For *frequency modulation* (FM), the frequency of the carrier wave is made to vary in synchrony with the displacement of the information signal. The amplitude of the carrier wave does not vary. The change in frequency of the carrier wave is a measure of the displacement of the information signal. The rate at which the carrier wave frequency is made to vary is equal to the (instantaneous) frequency of the information signal.

Note: The use of a carrier wave allows different radio stations in the same locality to transmit simultaneously. Each station transmits on a different carrier frequency and consequently the carrier waves do not, in effect, interfere with one another. This is because any one receiver is tuned to the frequency of a particular carrier wave. The receiver then responds to, and gives an output based on, the differences in displacement, or frequency, between the actual waveform and the 'underlying' carrier wave. In other words, the receiver recognises the information signal and rejects others.

(b) Candidates should be able to recall that a carrier wave, amplitude modulated by a single audio frequency, is equivalent to the carrier wave frequency together with two sideband frequencies.

(c) Candidates should be able to understand the term bandwidth.

Fig. 3.3 shows the waveform resulting from the amplitude modulation of a high frequency carrier wave by a signal that consists of a single audio frequency.

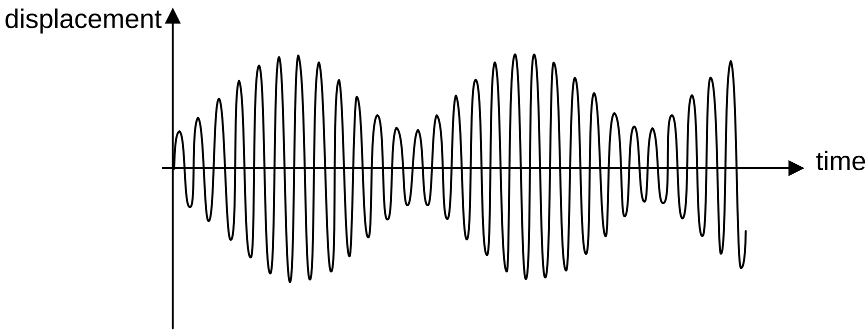

Fig. 3.3

When this waveform is analysed, it is seen to be composed of the sum of three waves of three separate frequencies. These waves are illustrated in the frequency spectrum of Fig. 3.4.

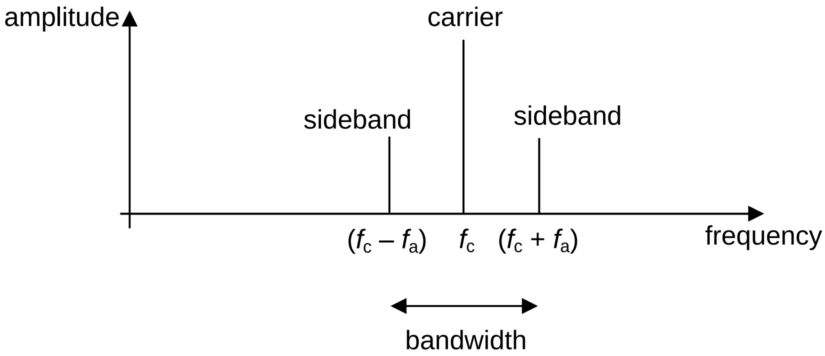

Fig. 3.4

The central frequency f_c is that of the high-frequency carrier wave. The other two are known as *sidebands* and for the AM waveform, they occur at frequencies given by $f_c \pm f_a$, where f_a is the frequency of the audio signal.

The relative amplitude of the sidebands and the carrier depends on the relative amplitudes of the audio and the carrier waveforms. If there is no audio frequency signal, there are no sidebands!

Bandwidth is the frequency range occupied by the AM waveform. This is equal to $2f_a$.

Fig. 3.5 illustrates the AM waveform and the corresponding frequency spectrum for a voice signal.

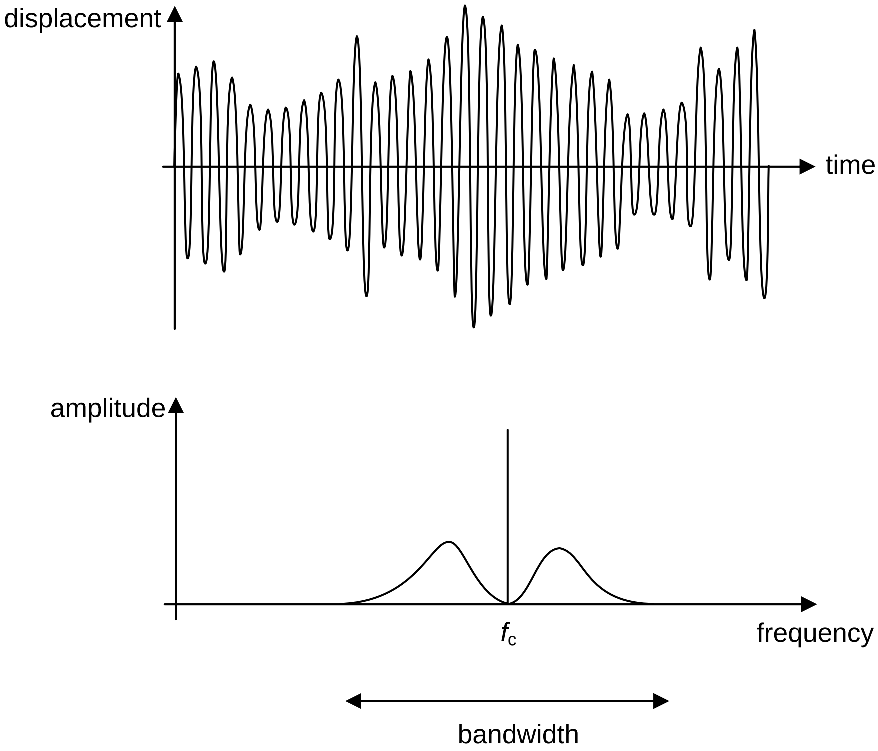

Fig. 3.5

Many audio frequencies are involved. It can be seen that the bandwidth for an AM waveform is the range of frequencies from the lowest to the highest component in the sidebands.

Note that the frequency spectrum of an FM waveform is not the same as that for an AM waveform because further side frequencies that are multiples of the audio frequencies are produced.

(d) Candidates should be able to demonstrate an awareness of the relative advantages of AM and FM transmissions.

An aerial receiving electromagnetic waves cannot distinguish between a genuine radio signal and, say, the interfering radiation from the ignition system of a passing motorbike. If the radio signal is AM, the interference would be considered to be part of the modulation and so it becomes audible in the output produced by the receiver. If, however, the radio signal is FM, the interference will not be picked up by the receiver because it is only variations in frequency that are important, not variations in amplitude. Thus, the quality, in terms of interference, of AM reception is generally poorer than that of FM.

On the long wave (LW) and medium wave (MW) wavebands, the bandwidth on an AM radio station is 9 kHz. This means that the maximum audio frequency that can be broadcast is 4.5 kHz. This frequency is well below the highest frequency audible to the human ear (about 15 kHz) and therefore such broadcasts lack higher frequencies and thus quality.

On the very-high frequency (VHF) waveband, the bandwidth of an FM radio station is about 200 kHz and the maximum audio frequency broadcast is 15 kHz. Thus, the quality of music received on AM is poorer than that of FM but in this case, on the basis of bandwidth.

The LW waveband occupies a region of the electromagnetic spectrum from 30 kHz to 300 kHz. The number of separate AM radio stations that could share this waveband is, theoretically, 270 / 9 = 30. However, the number of separate FM stations would be only 270 /200 = 1. So, more AM radio stations than FM radio stations can share any waveband. For this reason, FM is used only at frequencies in excess of 1 MHz.

The AM transmissions on the LW, MW and SW (short-wave) wavebands are propagated very large distances so that broadcasts can be made to a very large area from only one transmitter. FM transmissions have a range of only about 30 km by line-of-sight. To broadcast to a large area, many FM transmitters are required. It is, therefore, much cheaper and simpler to broadcast by AM than by FM.

AM transmitters and receivers are electronically simpler and cheaper and they also occupy a much smaller bandwidth than those of FM.

(e) Candidates should be able to recall the advantages of the transmission of data in digital form.

Much of the information that is to be communicated in the real world is analogue information (e.g. the voltage output of a microphone that varies with time in a similar manner to the sound waveform that caused it). If this analogue signal is to be transmitted over a large distance (either by radio or by cable) it will be attenuated and it will pick up noise.

Attenuation is a gradual reduction in signal power. This could be, for example, ohmic losses in a metal cable. In any electrical system there is always unwanted power present that adds itself in a random manner to the signal. This unwanted random power is called *noise* and it causes distortion of the signal. There are several sources of noise. One arises from the thermal vibrations of the atoms of the material through which the signal is passing. As a result, noise power cannot be totally eliminated.

Attenuation will mean that, eventually, the signal will have to be amplified so that it can be distinguished from the background noise. This is achieved using a repeater amplifier that amplifies the signal before passing it further on. The amplifier will, however, amplify the noise as well as the original signal. After several of these repeater amplifications (required for transmission over long distances), the signal will become very 'noisy'. This effect is illustrated in Fig. 3.6.

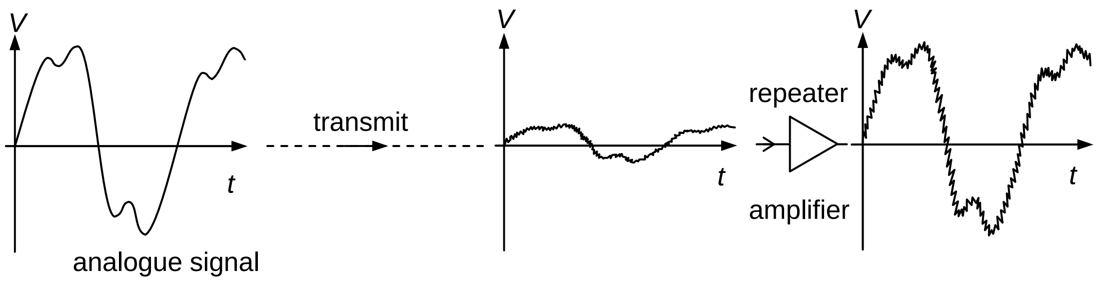

Fig. 3.6

If the signal is transmitted in digital form, then it also suffers from attenuation and the addition of noise. However, the amplifiers that are used for amplifying digital signals are required only to produce a 'high' voltage or a 'low' voltage. They are not required to amplify small fluctuations in amplitude, as is the situation for amplification of an analogue signal. Since noise consists, typically, of small fluctuations, the amplification of a digital signal does not also amplify the noise. Such amplifiers are called *regenerator amplifiers* and are able to reproduce the original digital signal and, at the same time, 'filter out' the noise. This is illustrated in Fig. 3.7.

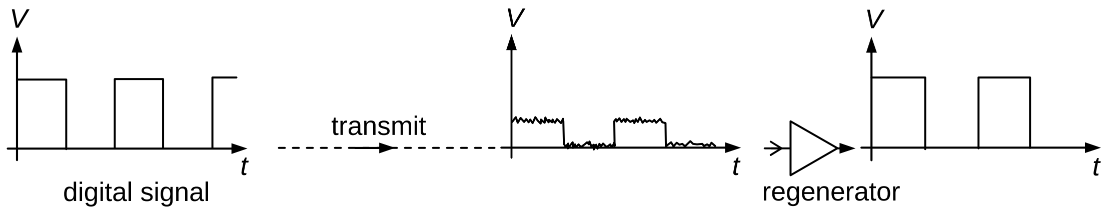

Fig. 3.7

As a result, a digital signal can be transmitted over very long distances with regular regenerations without becoming increasingly noisy, as would happen with an analogue signal.

A further advantage of digital transmissions is that they can have extra information – extra bits of data – added by the transmitting system. These extra data are a code to be used by the receiving system to check for errors and to correct them before passing the information on to the receiver.

Nowadays, digital circuits are generally more reliable and cheaper to produce than analogue circuits. This is, perhaps, the main reason why, in the near future, almost all communication systems will be digitally based.

(f) Candidates should be able to understand that the digital transmission of speech or music involves analogue-to-digital conversion (ACD) on transmission and digital-to-analogue conversion (DAC) on reception.

(g) Candidates should be able to show an understanding of the effect of the sampling rate and the number of bits in each sample on the reproduction of an input signal.

The electrical signals derived from speech or music are analogue audio-frequency signals. The voltage generated varies continuously. To convert an analogue signal into a digital signal involves taking samples of the analogue waveform (i.e. measuring its instantaneous voltage) at regular intervals of time. The instantaneous or sample voltage is converted into a binary number that represents its value.

For example, if the instantaneous value of the analogue signal is 6 V, the binary number could be 0110. For an instantaneous value of 13 V, the binary number could be 1101.

Note that a binary digit is referred to as a *bit*. The most significant bit (MSB) – the bit representing the largest decimal number is written first. The bit representing the lowest decimal number (1) is known as the least significant bit (LSB) and is written last.

A digital signal consists of a series of 'high' and 'low' voltages. A 1 represents a 'high' voltage and a 0 represents a 'low' voltage. A 4-bit system is used in the examples in this booklet. In reality, 8 or more bits would be used for any sampling.

Fig. 3.8(a) shows an analogue signal of frequency 1 kHz. This signal is sampled every 125 μs (a sampling frequency of 8 kHz). The sample voltages are shown in Fig. 3.8 (b). It should be noted that the value given to the sampled voltage is always the value of the nearest increment *below* the actual sample voltage. In this particular example, an analogue signal of 14.3 V would be sampled as 14 V and one of 3.8 V would be sampled as 3 V. The resulting digital signal is shown in Fig. 3.8 (c). Each number is a group of 4 bits and these groups are separated in time by 125 μs.

The choice of sampling frequency is important. A lower sampling frequency means that less information can be gathered from the analogue signal. More than eighty years ago, it was shown by Nyquist that, in order to be able to recover the analogue signal from its digital conversion, the sampling has to occur at a frequency greater than twice the highest frequency component in the original signal. As a result, in the telephone system, the highest frequency is restricted to 3.4 kHz because the sampling frequency is 8 kHz. In the manufacture of compact discs, the highest frequency is 20 kHz and the sampling frequency is 44.1 kHz.

After the analogue signal has been converted to a 4-bit digital signal by the analogue-to-digital converter (ADC), the digital signal is transmitted. The original signal can be recreated by passing the 4-bit numbers into a digital-to-analogue converter (DAC). This is illustrated in Fig. 3.8(d) where the original analogue signal of Fig. 3.8(a) has been recreated.

The output of the DAC is 'grainy' and is not smooth because the number of bits limits the number of possible voltage levels (with 4 bits there are $2^4 = 16$ levels; with 8 bits, there are $2^8 = 256$ levels). As described above, a higher sampling frequency also enables more detail of the analogue signal to be recovered.

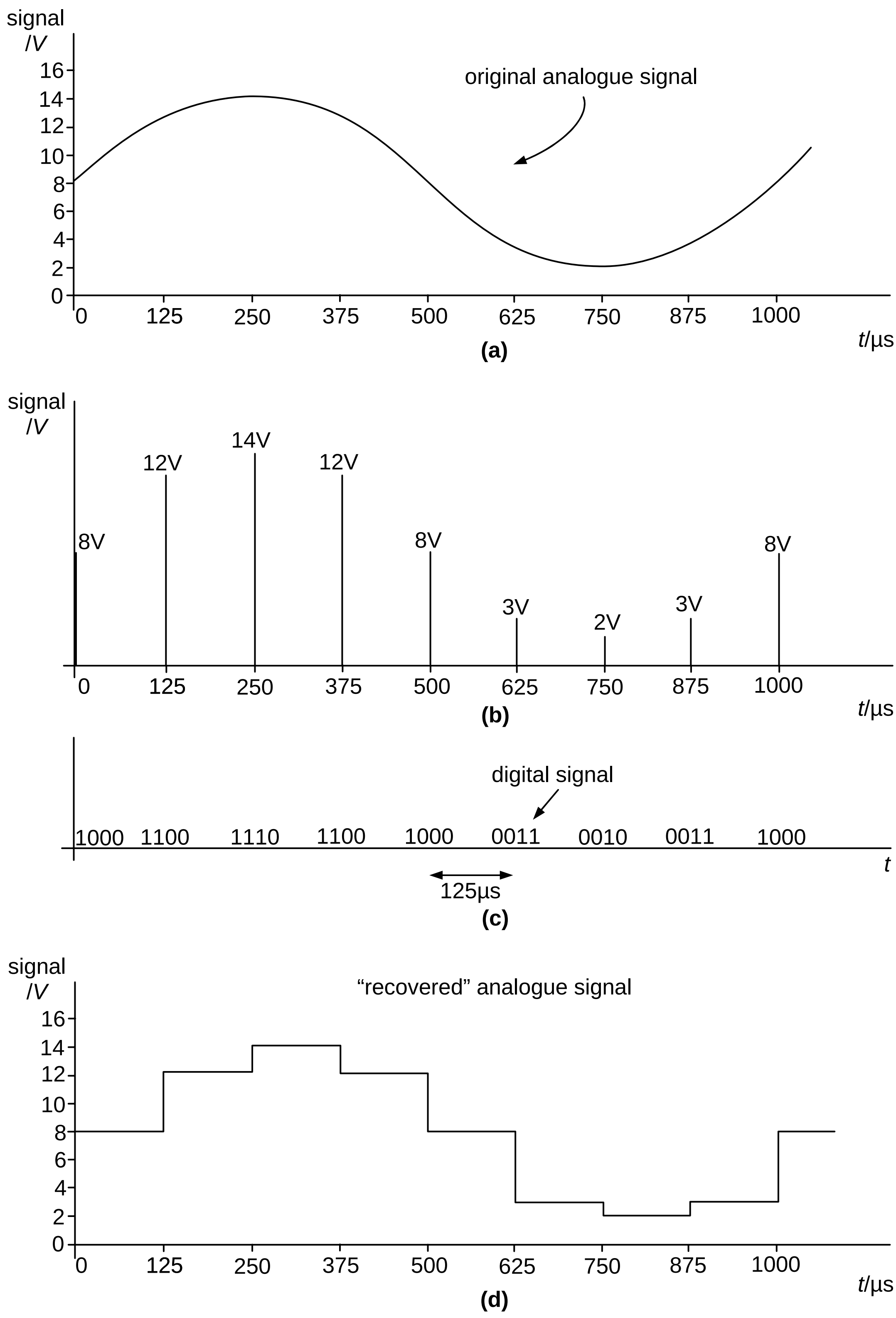
signal /V
16
14
12
10
8
6
4
2
0
original analogue signal
0
125
250
375
500
625
750
875
1000
t/µs
(a)
signal /V
8V
12V
14V
12V
8V
3V
2V
3V
8V
0
125
250
375
500
625
750
875
1000
t/µs
(b)
digital signal
1000
1100
1110
1100
1000
0011
0010
0011
1000
t
125µs
(c)
signal /V
"recovered" analogue signal
16
14
12
10
8
6
4
2
0
0
125
250
375
500
625
750
875
1000
t/µs
(d)

Fig. 3.8

(h) *Candidates should be able to appreciate that information may be carried by a number of different channels, including wire-pairs, coaxial cables, radio and microwave links, and optic fibres.*

(i) *Candidates should be able to discuss the relative advantages and disadvantages of channels of communication in terms of available bandwidth, noise, cross-linking, security, signal attenuation, repeaters and regeneration, cost and convenience.*

(l) *Candidates should be able to recall the frequencies and wavelengths used in different channels of communication*

Wire-pairs
In the early days of electrical communication, a transmitter was connected to a receiver by a pair of insulated copper wires. Fig. 3.9 illustrates an arrangement for transmitting information in digital code (Morse code).

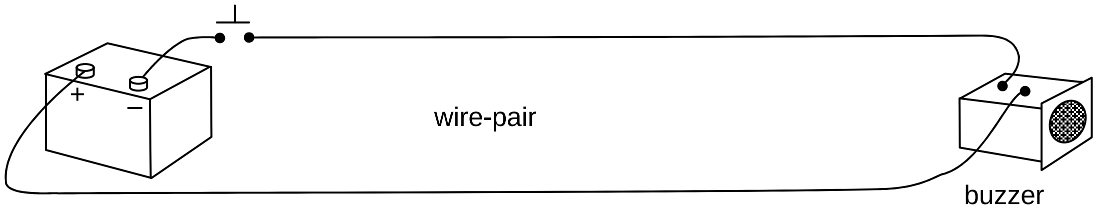

Fig. 3.9

Wire-pairs provide a very simple link. In modern communications, wire-pairs are used mainly for very short distances with low frequencies.

If high frequency signals are transmitted along a pair of wires over an appreciable distance, repeated amplification must be provided at regular intervals. This is due to the very high attenuation of the signal. Energy is lost as heat in the resistance of the wires and also as radiation since the wires act as aerials. A further problem is that the wires easily pick up external interference that degrades the original signal. If several wire-pairs are arranged next to one another, they will pick up each other's signals. This effect is known as *cross-talk* or *cross-linking* and gives very poor security as it is easy to 'tap' a telephone conversation.

The bandwidth of a pair of wires is only about 500 kHz. Consequently, as a means of carrying a large amount of information, it is extremely limited.

Coaxial cable
Coaxial cable is, essentially, a pair of wires arranged so that one wire is shrouded by the other, as illustrated in Fig. 3.10.

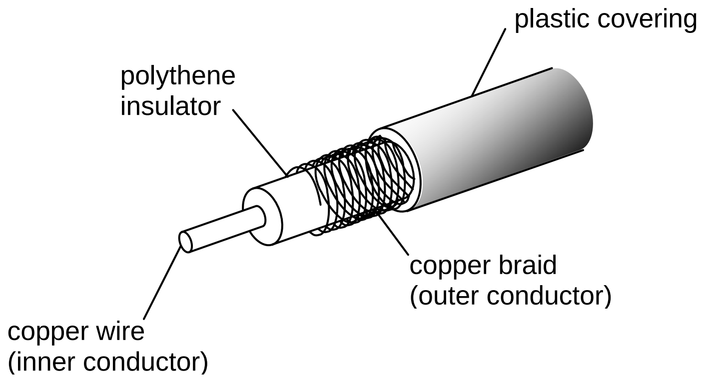

Fig. 3.10

The signal is transmitted down the inner conductor and the outer conductor acts as the return wire and also shields the inner one from external interference. The outer conductor is usually connected to earth.

Coaxial cable is more expensive than wire-pairs but causes less attenuation of the signal. This means that, for long distance communication, repeater amplifiers can be arranged further apart. Coaxial cables are less prone to external interference, though not immune to it, so they do offer slightly greater security.

The bandwidth of coaxial cable is about 50 MHz. It is capable of carrying much more information than a wire-pair.

Radio link

When radio was first developed, an electrical oscillation of a few kilohertz (the carrier wave) was linked to a long wire – the aerial. The oscillations were switched on and off. In this way, information was transmitted from the aerial in digital form (Morse code). It soon became possible to modulate the carrier wave (by AM or by FM) so that information could be sent at a much faster rate. Different carrier frequencies allowed different radio stations to share the same air space (frequency multiplexing).

Energy that is radiated from an aerial is in the form of electromagnetic waves and is propagated at the speed of light. If the frequency of the transmitted waves are somewhere in the range from 30 kHz to 3 GHz, then the waves are known as radio waves.

The electromagnetic radiation that is emitted from a transmitting aerial can be arranged (by suitable choice of the aerial) to radiate in all directions (e.g. for national broadcasting). For point-to-point communications, the aerial can be arranged to radiate mostly in one direction. No matter what aerial is used, there is always energy loss and the power of the signal picked up by a receiving aerial is reduced as the distance between the transmitter and the receiving aerial is increased. The actual distance any particular waves propagate is dependent on frequency, as illustrated in Fig. 3.11.

type of wave	frequency	range
surface wave	below 3 MHz	up to 1000 km
sky wave	3 MHz $\rightarrow$ 30 MHz	worldwide by means of reflection from ionosphere and ground
space wave	greater than 30 MHz	line of sight – including satellite communication

Fig. 3.11

As a means of communicating from a single transmitter over a large area, the AM broadcasts on the LW and MW are relatively cheap and technically simple, as explained in 3(d).

In modern communication, considerable use is made of the VHF and UHF wavebands for mobile phones, walkie-talkie radio etc. This is due to the fact that, at these frequencies, the wavelength is relatively small and hence the aerial can be made conveniently short.

The part of the electromagnetic spectrum used for radio communication is shown in Fig. 3.12.

	frequency band	frequencies	wavelengths (in a vacuum)
LW radio	low frequencies LF	30 kHz $\rightarrow$ 300 kHz	10 km $\rightarrow$ 1 km
MW radio	medium frequencies MW	300 kHz $\rightarrow$ 3 MHz	1 km $\rightarrow$ 100 m
SW radio	high frequencies HF	3 MHz $\rightarrow$ 30 MHz	100 m $\rightarrow$ 10 m
FM radio	very high frequencies VHF	30 MHz $\rightarrow$ 300 MHz	10 m $\rightarrow$ 1 m
TV broadcast	ultra-high frequencies UHF	300 MHz $\rightarrow$ 3 GHz	1m $\rightarrow$ 10 cm
microwave/satellite	super-high frequencies SHF extra-high frequencies EHF	3 GHz $\rightarrow$ 30 GHz 30 GHz $\rightarrow$ 300 GHz	10 cm $\rightarrow$ 1 cm 1 cm $\rightarrow$ 1mm

Fig. 3.12

The bandwidth of a radio link increases as the frequency of the carrier wave increases.

Microwave link

Microwaves are radio waves in the SHF waveband from 3 GHz to 30 GHz with wavelengths of only a few centimetres. They are generally used for point-to-point communication, as illustrated in Fig. 3.13.

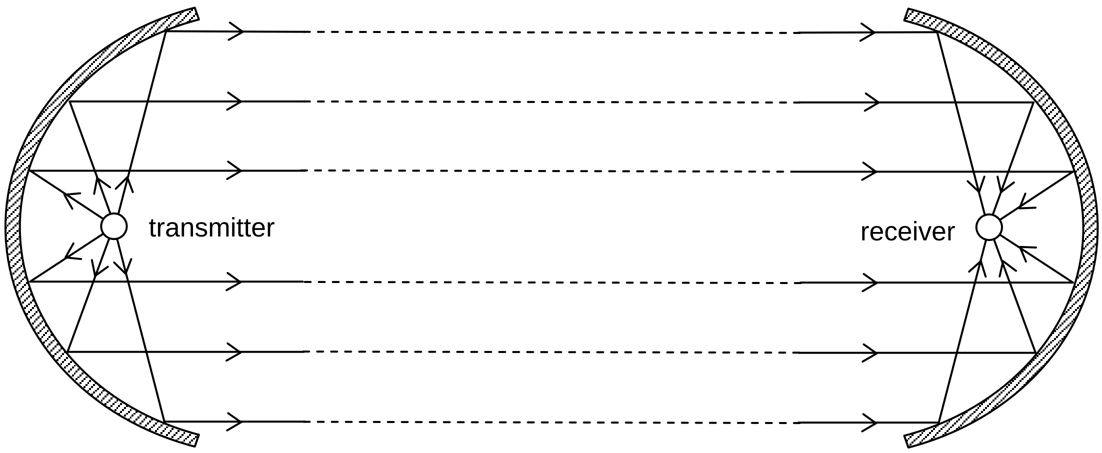

Fig. 3.13

The transmitting element is placed at the focus of a parabolic mirror. This causes the wave power to be radiated in a parallel beam. A parabolic reflector, placed in the path of this beam, reflects and focuses the wave power on to a receiving element. The reflecting parabolic dishes are the not aerials themselves. They are a means of directing as much power as possible into a parallel beam and then collecting this power and directing it to the receiving aerial or element. Parabolic dishes are most useful with short wavelengths where the spread of the waves due to diffraction is less pronounced.

The bandwidth of a microwave link is of the order of GHz. Consequently, microwave links have a very large capacity for carrying information. However, for terrestrial use, the range of the transmissions is limited to line-of-sight. For long-distance transmissions, many repeater stations are required.

Optic fibres

Optic fibres carry digital information in the form of pulses of light or infra-red radiation. These pulses are provided by lasers and the light produced has very high frequencies of the order of 10^8 MHz. In theory, a bit or individual light wave could last for only 10^{-14} s. This would allow hundreds of thousands of individual telephone calls to share the same optic fibre. However, present technology does not allow control at such high frequencies. The duration of a bit is governed by how fast the laser providing light to the fibre can be switched on and off. This is, at present, of the order of GHz but is increasing as technology develops.

The advantages of transmission using optic fibres are indicated below.

- Optic fibres have a wide bandwidth. This gives rise to a large transmission capacity.
- Signal power losses in optic fibres are relatively small. This allows for longer uninterrupted distances between regenerator amplifiers and reduces the costs of installation.
- The cost of optic fibre is much less than that of metal wire.
- The diameter and weight of fibre optic cables is much less that that of metal cables. This implies easier handling and storage.
- Optic fibres have very high security since they do not radiate energy and thus there is negligible 'cross-talk' between fibres.
- Optic fibres do not pick up electromagnetic interference. This means they can be used in electromagnetically 'noisy' environments, for example alongside electric railway lines. In fact, optic fibre cables are installed along the routes of the National Grid.
- Optic fibre is ideal for digital transmissions since the light is obtained from lasers that can be switched on and off very rapidly.

(j) Candidates should be able to describe the use of satellites in communication.

(k) Candidates should be able to recall the relative merits of both geostationary and polar orbiting satellites for communicating information.

The basic principle of satellite communication is illustrated in Fig. 3.14.

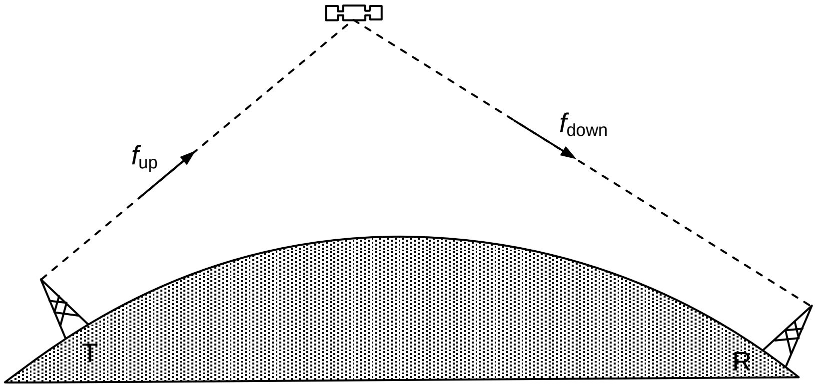

Fig. 3.14

A transmitting station T directs a carrier wave of frequency f_{up} towards the satellite. The satellite receives this signal, amplifies it and changes the carrier frequency to a lower value f_{down} before directing it towards a receiver R back on Earth. Typically the uplink would have a frequency f_{up} of 6 GHz and the downlink would have a frequency f_{down} of 4 GHz (the 6/4 GHz band). Alternatives are the 14/11 GHz band and the 30/20 GHz band. The two carrier frequencies are different to prevent the satellite's high power transmitted signal swamping its reception of the very low power signal that it receives. There is no interference of the actual information being carried by the waves because this is stored as a modulation of the carrier waves.

Although the transmitter in Fig. 3.14 could transmit more or less directly to the receiver without the use of a satellite, it could only do so on the SW or MW wavebands, as described in section 3(h). However, in modern communication systems, this is not done for three reasons.

(i) Long-distance communication on these wavebands is unreliable. Sky waves rely on ionospheric reflection. These layers of ions vary in height and density according to the time of day. In hilly areas, surface waves give rise to regions of poor reception where there are 'shadows'.

(ii) The wavebands are already filled by existing broadcasts.

(iii) The available bandwidths are too narrow to carry the required amount of information.

Satellites may orbit the Earth in polar orbits, as illustrated in Fig. 3.15.

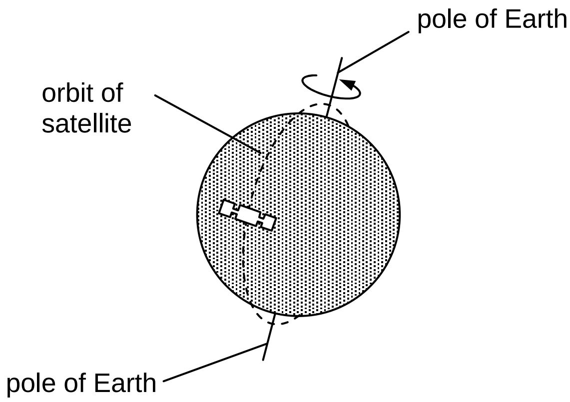

Fig. 3.15

Polar orbits are relatively low with a period of rotation of the order of 90 minutes. Such satellites will, as a result of the rotation of the Earth, at some time each day orbit above every point on the Earth's surface. For a satellite having a period of 90 minutes, each orbit crosses the Equator 23° to the west of the previous orbit.

It is not possible to have continuous communication links with one such satellite because, from Earth, the satellite appears to move rapidly across the sky and, for part of the time, is below the horizon. Polar orbiting satellites are used, as well as for communications, for monitoring the state of the Earth's surface, weather forecasting, spying etc.

Alternatively, satellites may be placed in geostationary orbit, as illustrated in Fig. 3.16.

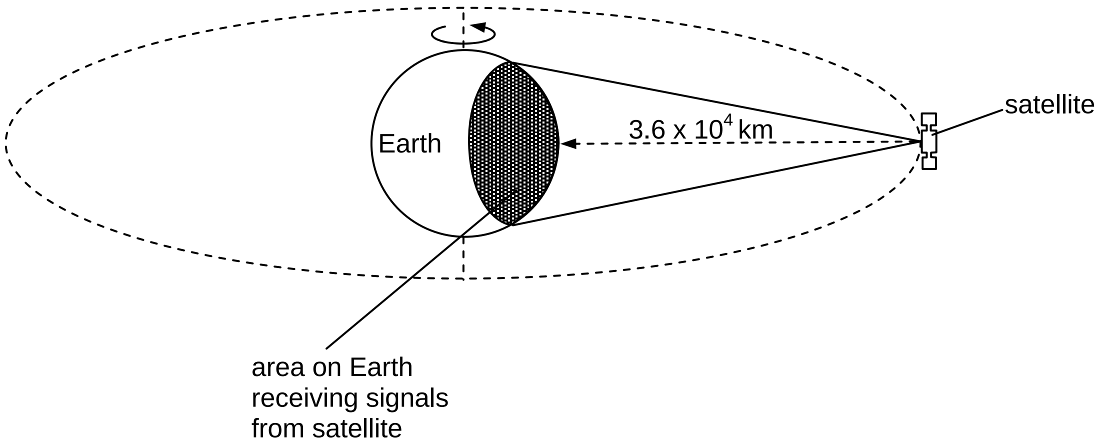

Fig. 3.16

Geostationary satellites orbit the Earth above the Equator with a period of 24 hours at a distance of 3.6×10^4 km above the Earth's surface. If the satellite is orbiting in the same direction as the direction of rotation of the Earth, then, for an observer on the Earth, the satellite will always appear to be above a fixed position on the Equator.

The satellite can allow for continuous communication between a ground station and anywhere on the surface of the Earth that can receive the signal from the satellite. A number of such satellites with over-lapping areas of communication are used for trans-oceanic telephone calls, removing the need for long-distance submarine cables. International television broadcasts are possible, enabling viewers in one country to watch television broadcasts from another.

It should be remembered that geostationary satellites are in Equatorial orbit and thus polar regions may not be in line of sight with a satellite.

Where communication is possible, the height above the Earth's surface of the satellites gives rise to delays in conversation between two people using a satellite link. This delay would be unacceptable where several satellites provide the complete link. For this reason, geostationary satellites may be used in conjunction with optical fibres.

Polar orbiting satellites are used for communication since they are in low orbits, resulting in short time delays between transmission and receipt of a signal. Furthermore, total global coverage is possible. However, a network of such satellites is required in order to maintain continuous links. The satellites must be tracked and the link switched from one satellite to another. Geostationary satellites have the advantage that they do not need to be tracked.

(m) Candidates should be able to understand and use signal attenuation expressed in dB and dB per unit length.

(n) Candidates should be able to recall and use the expression number of dB = 10lg(P_1/P_2) for the ratio of two powers.

When an electrical signal is transmitted along a metal wire, it gradually loses power, mostly as thermal energy in heating the wire. Similarly, a light pulse travelling along an optic fibre loses power, mostly by absorption due to impurities in the glass and by scattering due to imperfections. Electromagnetic waves lose power by absorption and dispersion. A reduction in signal power is referred to as *attenuation*.

In order that a signal may be detected adequately, its power must be a minimum number of times greater than the power associated with noise (see the section on 30(e)). Typically, this signal-to-noise ratio could be 100.

Repeater amplifiers may be required to increase the power of a signal that is being passed along a transmission line. The gain of such an amplifier (the ratio of the output power to the input power) could be 100000. For a radio link between Earth and a geostationary satellite, the power received by the satellite may be 10^{19} times smaller than that transmitted from Earth.

It can be seen from the above examples that the ratio of the two powers may be very large. Consequently, an extremely convenient unit by which power levels, or any other quantities, may be compared is the bel (B). The number of bels is related to the ratio of two powers P_1 and P_2 by the expression

$$\text{number of bels} = \lg(P_1/P_2).$$

In practice, the ratios are usually expressed in decibels (dB), where 10 dB = 1 B. Consequently,

$$\text{number of decibels} = 10 \lg(P_1/P_2).$$

Example

The gain of an amplifier is 45 dB. Calculate the output power P_{out} of the amplifier for an input power P_{in} of 2.0 μW.

number of decibels = $10 \lg(P_1/P_2)$

$$\begin{aligned} 45 &= 10 \lg(P_{out} / 2.0 \times 10^{-6}) \\ 4.5 &= \lg(P_{out} / 2.0 \times 10^{-6}) \\ 10^{4.5} &= (P_{out} / 2.0 \times 10^{-6}) \\ P_{out} &= 10^{4.5} \times 2.0 \times 10^{-6} \\ P_{out} &= 6.3 \times 10^{-2}\ \text{W} \end{aligned}$$

A transmission line has an input power P_2 and the power at a point distance L along the line is P_1 as illustrated in Fig. 3.17.

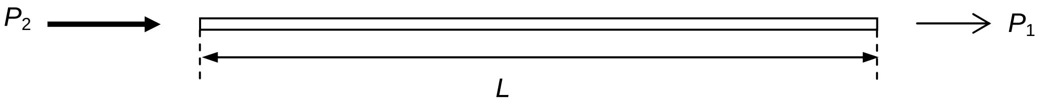

Fig. 3.17

Then, attenuation in the line = $10 \lg (P_2 / P_1)$ dB.

Since a transmission line may vary in length, an important feature of a transmission line is its attenuation per unit length.

$$\text{attenuation per unit length} = \frac{1}{L} 10 \lg \frac{P_2}{P_1}$$

Example

The input power to a cable of length 25 km is 500 mW. The attenuation per unit length of the cable is 2 dB km^{-1}. Calculate the output power of the signal from the cable.

signal loss in cable = $2 \times 25 = 50$ dB

$50 = 10 \lg(500 \times 10^{-3} / P_{out})$, where P_{out} is the output power.

$P_{out} = 500 \times 10^{-3} \times 10^{-5} = 5 \times 10^{-6}$ W.

The signal cannot be allowed to travel indefinitely in the cable because, eventually, it will become so small that it cannot be distinguished from background noise. An important factor is the minimum *signal-to-noise ratio* that effectively provides a value for the lowest signal power allowed in the cable.

In the above example, the background noise is 5×10^{-13} W and the minimum signal-to-noise ratio permissible is 20 dB. Then if P_M is the minimum signal power,

$20 = 10 \lg(P_M / 5 \times 10^{-13})$

$P_M = 5 \times 10^{-13} \times 10^2 = 5 \times 10^{-11}$ W.

This enables the maximum uninterrupted length of cable along which the signal can be transmitted to be determined.

Maximum loss in cable = $10 \lg(500 \times 10^{-3} / 5 \times 10^{-11}) = 120$ dB

Maximum distance = 120 / 2 = 60 km.

(o) *Candidates should be able to understand that, in a mobile-phone system, the public switched telephone network (PSTN) is linked to base stations via a cellular exchange.*

In the early days of telephones, each telephone user was connected to all other users by their own cables. This was feasible only where the number of users was small as in, for example, a single building.

As telephones became more popular and widespread, connections between individual users became impractical. Consequently, the telephone exchange was developed. The caller would contact the telephone exchange and, at the exchange, the connection to the other user would be made by a person known as an ‘operator’. If the call was not a local call served by that particular exchange, then the local exchange would contact the other user’s local exchange via a trunk exchange. Trunk exchanges were connected via trunk lines and hence the expression ‘trunk call’ for any long-distance call.

In essence, the Public Switched Telephone Network (PSTN) uses the same principles of exchanges but has developed with modern technology and the number of users. For example, switching is no longer achieved using operators. Electromagnetic relays have been replaced by solid-state devices and international ‘exchanges’, or gateways, have been introduced. The PSTN is illustrated in Fig. 3.18.

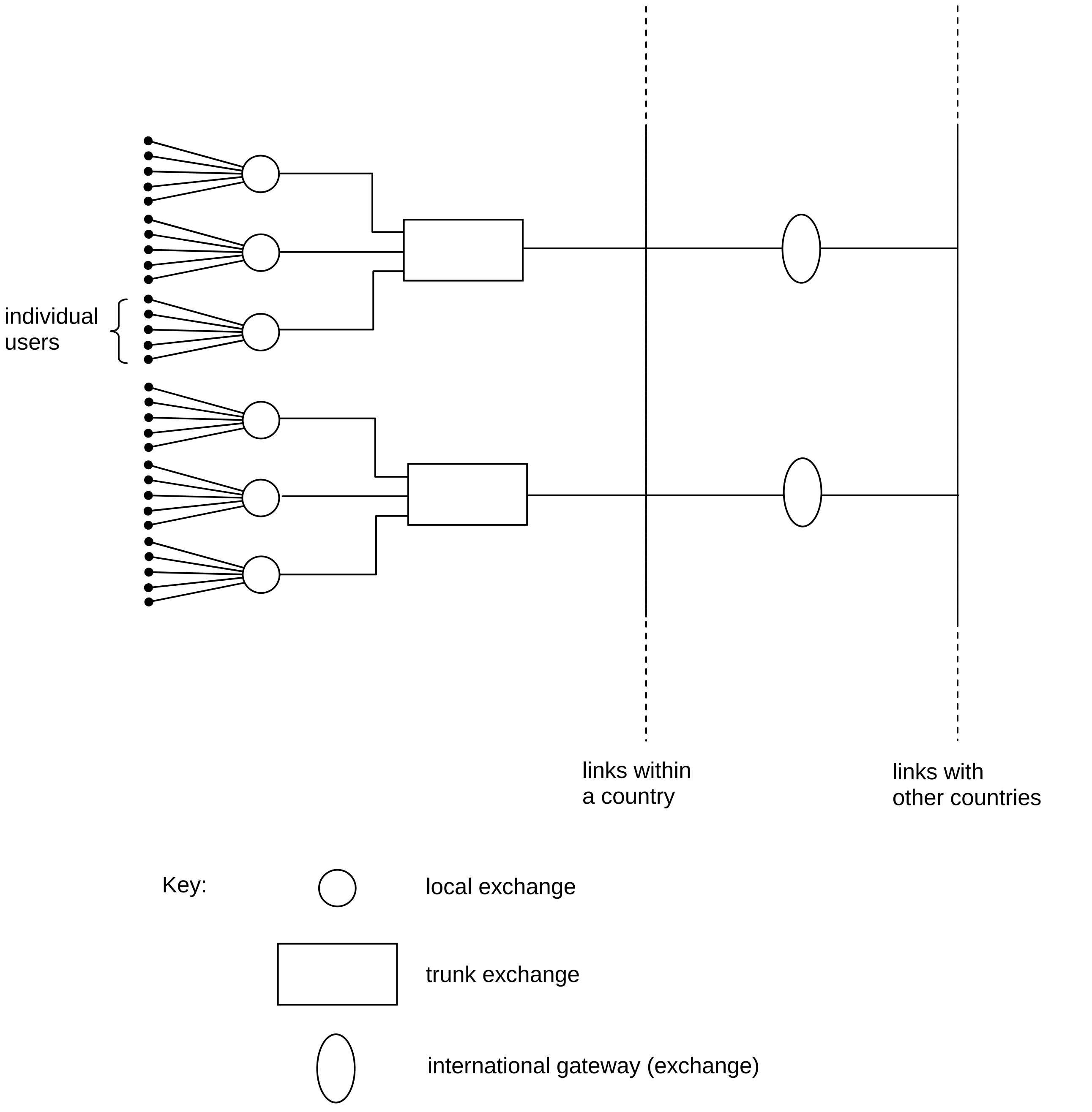

Fig. 3.18

In the system illustrated in Fig. 3.18, the user is connected to the PSTN via the local exchange. Each user has a ‘fixed line’ to the local exchange, resulting in the user having limited mobility whilst making the call.

During the 1970s and 1980s, mobile phone systems were developed that did not have a permanent link to a local exchange. Basically, a mobile phone is a *handset* that is a radio transmitter and receiver. When a call is to be made, the user makes a radio-wave link with a nearby *base station*. This base station is connected by cable to a *cellular exchange*. The cellular exchange then allows connection to be made to the PSTN. This is illustrated in Fig. 3.19.

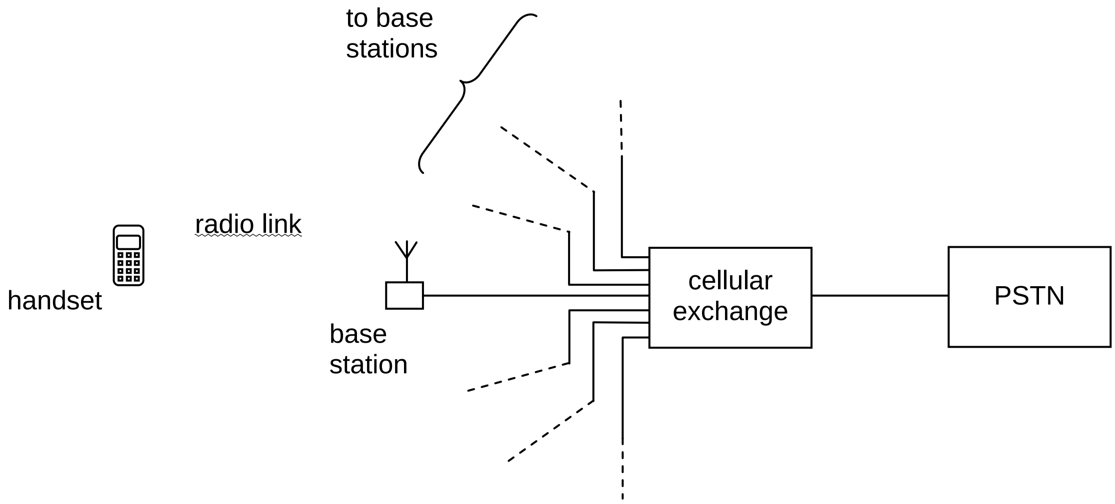

Fig. 3.19

(p) Candidates should be able to understand the need for an area to be divided into a number of cells, each cell served by a base station.

(q) Candidates should be able to understand the role of the base station and the cellular exchange during the making of a call from a handset.

The popularity of mobile phones means that large numbers of people use a mobile phone system at the same time. However, the range of carrier-wave frequencies for linking between the mobile phone and the base station is limited. Consequently, each mobile phone cannot have its own carrier frequency. This means that the same carrier frequencies must be used by many mobile phones at the same time. This is achieved using a network of base stations.

The base stations operate on UHF frequencies so that they have a limited range (see the section on 30(h)) and are low-power transmitters. The UHF frequencies also mean that the aerial in the mobile phone is conveniently short! A country is divided into areas or *cells*, with each cell having its own base station, usually located near the centre of the cell. The aerial at the base station transmits in all directions so as to cover the whole cell, but not to overlap too far into neighbouring cells. In this way, the whole country is 'covered'. Neighbouring cells cannot use the same carrier frequencies, otherwise interference would occur at the boundaries between cells. A possible arrangement of cells is shown in Fig. 3.20.

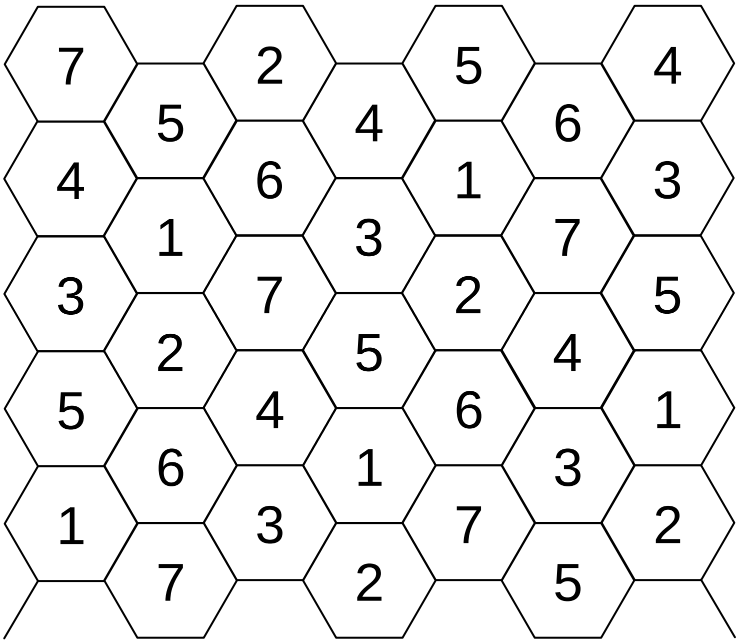

Fig. 3.20

Although each cell is approximately circular (depending on the flatness of the land), the cells are shown as a 'honeycomb' so that the cells fit together. The number in each cell represents a particular range of carrier frequencies that would be allocated to each cell. Neighbouring cells do not have the same range of carrier frequencies.

When a handset is switched on, it transmits a signal to identify itself. This signal is received by a number of base stations, from where it is transferred to the cellular exchange. A computer at the cellular exchange selects the base station with the strongest signal from the handset. The computer also allocates a carrier frequency for communication between the base station and the handset. During communication between the handset and the base station, the computer at the cellular exchange monitors the signal from the handset. If the user of the handset moves from one cell to another, the signal strength changes. The call from the handset is then re-routed through the base station with the greater signal.

(r) *Candidates should be able to recall a simplified block diagram of a mobile-phone handset and understand the function of each block.*

A mobile-phone handset is, in its simplest form, a radio transmitter and receiver. A simplified block diagram of its circuitry is shown in Fig. 3.21.

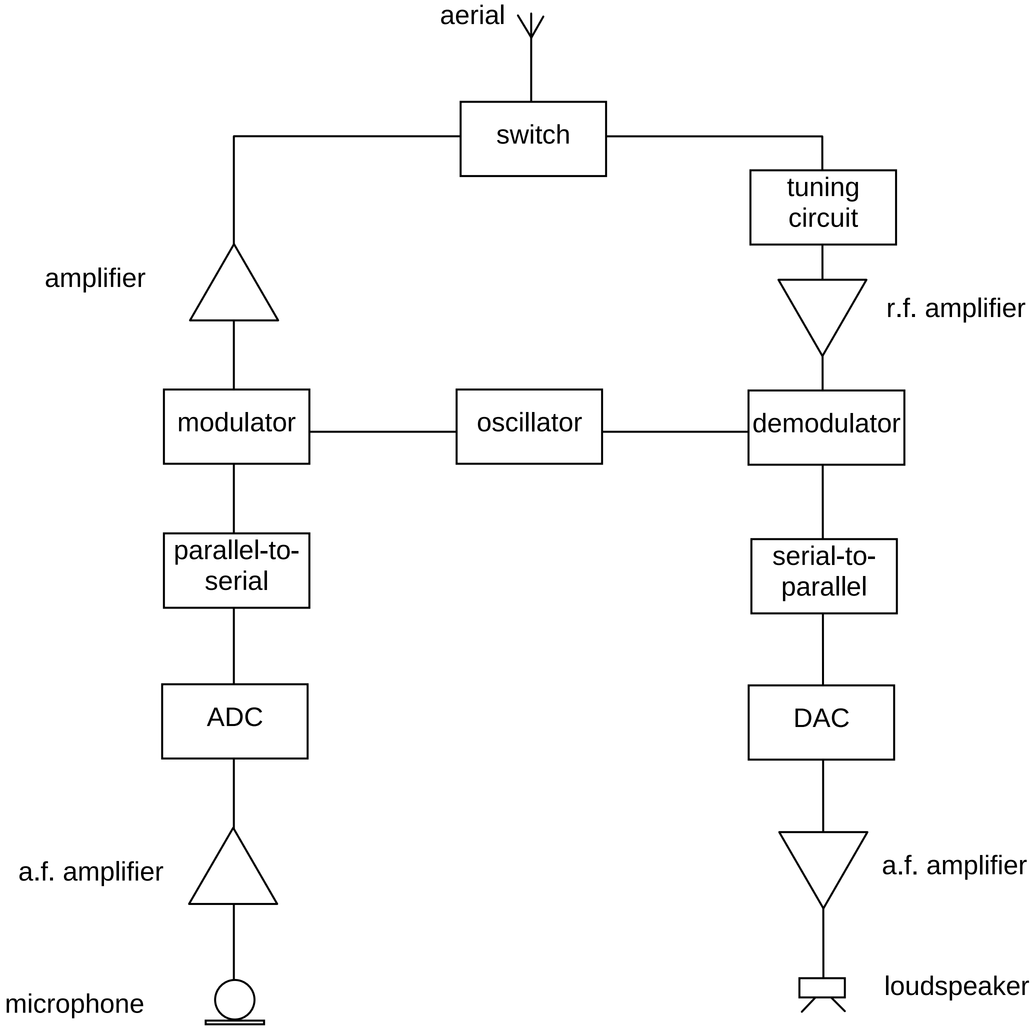

Fig. 3.21

The caller speaks into the microphone. This produces a varying signal voltage that is amplified and converted to a digital signal by means of the ADC. The parallel-to-series converter takes the whole of each digital sample voltage and then emits it as a series of bits. The series of bits is then used to modulate the chosen carrier wave. After further amplification, the modulated carrier wave is switched to the aerial where it is transmitted as a radio wave.

On receipt of a signal at the aerial, the signal is switched to a tuning circuit that selects only the carrier-wave frequency allocated to it by the computer located at the cellular exchange. This selected signal is then amplified and demodulated so that the information signal is separated from the carrier wave. This information signal is in digital form. It is processed in a series-to-parallel converter to produce each sample digital voltage and then in a digital-to-analogue converter (DAC) to provide the analogue signal. After amplification, the analogue signal is passed to a loudspeaker.