OPTION M

Medical Physics

M1. Medical Imaging

1 (a) *Candidates should be able to describe in simple terms the need for non-invasive techniques of diagnosis.*

Historically, diagnosis consisted of two techniques – observing the patient outwardly for signs of fever, vomiting, changed breathing rate etc, and observing the patient inwardly by surgery. The first technique depended greatly on experience but was still blind to detailed internal conditions. The second quite often led to trauma and sometimes death of the patient. In earlier times there was also the significant risk of post-operative infection.

Modern diagnostic techniques have concentrated on using externally placed devices to obtain information from underneath the skin. X-rays have been used for a century. More recently, ultrasonic waves have been used – especially in cases of pregnancy where X-rays may be harmful. Other techniques use lasers which can shine through a finger or are used in a very narrow tube which is inserted into the body via various orifices. Another technique utilises a response to the presence of a large magnetic field (magnetic resonance imaging).

In all these cases, the aim is to obtain detailed information concerning internal structures of the body, thereby outlining progress or searching for signs of abnormality. This is achieved without the need of the surgeon's knife and is described as a *non-invasive* technique. Non-invasive techniques are designed to cause as little risk as possible to the patient and are much safer and far less traumatic than surgery.

1 (b) *Candidates should be able to show a qualitative understanding of the importance of limiting exposure to radiation with particular reference to the type of radiation.*

In each of the non-invasive techniques discussed, it is necessary to obtain a clear picture of the region under investigation. When using radioactive tracers or X-rays, radiation exposure must continue until that objective is achieved. Effectively, energy is transmitted and selectively absorbed. The transmitted radiation is viewed – either directly, or by photographic plate – or detected by some electronic means and subsequently converted to visual form.

In each case, radiation energy is involved. It is essential to restrict absorbed energy to a minimum, as localised energy absorption can cause harmful effects to the body (local concentrations cause burns whilst low doses may cause abnormalities within cells).

Energy associated with infra-red radiation and the visible spectrum is not considered to be harmful unless intense beams are involved. Ultrasonic waves are widely used in diagnosis and therapy. Whenever energy is deposited in cells there is some risk of damage to the cells. However, by suitable choice of frequency, intensity and exposure, any risk of damage is minimal. As is the situation for other types of radiation, the use of ultrasound is governed by a Code of Practice.

With magnetic resonance imaging – a relatively new technique – detailed knowledge of the internal structure of the skull is thought to be well worth any possible but as yet unknown risk in exposing the head to magnetic fields of flux densities in excess of one tesla. The alternative is the much greater risk associated with surgery. In both these cases (ultrasound and magnetic resonance), investigations are continuing into any possible long term effects.

X-rays are sufficiently energetic to cause ionisation in matter – with the possible consequent destruction of cells or damage to DNA. Exposure time is curtailed so that damage is kept to a minimum. The outcome is sufficient diagnostic detail so that the correct treatment can be specified with as little risk as possible to the patient.

Radioactive tracers can be injected or ingested. Ionising radiation is again involved and the dose applied is kept to a minimum so that the harmful effects of radiation and risk of long term effects are minimised (see also section *2(b)*).

1 *(c) Candidates should be able to explain the principles of production of X-rays by electron bombardment of a metal target.*

X-rays are produced by bombarding metal targets with high speed electrons. A simple X-ray tube is shown in Fig. 1.1.

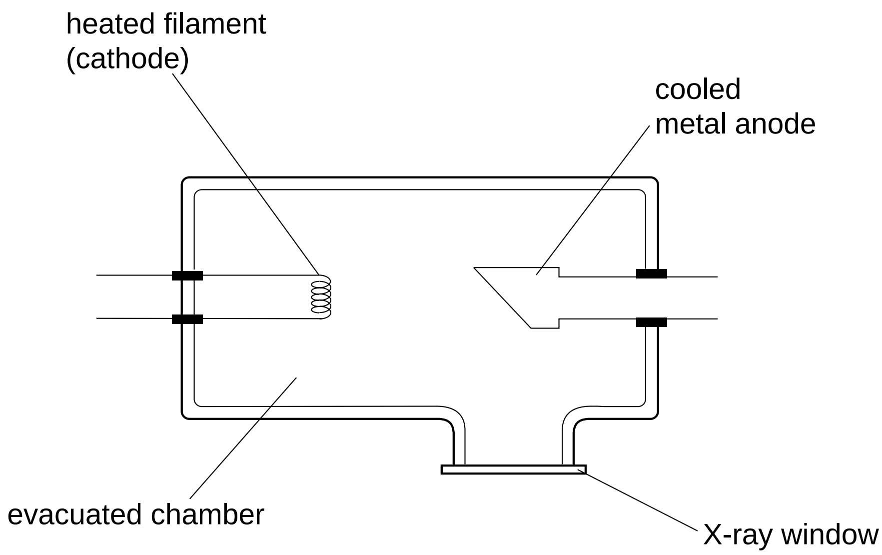


Fig. 1.1

In order to provide the electrons with the necessary energy, they are accelerated through large potential differences (many kV). The electrons are produced from a heated metal filament.

Whenever a charged particle is accelerated, electromagnetic radiation is emitted: the greater the acceleration, the shorter is the wavelength of the emitted radiation. This radiation is known as Bremsstrahlung radiation. When electrons bombard the target, they suffer large accelerations, and radiation in the X-ray band is emitted. Since the electrons have many different accelerations, a continuous distribution of wavelengths is emitted.

As well as the continuous distribution of wavelengths, sharp peaks may be observed in the X-ray spectrum, as shown in Fig. 1.2.

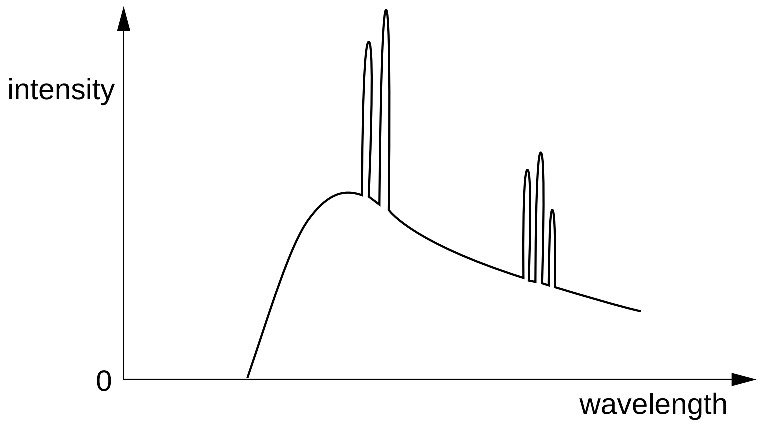


Fig. 1.2

These peaks correspond to the emission line spectrum of the atom of the target. The electrons which bombard the target excite atomic electrons in the lower energy levels of the atoms and subsequent de-excitation results in the line spectrum.

1 (d) *Candidates should be able to show an understanding of the use of X-rays in imaging internal body structures, including a simple analysis of the causes of the sharpness and contrast in X-ray imaging.*

X-rays affect photographic plates in much the same way as visible light. A photographic plate, once exposed, will appear blackened on development. The degree of blackening is dependent on the total X-ray exposure. X-rays also produce fluorescence in certain materials. The mechanism is similar to that by which visible light is produced on the screen of a cathode-ray oscilloscope.

X-rays are used to obtain ‘shadow’ pictures of the inside of the body to assist doctors with the diagnosis or treatment of illness. If a picture is required of the bones, this is relatively simple as the absorption properties of bone are about twice as great as those of the surrounding muscles and tissues. X-ray pictures of other parts of the body can be obtained if there is a difference between the absorption properties of the organ under review and the surrounding tissues.

The quality of the shadow picture (the image) produced on a photographic plate depends on its sharpness and contrast. Sharpness is concerned with the ease with which the edges of structures can be determined. A sharp image implies that organs etc are outlined clearly. An image may be sharp but, unless there is a marked difference in the degree of blackening of the image between one organ and another (or different parts of the same organ), the information that can be gained is limited. An X-ray plate with a wide range of exposures, having areas showing little or no exposure as well as areas of heavy blackening, is said to have good contrast.

In order to achieve a sufficiently sharp photographic image, the X-ray tube is designed to generate as narrow a beam of X-rays as possible. Factors in the design of the X-ray apparatus that may affect sharpness are:

(i) the area of the target anode in the tube (see Fig. 1.3),

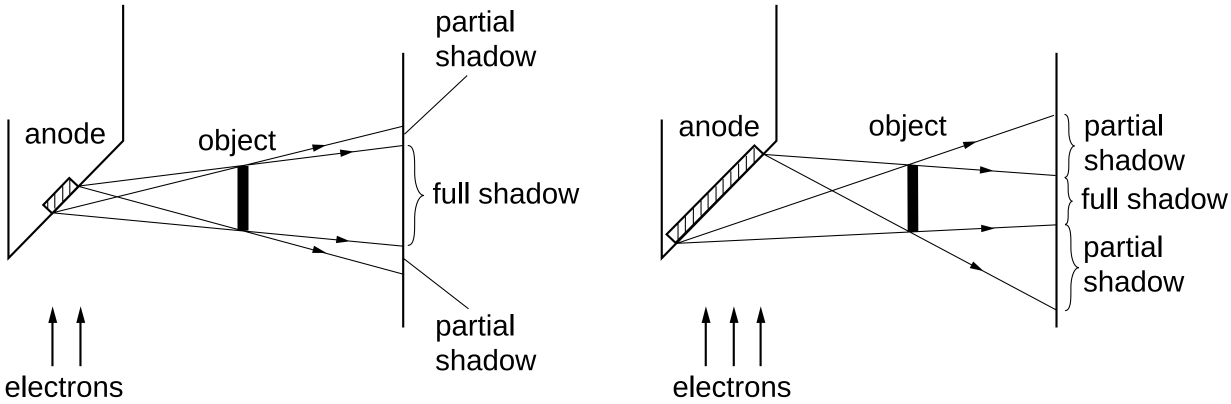


Fig. 1.3

(ii) the size of the aperture, produced by overlapping metal sheets (see Fig. 1.4), through which the X-rays are emitted from the tube,

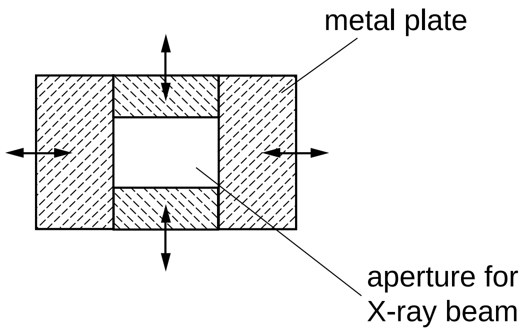


Fig. 1.4

(iii) the use of a lead grid in front of the photographic film to absorb scattered X-rays, as illustrated in Fig. 1.5.

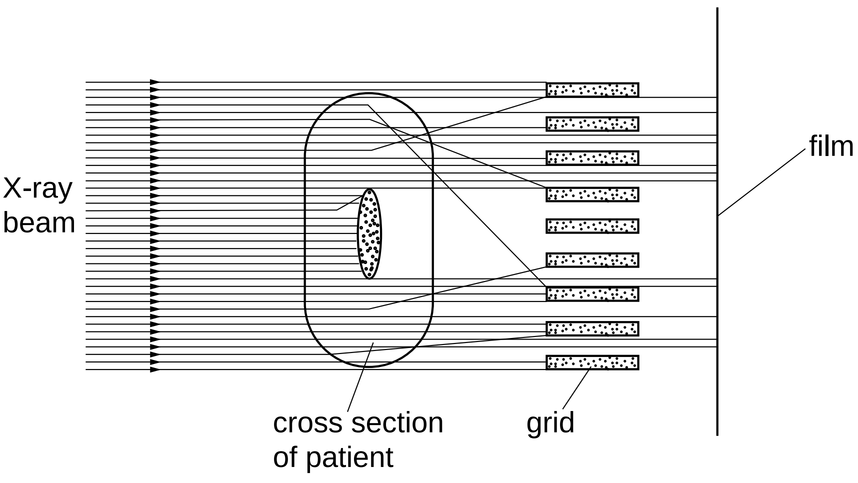


Fig. 1.5

In order to provide contrast, a barium sulphate meal is sometimes taken when observing the digestive tract or a 'contrast' medium is injected into an organ before exposure to X-rays. In these circumstances, the barium sulphate or 'contrast' medium strongly absorbs the X-radiation.

The contrast of the image produced on the photographic plate is affected by a number of other factors, including exposure time, penetration of the X-rays and scattering of the X-ray beam within the patient's body. Contrast may be improved by backing the photographic film with a fluorescent material.

1 (e) *Candidates should be able to recall and solve problems using the equation $I = I_0 e^{-\mu x}$ for the attenuation of X-rays in matter.*

When the energy of X-rays radiates from the source in all directions in a vacuum, the intensity decreases in proportion to the square of the distance from the source. This is a simple geometrical consequence (the surface area of a sphere is $4\pi r^2$). Thus, intensity $I \propto \frac{1}{r^2}$ in a vacuum. The law also applies to γ-rays and β-radiation and may also be applied, with little error, to attenuation of X- and γ-rays in air.

However, in a medium, where absorption processes are occurring, the intensity I of a parallel beam is considered to fall by a small constant fraction each time the radiation travels through equal small distances. This is effectively an exponential decrease in the transmission of the radiation. If the thickness of the absorber is x then

$$\frac{I}{I_0} = e^{-\mu x},$$

where μ is a constant (dependent on the medium and also on the photon energy of the X-rays) and I_0 is the incident intensity.

The penetrating power, or quality, of the X-radiation may be described in terms of the thickness of material needed to reduce the transmitted intensity to half its former value. This is called the *half-value thickness* $x_{\frac{1}{2}}$ (HVT), and is illustrated in Fig. 1.6.

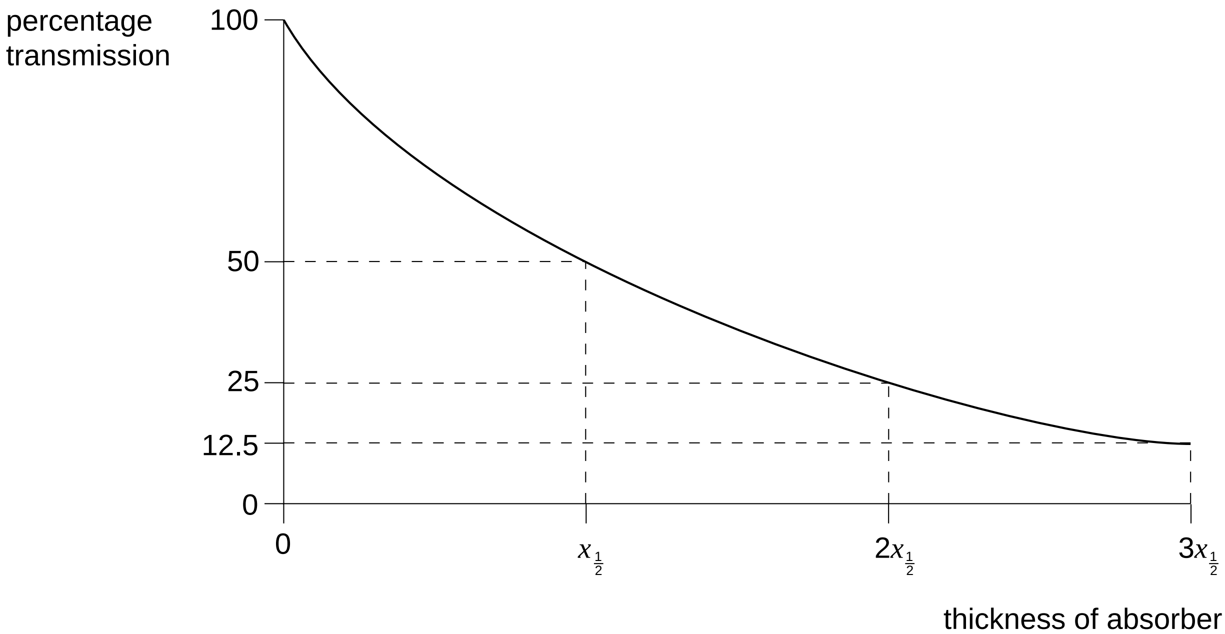


Fig. 1.6

For the half-value thickness $x_{\frac{1}{2}}$, $I = \dfrac{I_0}{2}$. Thus,

$$I = \frac{I_0}{2} e^{-\mu x_{\frac{1}{2}}}$$

$$2 = e^{\mu x_{\frac{1}{2}}}$$

$$\ln 2 = \mu x_{\frac{1}{2}}$$

In practice, HVT is not precise as the relationship is strictly valid only for a narrow homogeneous X-ray beam.

1 (f) *Candidates should be able to explain the principles of generation of ultrasonic waves using piezo-electric transducers.*

Ultrasonic waves are produced by a piezo-electric transducer. The basis of this is a piezo-electric crystal such as quartz, coated on two sides with thin layers of silver, as illustrated in Fig. 1.7.

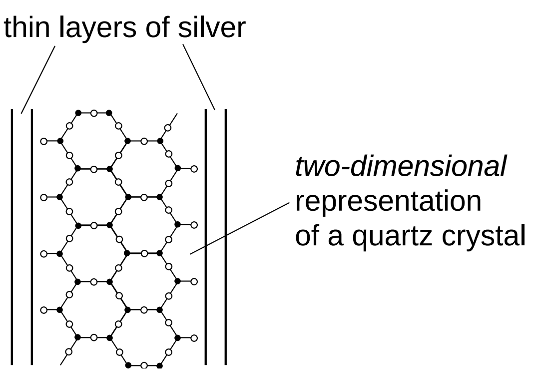


Fig. 1.7

Quartz has a complex structure made up of a large number of repeating tetrahedral silicate units, as illustrated in Fig. 1.8.

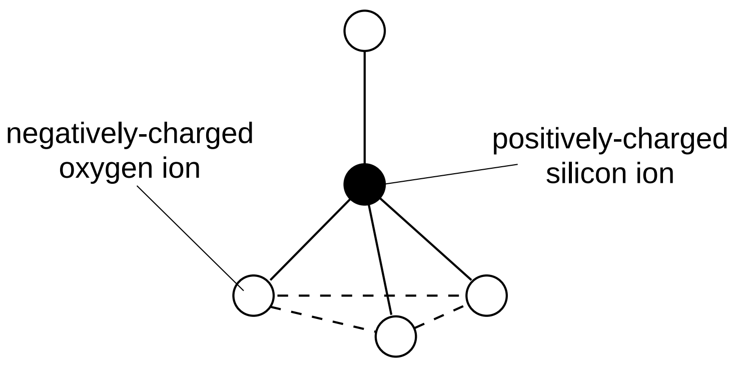
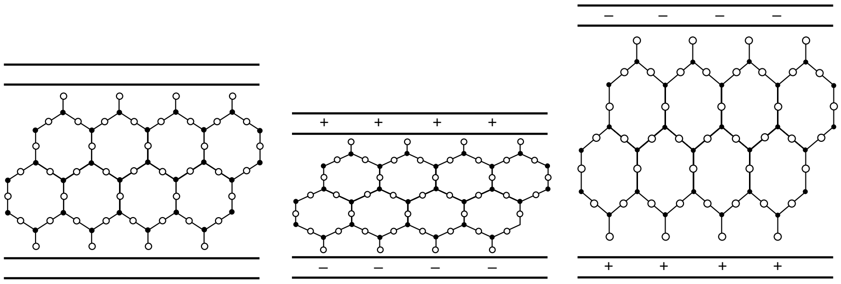


Fig. 1.8

The positions of the oxygen links are not rigidly fixed in these units and since the oxygen ions are negatively charged, movement can be encouraged by applying an electric field.

When the crystal is unstressed, the centres of charge of the positive and negative ions bound in the lattice of the piezo-electric crystal coincide, so their effects are neutralised as in Fig. 1.9(a).



(a) unstressed (b) compressed (c) extended

Fig. 1.9

If a constant voltage is then applied across the electrodes (i.e. the layers of silver), the positive silicon ions are attracted to the cathode and the negative oxygen ions to the anode, causing distortion of the silicate units. Depending on the polarity of this applied voltage, the crystal becomes either thinner or thicker due to the altered charge distribution. These effects are illustrated in Fig. 1.9(b) and Fig. 1.9(c).

An alternating voltage applied across the electrodes sets up mechanical vibrations in the crystal. Should the frequency of the applied voltage be the same as the natural frequency of the crystal, resonance occurs, i.e. the oscillations have maximum amplitude. The dimensions of the crystal can be such that the oscillations produced are in the ultrasonic range (i.e. greater than 20 kHz), thus causing ultrasonic waves in the surrounding medium.

Ultrasonic transducers can also be used as receivers since, when an ultrasonic wave strikes an unstressed crystal, it alters the positions of positive and negative ions within the crystal. This induces opposing charges on the electrodes, producing a potential difference between them.

A simplified diagram of a typical piezo-electric transducer/receiver is shown in Fig. 1.10.

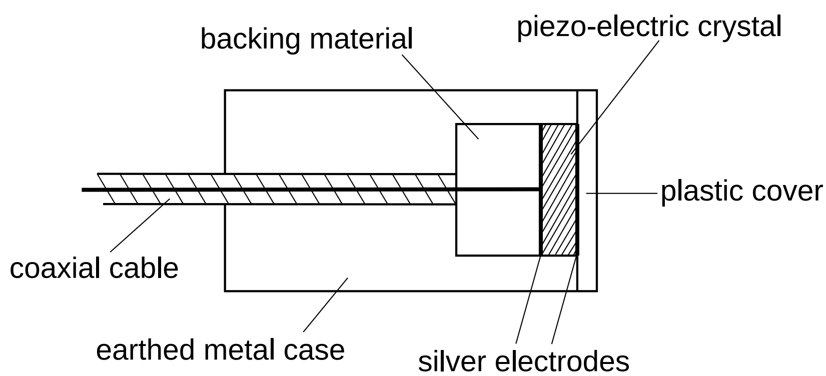


Fig. 1.10

Such devices operate in the MHz frequency range up to a maximum of around 600 MHz.

1 (g) *Candidates should be able to identify and explain the main ideas behind the use of ultrasound to obtain diagnostic information about internal structures.*

Ultrasound obeys the same laws of reflection and refraction as audible sound and light waves. When an ultrasonic wave meets a boundary between two media, some of the wave energy is reflected and some is refracted, as illustrated in Fig. 1.11.

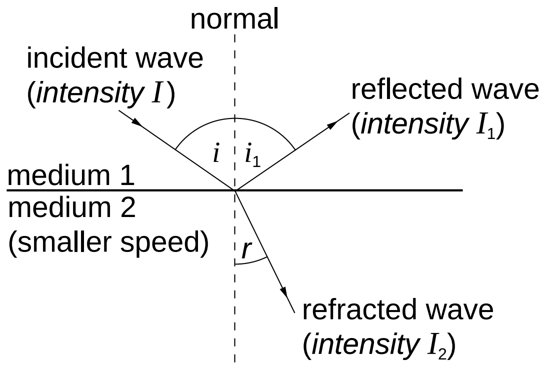


Fig. 1.11

Energy considerations make it possible to state that I (the input intensity) is equal to the sum of the reflected and refracted intensities I_1 and I_2, i.e. $I = I_1 + I_2$. The relative magnitudes of the reflected and refracted intensities depend on a property of each of the media known as the specific acoustic impedance Z. This is the ratio of acoustic pressure to particle velocity but may be taken to be approximately equal to the product of the medium's density ρ and the speed c of the wave through the medium. That is,

$$Z = \rho c.$$

If the ultrasonic waves were to try to pass from one medium into another of much larger impedance, the waves would have difficulty in making the molecules in the denser material vibrate, so most of the wave energy would be reflected and little would be transmitted. Alternatively, if the two media were of similar impedance, the majority of the wave energy would be transmitted with very little reflection occurring. In medical terms, a good example of two media with greatly differing impedances are human tissue and bone, where

$$\frac{Z_{\text{tissue}}}{Z_{\text{bone}}} \simeq \frac{1}{4}$$

Ultrasonic waves striking a tissue-bone boundary from the tissue would be mainly reflected. However, tissue and water have similar impedances, so most ultrasonic waves would be transmitted straight through a tissue-water boundary.

Another factor which affects the intensity of ultrasonic waves passing through a medium is absorption. As the waves travel through a medium, the intensity falls exponentially, because energy is absorbed by the medium. This causes the temperature of the medium to increase. The intensity I of a parallel beam at a particular point is given by

$$\frac{I}{I_0} = e^{-kx},$$

where x is the distance travelled in the medium by the wave, I_0 is the original intensity and k is an absorption constant (see also section *1(e)*).

Diagnostic information about internal structures of the body can be obtained by monitoring the pulses reflected back when short bursts of ultrasound have been transmitted into the body. Two techniques, A-scan and B-scan, are in common use. However, initially, there is a major problem in actually transmitting ultrasound into the human body. This is because the impedances of air and skin are vastly different and this causes the majority of ultrasound wave energy incident on the air-skin boundary to be reflected without entering the body. One method originally used to overcome this problem is to immerse a portion of the body and the transducer in water. Then, since the impedances of water and skin are similar, reflection is minimal and most of the wave energy is transmitted. Nowadays, a material called a coupling medium is inserted between ultrasonic transducer and skin to reduce the size of the impedance change between boundaries. At present, a film of oil or water-based cellulose jelly is used as a coupling medium for general use.

The A-scan system basically measures the distance of different boundaries from the transducer, with the transducer kept in one position. A short burst of ultrasound is transmitted to the body through the coupling medium. At each boundary between different media in the body, some ultrasound is reflected and this reflected pulse is picked up by the transducer which now acts as a receiver. The signal from the transducer is amplified and displayed on a cathode-ray oscilloscope (c.r.o.). The reflected ultrasound also meets boundaries on its return trip to the transducer, causing some energy to be reflected and some transmitted. As one consequence of this, echoes from boundaries deeper in the body tend to be weaker. To compensate for this loss in energy, the later the echo is received at the transducer, the more it is amplified before display on the c.r.o. A line appears on the c.r.o. screen each time an echo returns to the receiver. The time-base on the X-plates is adjusted so that a complete picture of the reflections is received on one scan. Then, by knowing the speed of transmission of the ultrasound, this time-base display can be calibrated as distance. A typical A-scan for the brain is shown in Fig. 1.12.

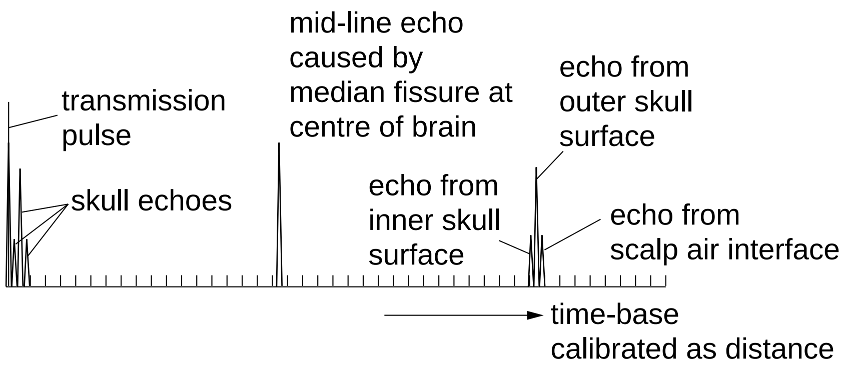


Fig. 1.12

The B-scan technique basically combines a series of A-scans, taken over a range of angles, to form a two-dimensional picture. As before, each A-scan corresponds to a single ultrasound pulse being emitted by the transducer and producing a series of reflected pulses from boundaries within the body. The ultrasound probe consists either of one crystal which can be moved rapidly through an arc or a series of crystals, each having a different orientation. The signals received from these probes are processed by a computer and displayed on a c.r.o. To make the combinations of these A-scans easy to interpret, the pulses are displayed as bright spots in the correct orientation of the A-scan. Consequently, the completed pattern of spots on the screen gives a two-dimensional representation of the boundary positions in the body being scanned, which can then be photographed to give a permanent record. Problems can arise trying to move the transducer over the surface of the body without introducing air between it and the skin. Coating the body with oil-based jelly has proved successful in overcoming these difficulties.

The main advantage of ultrasonic scanning is that the health risk factor to human patients, and to those operating the system, is considered to be much less than in X-ray diagnosis. Other advantages are that the equipment is portable and relatively simple to use. Ultrasound systems have now been devised which can pick up and amplify the very small echoes produced by soft tissue boundaries within the body as well as those between hard and soft tissues.

The information obtainable by the use of ultrasonic scanning is potentially far greater than for X-rays.

1 (h) *Candidates should be able to identify and explain the main ideas behind the use of magnetic resonance to obtain diagnostic information about internal structures.*

Many atomic nuclei behave as if they possess a ‘spin’. Such nuclei have an odd number of protons and/or an odd number of neutrons. Their ‘spin’ causes such atoms to behave as tiny magnets (the moving charge in the nucleus creates a magnetic field). If an external magnetic field is applied to the atoms, they will line up in the field. However, this alignment is not perfect and the nuclei rotate about the direction of the magnetic field. This type of motion is known as *precession* and the frequency of precession (the Larmor frequency) depends on the nature of the nucleus and the strength of the magnetic field. The Larmor frequency is found to lie in the radio-frequency (RF) region of the electromagnetic spectrum.

If a short pulse of radio waves of frequency equal to that of the Larmor frequency is applied, the atoms will resonate and, when the pulse ends, the atoms will return to their equilibrium state. In this process, RF radiation is emitted. The time taken between the end of a pulse and the emission of the RF radiation forms the basis of magnetic resonance imaging (MRI).

A schematic diagram of an MR scanner is shown in Fig. 1.13.

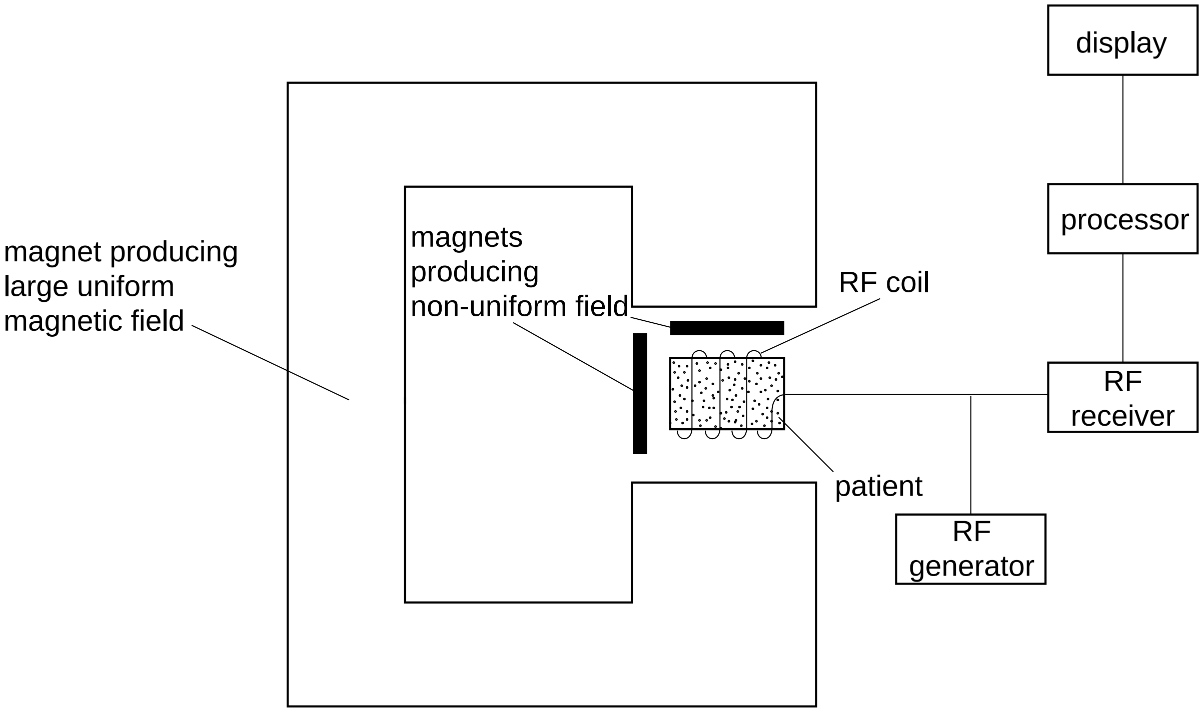


Fig. 1.13

The subject of the investigation is placed between the poles of a very large magnet which produces a uniform magnetic field in excess of one tesla. Since hydrogen atoms are in great abundance in body tissue, hydrogen nuclei are used for imaging. All these nuclei within the uniform magnetic field would have the same Larmor frequency. In order to locate the position within the patient of particular hydrogen atoms, a non-uniform (but accurately calibrated) magnetic field is also applied. This provides a unique value for the magnetic field at any point and, coupled with the particular value of the Larmor frequency, enables the hydrogen nuclei to be located.

RF pulses are transmitted to the patient by means of suitable coils. These coils are also used to detect the RF emissions from the patient. The received signals are processed in order to construct an image. One such MR scan, which shows a section through the spine and the back muscles, is shown in Fig. 1.14.

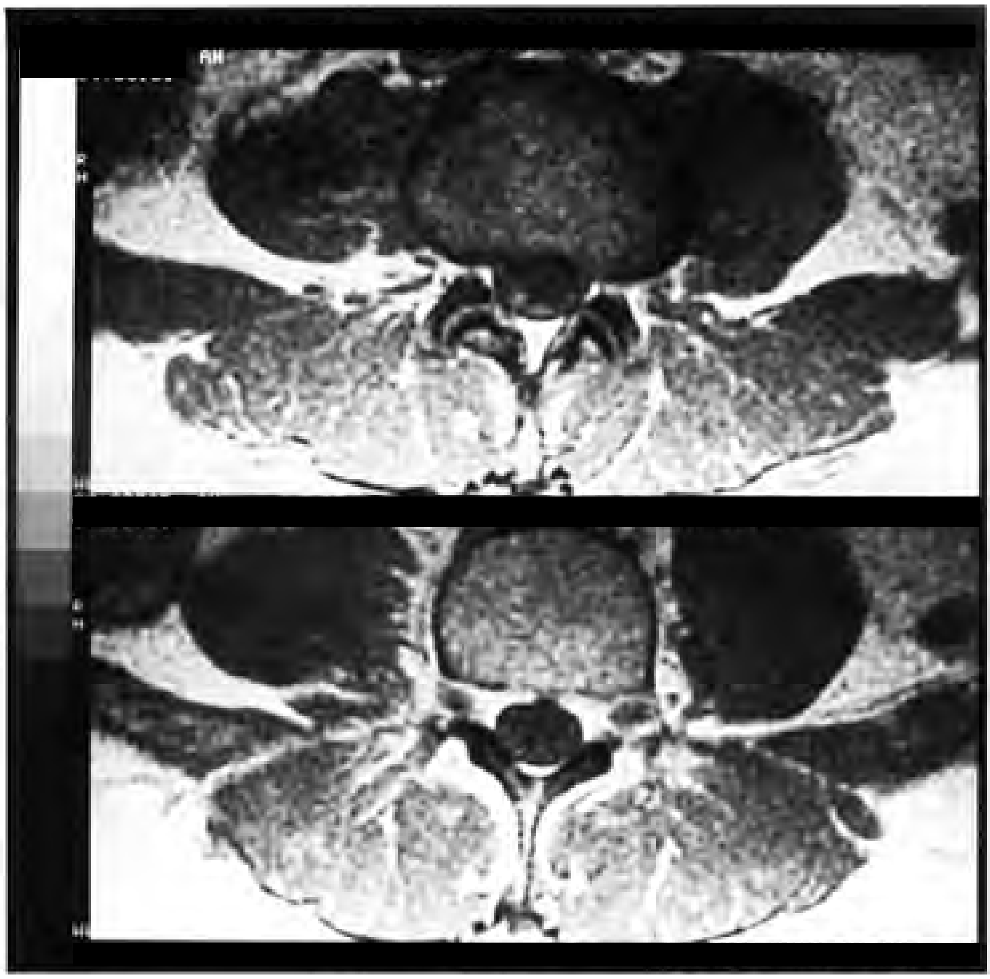


Fig. 1.14

Modern developments have allowed for further refinement of the information available so that not only can detailed images of body structures be obtained but also "image flow". Images showing only arteries and veins have been created.

1 (i) *Candidates should be able to identify and explain the main ideas behind the use of lasers in diagnosis, e.g. in pulse oximetry and in endoscopes.*

Magnetic resonance imaging can be used for the brain but, in the special case of new-born babies, the use of magnetic resonance can cause problems. It requires the child to remain very still during the technically complex and long assessment procedures. The outcome is likely to provide only intermittent and not continuous information on brain metabolism. Also, magnetic resonance does not measure directly the supply of oxygen to the child's brain tissues.

However, tissues change colour as their oxygenation and blood supply changes, and thus provides the basis for *pulse oximetry*. Colour change is measured in the infra-red region because tissue is much more transparent to radiation in this region as compared with the visible region.

Semi-conductor laser-light sources together with light detectors (see also section *2(e)*) are used, and it is possible to have a portable instrument which measures colour changes across the width of a baby's head, some 10 cm in diameter. In practice, the only physical contacts with the head are optical fibres. These allow a pulse of laser light to enter the scalp. Each pulse is about 4 ps duration. Detectors then pick up the light scattered from skin tissue and the results produce quantifiable measurements on the colour changes observed. With this technique, it is possible to measure quantities such as the volume of blood in the brain, the rate of flow of blood and the total oxygen delivery. With the use of further analysis, the availability of oxygen inside cells may be determined.

The technique is being extended to the study of oxygenation of the foetus's brain during labour, the brain's oxygen supply during cardiac by-pass surgery in both children and adults, and to the monitoring of the prevention and treatment of brain damage.

An endoscope is an instrument designed to provide visual images of internal body structures and to enable small-scale surgical procedures to be undertaken. The endoscope consists of a flexible tube which can be inserted into the body. The tube contains two groups of optical fibres – one by which internal body structures may be illuminated, the other by which these structures may be viewed or photographed. The tube also allows for the passage of water, gases and small surgical instruments. A schematic diagram of an endoscope is shown in Fig. 1.15.

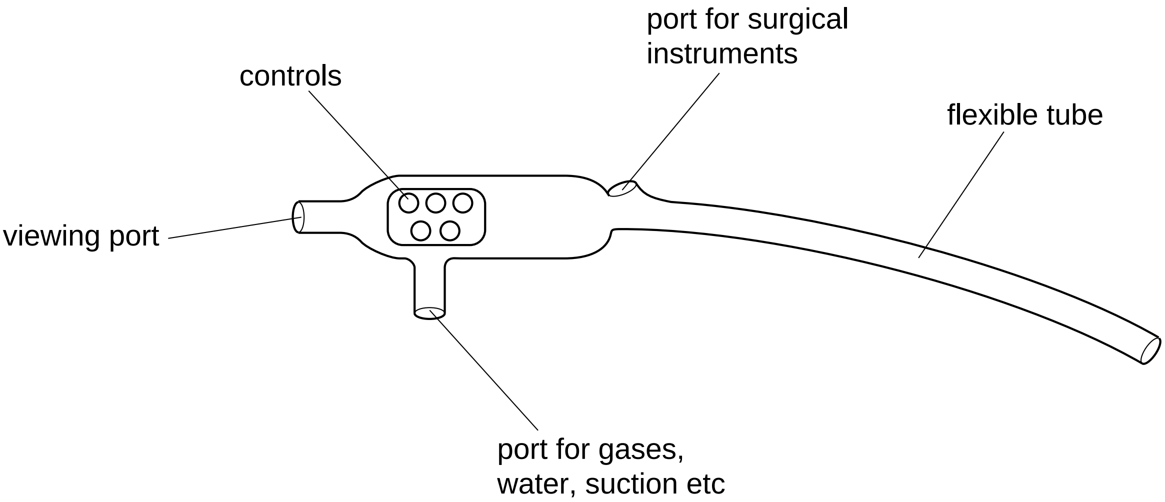


Fig. 1.15

The light source for the endoscope is a laser. The laser light is transmitted along the bundle of optical fibres and illuminates the site under investigation. The endoscope is able to probe many passages within the body including the gastrointestinal tract and has been used to diagnose problems such as ulcers, cancers, constrictions and internal bleeding.

1 (j) *Candidates should be able to describe examples of the use of radioactive tracers in diagnosis.*

Any radioactive nuclide which is to be used as a tracer for diagnostic purposes must be incorporated into a pharmaceutical product which can be administered to the patient. The pharmaceutical product, and the radioactive nuclide which it incorporates, should have the following properties.

(i) It should concentrate in the organ or system which it is required to image or assess.

(ii) It should not change the functioning of the organ, i.e. the pharmaceutical behaves in the same way as the substance it is replacing.

(iii) The radioactive half-life should be as short as possible to avoid radiation damage but it must be sufficiently long to enable measurements to be made.

(iv) The type of radiation emitted and its energy should be such that detection is possible.

Some examples of the use of radioactive tracers are listed below.

(i) Thrombosis (blood clots). A radioactive tracer, iodinated fibrinogen, is introduced into the blood. If there is a blood clot forming in some part of the body, a concentration of radioactive fibrinogen will occur at that point. The concentration can be located using a radiation detector.

(ii) Assessment of thyroid function. The thyroid gland is a ductless gland in the neck. Iodine, which is absorbed readily into the blood stream, is collected by the thyroid gland for use in making thyroid hormones. Both over-active and under-active thyroids, if untreated, can lead to medical problems. To investigate the behaviour of the thyroid, tests are carried out to measure the accumulation of radioactive ^{131}I by the gland. The patient is given a tracer dose of a dilute sodium iodide solution and the count rate from the thyroid is taken over a period of time and compared with that of a normal thyroid. This is illustrated in Fig. 1.16.

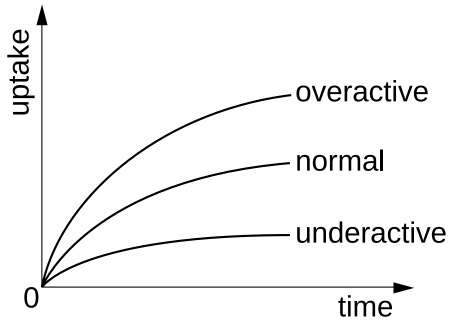


Fig. 1.16

(iii) Radiocardiography. A suitable radiation detector is positioned above the heart before a tracer is injected into the blood stream. A typical radiocardiograph is shown in Fig. 1.17.

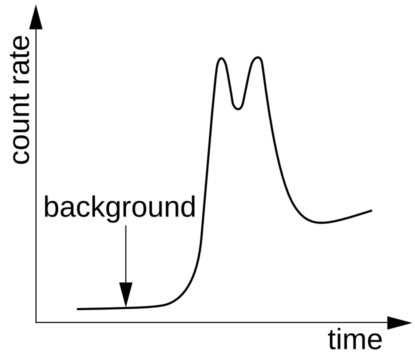


Fig. 1.17

The two peaks correspond to the passage of the tracer through the right and left chambers of the heart, and the dip is due to the passage of the tracer through the lungs.

More recently, $^{99}Tc^{m}$ (the metastable isotope of technetium) has been used instead of iodine. The patient is given sodium pertechnate which is taken up in a similar way to iodine but is more easily released from the body. In addition, the half-life of $^{99}Tc^{m}$ is much shorter – ^{131}I is 8 days but $^{99}Tc^{m}$ is only 6 hours. The person thus remains radioactive for a much shorter period of time.

M2. Medical Treatment

2 (a) Candidates should be able to explain in simple terms the effects of ionising radiation on living matter.

Much experimental work has been conducted into the nature of the damage to cells caused by ionising radiation. The nucleus of the cell has been shown to be the part of the cell most vulnerable to radiation damage and, within the nucleus, the DNA molecules. Damage to the DNA not only affects the functioning of the cell itself but also its ability to divide.

Damage to DNA may be termed “direct” or “indirect”. Ionisation causes direct damage to DNA, in which case the whole molecule may either be broken into fragments or have sections removed. Other molecules may also be ionised. In particular, water molecules are in great abundance and consequently the production of hydrogen (H^+), and hydroxyl (OH^-), ions is common. Indirect damage occurs when OH^- ions react with DNA molecules.

If the damage within the cell is great, then cell death occurs immediately or within a very short time. It is possible for a cell to be damaged only slightly so that its function is not impaired too badly. Even at low doses, the DNA in the cell nucleus may have been damaged so badly that the cell is unable to replicate itself by cell division. This may have serious long term consequences for the organism. At yet lower damage levels, cellular reproduction may continue for many cell generations, giving rise to delayed effects such as enhanced risk of developing cancer. Thus, rapidly multiplying cells are far more susceptible to damage and, in consequence, special attention is paid in the use of ionising radiation on unborn children and young persons. The reproductive system may also be affected, giving rise to an increased risk of genetic mutations. If the damage is severe enough, it can cause sterility. This effect has been used beneficially by releasing sterile pest insects to prevent their breeding and multiplying. It is also possible to irradiate food so as to kill bacteria in it and thus prolong its shelf-life.

The total dose of radiation determines the amount of ionisation per unit volume of tissue. Consequently, radiation damage increases as the total dose increases. Furthermore, for the same dose, the density of ionisation varies according to the nature of the radiation. A greater density of ionisation is likely to give rise to more damage. Thus, for the same dose, α-radiation is potentially more dangerous than β-radiation or γ-rays.

Dose rate is also of importance. Some damage to DNA may be irreparable. However, mechanisms do exist by which damage to DNA molecules may be repaired. If the radiation dose is given over a long period of time, the repair mechanism has more time to act than when the same dose is given over a short period and thus, in the former case, the damage to the organism is likely to be less.

Location of the damage depends on the type of radiation. α- and β-radiation are unable to penetrate far through organic matter so external radiation sources tend to produce severe damage near the skin. γ-ray photons pass easily through the human body so they produce less severe ionisation but are capable of causing direct damage to internal parts of the body. The severity of the biological effects depends amongst other factors on the energy absorbed.

One of the most dangerous possibilities is that radioactive material may be ingested. For example, ^{131}I produced in fission reactions (nuclear bomb tests or reactor accidents) may get into the food chain and be absorbed with food. Once in the body, it is possible for radioactive material to become concentrated in a particular part (iodine builds up in the thyroid gland) and this may increase the activity in that region. To reduce the problem at a time of crisis, iodine tablets are taken so that the body already has an excess of iodine, increasing the likelihood of excretion of radioactive iodine.

A different example of ingestion is radon gas. This gas is released from rocks in minute quantities but it may accumulate in still air and then be breathed into the lungs and absorbed.

The severe medical effects of high dose and high dose rates are well understood. The long-term effects of very low doses and very low dose rates are not fully understood and so remain highly controversial.

Radiation can be used beneficially in medicine by being used to destroy malignant cells. Different types of radiation are used to treat different types of cancers. The patient often suffers side-effects known as “radiation sickness” which may make the patient very ill.

2 (b) *Candidates should be able to show a qualitative understanding of the importance of limiting exposure to ionising radiation.*

The long-term medical and genetic effects of low radiation levels are not properly understood so it is important to keep radiation doses as small as possible. On the other hand, we are surrounded by natural radiation all the time (cosmic radiation, radiation from rocks, etc), so adding a small fraction more should not alter the risk appreciably. In some circumstances there are clear benefits to radiation, for example, in having a fractured limb X-rayed, or in having an internal problem diagnosed, or in having radiation therapy. In these cases, the benefit to the patient will outweigh the increased risk significantly. Radiation safety levels are set higher for people who work regularly with radiation than for the general public, but these people must always be especially careful to avoid unnecessary exposure. Some people are considered to be especially at risk, such as young children who are still growing quickly, and also expectant mothers. In these cases, cell division is rapid, so risks from radiation are especially high.

Natural background radiation comes mainly from two sources – firstly from outer space (mostly gamma-radiation) but also from minerals in the Earth. The latter varies widely across the country because although some minerals are almost free of radiation (for example limestone and chalk), others are appreciably radioactive (particular volcanic rocks such as granite and basalt). People who live in regions where granite is common may receive two or three times the annual radiation dose compared with those who live on non-radioactive land.

Radiation from artificial sources should ideally be kept well below the natural background level. This includes medical and dental radiation, the nuclear industry, plus fall-out from nuclear weapons and also radioactivity ingested with air and food. Unnecessary radiation exposure should be avoided as, for example, the old fashioned practice of X-raying feet inside shoes to check the fitting. Rutherford advised his students to treat radiation 'like poison' and it is certain that excessive exposure to radiation shortened Marie Curie's life.

When working with radioactive sources, written precautions which come with the sources should be followed. The source should be kept well away from humans, suitable shielding should be arranged – dependent on the type of radiation – and a source handling tool at arm's length should be used, if the source has to be moved. Sources should never be removed from the laboratory. The safety regulations, provided by the person responsible for the laboratory, must always be observed.

2 (c) *Candidates should be able to distinguish between dose rate and dose, paying particular attention to the type of incident radiation.*

Total radiation dosage of a sample is assessed by calculating the ionising energy delivered into the sample by the radiation. The absorbed dose D is the ratio of the total radiation energy absorbed to the mass of the sample. The SI unit is the gray (1 Gy = 1 joule of energy deposited for each kilogram of the sample).

Given the same absorbed dose, some types of radiation are more damaging than others as regards biological effects. This is due to a greater concentration of ionisation having a larger effect. This is allowed for by assigning a 'quality factor' to each type of radiation. To find a measure of biological damage, the absorbed dose is multiplied by the *quality factor*. The SI unit for this modified dosage is the sievert (Sv), which is a large dose to have received. A more useful unit is the milli-sievert, mSv. Some quality factors are as follows:

X-rays, γ-rays, β-particles	1 or 2
slow neutrons	5
fast neutrons, protons, singly charged particles	10
α-particles, multiply charged particles	20

In order to calculate the dose rate near a radioactive source, account should be taken of how effective the sample is at absorbing radiation. This depends on the type of radiation, so an α- or β-source could be more dangerous than a γ-source of the same total activity. It also depends on the composition of the sample. Air in the lungs does not absorb strongly whereas water and soft body tissue are more absorbent and bone is particularly absorbent. Dense metal such as steel (artificial joints) and lead are extremely absorbent. A final complicating factor is that some parts of the body are much more susceptible to radiation damage than others. In cases where radiation is concentrated on to particular organs, a risk factor has been assessed and the dose in mSv can be multiplied by this risk factor.

The absorbed dose rate is simply the rate at which radiation energy is absorbed per unit mass of tissue. The rate is measured in gray per second (Gy s^{-1}).

2 (d) *Candidates should be able to explain the use of X-rays and of implanted sources in the treatment of malignancy.*

All cancers have one thing in common. They are caused by the uncontrolled division of cells. Most cells in the body divide into two, from time to time, in order to replace cells that have died. A cancer cell is one that divides too frequently, each of the new cells produced by the division having the same fault. As a consequence, and in time, a high concentration of faulty cells (a tumour) develops. Not all tumours, however, are malignant. Malignant tumours grow more rapidly, invading surrounding tissues, and affected cells break away from the tumour and are carried to other parts of the body, becoming the nucleus of a secondary tumour.

Radiotherapy is the treatment of disease, especially cancer, by ionising radiations such as X-rays, γ-rays or electron beams. Radiation destroys normal cells but cells which are dividing rapidly are far more susceptible.

In radiation treatment, when the position of the tumour has been found accurately, a radiation beam is directed on to the tumour. If only one beam were to be used, as shown in Fig. 2.1(a), considerable damage would be caused to the skin and other tissues between the skin and the tumour.

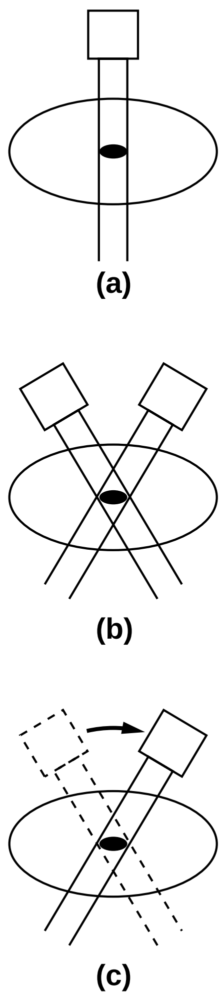


Fig. 2.1

In order to prevent this, either several beams are directed on to the tumour (Fig. 2.1(b)), or a single radiation beam is rotated around the patient so that the beam always passes through the tumour site (Fig. 2.1(c)) but other tissues receive a smaller dose. In these ways, the beam can be of relatively low intensity but its cumulative effect in the tumour is far greater than in the surrounding tissue.

The method of implanting radioactive sources directly into tumours is another means of treatment. ^{137}Cs has been found to be very effective in the treatment of cervical and uterine cancers. Similarly, in the case of cancer of the thyroid, ^{131}I can be used, since iodine will accumulate in the thyroid gland as a natural course of events. Skin cancers may be treated directly by attaching a radioactive source to the affected area.

2 (e) *Candidates should be able to describe examples of the use of lasers in clinical therapy, e.g. as a scalpel or as a coagulator.*

Lasers – light amplification by stimulated emission of radiation – were first made from synthetic ruby in 1960. Essentially, a laser is a source of light which may be monochromatic and is highly collimated. Its most useful feature here is its intensity, i.e. its power per unit area, because the beam is highly coherent. Laser beams have little divergence. As a consequence, even though a laser power source may be much lower than, say, a fluorescent lamp, its power per unit area (intensity) can remain high even at considerable distances.

The laser beam may also be focussed on to a very small spot. The minimum diameter of the focussed beam is a function of the angle of divergence of the laser and the focal length of the lens. The ultimate spot size is determined by the wavelength of the light – i.e. the spot is limited by diffraction effects.

Spot diameters of the order of a few hundredths of a millimetre can be obtained. The intensity at the focus may be some 10^{12} times greater than that in an oxy-acetylene flame. Such light focussing can be used for internal cauterisation of tissue. Thus the laser acts as a surgeon’s knife, but with the added advantage of no or minimal bleeding and the possibility of very accurate guidance.

A particular example is the use of the carbon dioxide (CO_2) laser. Its infra-red light is strongly absorbed by water and the power density can be so high that there arises explosive vaporisation of intracellular water. This is the effective scalpel. Another more penetrating and powerful laser (the Nd:YAG – yttrium aluminium garnet doped with neodymium) heats a greater volume of tissue more gradually, i.e. coagulation occurs. This is also the situation where an argon laser is used to repair small tears in the eye’s retina. Photocoagulation takes place and this is often referred to as “spot welding”.

The monochrome nature of lasers is used in another example – the argon-ion laser – where the blue-green light is strongly absorbed by red blood but weakly absorbed by the dermis and epidermis, i.e. skin layers. The thermal energy produced may raise the temperature of capillary blood much higher than that of the dermis and epidermis. Thus port wine stains – due to abnormal capillary blood vessels – may be destroyed with minimal damage to the surroundings.

Finally, very powerful pulses may be created by a switching technique, e.g. 10^9 W in an optical pulse of 10^{-8} s. When these are conveyed by a suitable fibre optic system to an internal organ, the destructive shock waves arising are sufficient to fragment bladder- and gall-stones.

Lasers are hazardous. Extreme caution is necessary in their operation. Clinical lasers exceed by some three orders of magnitude the critical power density necessary to scar the cornea permanently. The eye never recovers. Protective eyewear is essential and, in addition, lasers with endoscopes have filters and shutters in the path line to the operator’s eyepiece. Skin may also be burned and there is a potential fire hazard. Operators should be aware of all these hazards and take all the necessary steps to minimise the risk of accidents.

M3. The Physics of Sight

Before discussing the eye and the eye's ability to form images, it is pertinent to consider lenses in general and the relation between focal length and power in particular.

3 (d) *Candidates should be able to distinguish between converging and diverging lenses and show an understanding of the significance of focal length.*

3 (g) *Candidates should be able to relate the focal length of a lens to its power in dioptres.*

A converging (convex) lens is a lens shaped so that it is thicker in the middle than at the edges. Such a lens refracts a beam of light which is parallel to its principal axis so that it converges to a point, as illustrated in Fig. 3.1.

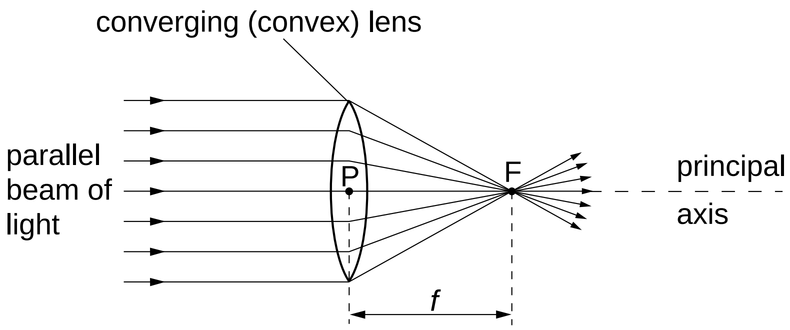


Fig. 3.1

This point of convergence is called the principal focus (focal point) F. The focal length *f* is the distance from the principal focus to the optical centre (pole) P of the lens. The diagram shows that all the rays of light in the parallel beam striking the lens actually pass through F. Thus, the image at F is called a *real* image, in that rays actually form an image which can be picked up on a screen held at that point. The image formed on a film in a camera is a practical example of a real image.

A diverging (concave) lens is shaped to be thicker around the edges of the lens than in the middle. This type of lens refracts a beam of light parallel to its principal axis so that it appears to have diverged from a point F, as illustrated in Fig. 3.2.

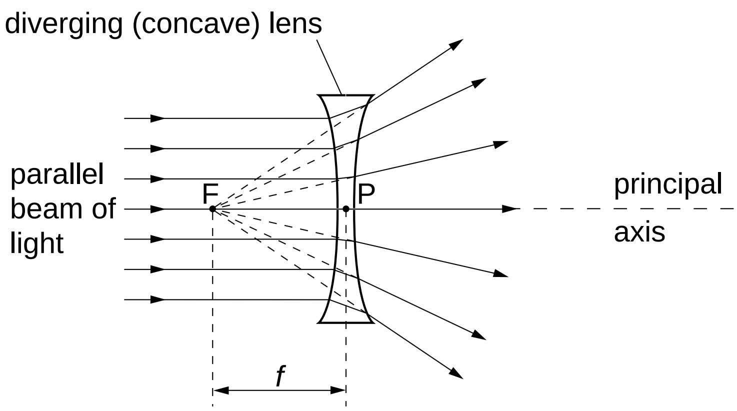


Fig. 3.2

This time, the rays do *not* actually pass through the point F. Hence a new name is required for this principal focus image from which the rays appear to have diverged. It is called a *virtual* image, implying that rays do not pass through this point, so that there would be no image formed on a screen placed at F. The image formed using a plane mirror is a useful practical example of a virtual image. The optical centre P of the lens and the focal length *f* are also shown on Fig. 3.2.

These ray diagrams are only strictly correct providing the lenses are thin, and the rays are parallel and close to the principal axis. There is an added complication that the different colours in white light refract by different amounts, leading to an effect called chromatic aberration where the edges of images show red and blue colouration.

The most fundamental property of a lens is its ability to form an image. This is what lenses are used for in cameras, projectors, telescopes and microscopes. In the camera, for example, adjustments may have to be made in one or more of the following; the object distance from the lens, image distance from the lens, or focal length of the lens, before a clearly focused image is obtained. A lens formula can be used to predict object or image distances, or appropriate lens focal lengths. The lens formula is

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} ,$$

where *f* is the focal length of the lens, (the distance FP),
u is the distance of the object from the optical centre P of the lens
and *v* is the distance of the image from the optical centre P of the lens.

These distances are illustrated on Fig. 3.3.

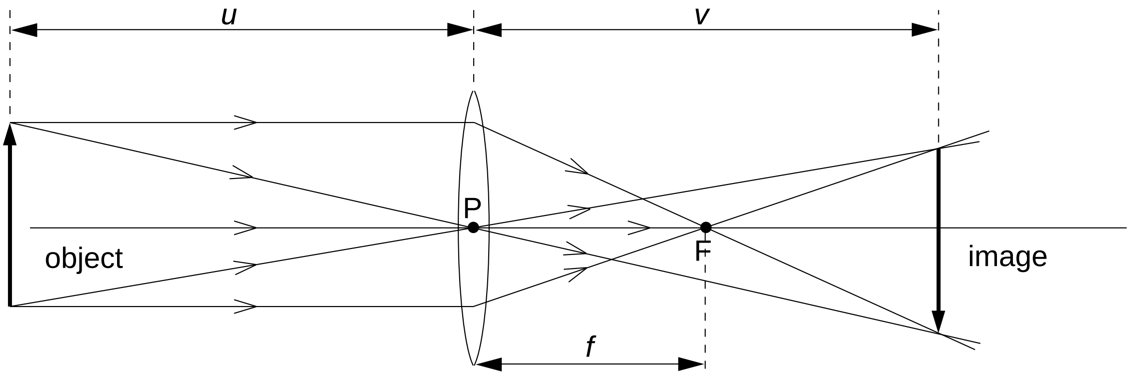


Fig. 3.3

In order to enable the same formula to be used for all types of lens, object and image, a sign convention is required. This determines which of the factors *u*, *v*, and *f* are positive and which are negative. One such convention is the *real-is-positive* convention in which

real object and real image distances are labelled positive,
virtual object and virtual image distances are labelled negative,
the focal length of a converging lens (its principal focus is real) is considered positive,
the focal length of a diverging lens (its principal focus is virtual) is considered negative.

The following worked example shows how the lens formula plus sign convention can be used to solve a lens problem.

Example

(a) An illuminated object is placed 1.20 m from a screen. A converging (convex) lens is then moved between the object and the screen until a clear image is seen on the screen. If the lens is then 0.35 m from the object, calculate the focal length of the lens.

(b) The object is now moved towards the lens until it is placed 0.20 m from the lens. Where is the new position and what is the nature of the image formed by the lens?

A suitable diagram to illustrate part (a) is given in Fig. 3.4.

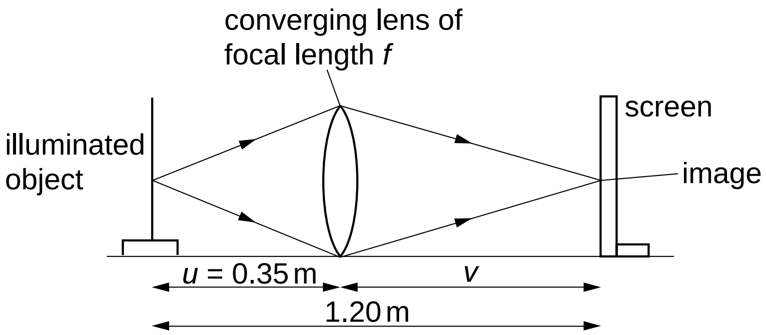


Fig. 3.4

Both object and image are real so u = +0.35 m and v = + (1.20 – 0.35) m = 0.85 m

Using the formula

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v},$$

$$\frac{1}{f} = \frac{1}{+0.35} + \frac{1}{+0.85} = 2.86 + 1.18 = +4.04$$

Therefore $f = \dfrac{1}{+4.04} = +0.25$ m. Thus the focal length is 0.25 m, the positive sign confirming that it is a converging lens.

Now, using this information in part (b), f = +0.25 m and u = +0.20 m, it is required to find the new image distance.

Using the lens formula

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v},$$

$$\frac{1}{+0.25} = \frac{1}{+0.20} + \frac{1}{v}$$

$$\frac{1}{v} = 4.0 - 5.0 = -1.0$$

$$\text{So, } v = -1.0\text{ m}$$

Thus, the image is 1.0 m from the lens. The negative sign indicates that this image is virtual.

To understand *where* this virtual image lies, it must be remembered that the object was closer to the lens than the principal focus. This leads to diverging rays leaving the lens after refraction, as shown in Fig. 3.5.

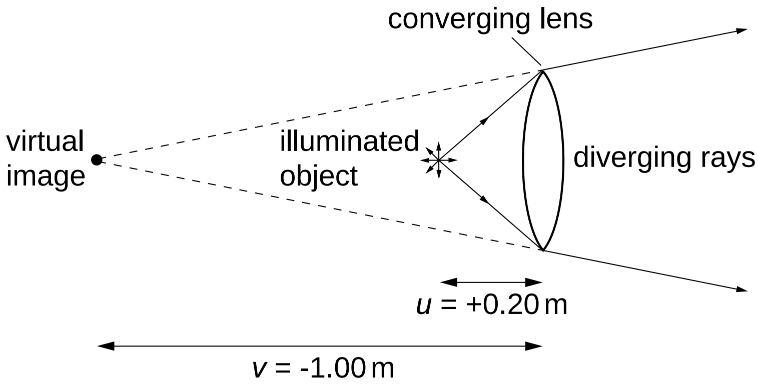


Fig. 3.5

The virtual image must be situated at the point from which these rays appear to diverge, i.e. the image is on the same side of the lens as the object.

The shorter the focal length of a lens the more converging or diverging the lens is said to be. This is indicated in Fig. 3.6.

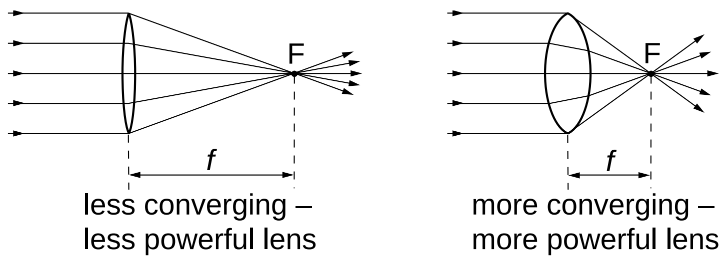


Fig. 3.6

This shows that for the more converging lens or more powerful lens, of shorter focal length, the principal focus F is nearer the lens. Parallel rays are refracted to a greater extent due to the increased curvature of the lens surfaces.

The defining equation for lens power is

$$\text{power of lens in dioptres} = \frac{1}{\text{focal length in metres}}.$$

The symbol for the unit dioptre is D and just as diverging lenses have negative focal lengths so they have negative powers. Thus,

(i) a lens of focal length +0.50 m has a power of $\frac{1}{+0.50}$ D = +2.0 D. This is a converging lens.

(ii) a lens of focal length −0.25m has a power of $\frac{1}{-0.25}$ D = −4.0 D. This is a diverging lens.

This is a further example of the operation of the real-is-positive convention.

3 *(a)* *Candidates should be able to explain how the eye forms focused images of objects at different distances.*

As a ray of light enters the eye, it passes through several media before reaching the retina. From air, the light passes through the cornea, the aqueous humour, the lens, and finally through the vitreous humour to the retina. This is illustrated in Fig. 3.7.

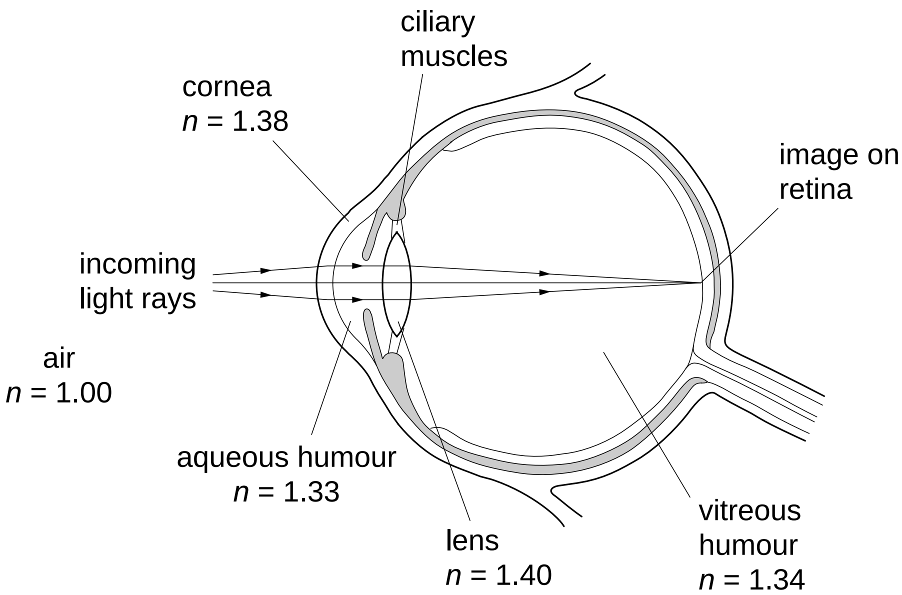


Fig. 3.7

At each boundary between two media refraction may occur, with greater deviation where the refractive index change is greater. The most significant media change occurs as light rays enter the cornea from air. Since the refractive index of air is 1.00 and that of the cornea is 1.38, the ray will be refracted quite considerably towards the normal. As the ray of light leaves the cornea, however, it will be bent only slightly away from the normal as the refractive index of the aqueous humour is close to that of the cornea. So the effect of the cornea is to converge the light considerably. This is illustrated in Fig. 3.7.

Although the lens has a higher refractive index than either the aqueous humour or the vitreous humour, the difference is not great. Hence refraction at the lens boundaries contributes only slightly to increasing the convergence of the incoming light rays. The major task of the lens is *not* to converge the light but to produce slight adjustments to the focusing. To achieve this, the lens is suspended by ligaments which are attached to a circular ring of muscles called the ciliary muscles. The ciliary muscles control the shape of the lens. When the muscles are relaxed, the lens is thinner and less refracting, but contraction of the ciliary muscles makes the lens more curved and thus more refracting and of shorter focal length.

3 (b) *Candidates should show an understanding of the terms depth of focus and accommodation.*

The lens-cornea combination is responsible for focusing rays of light on to the retina. But within this combination, it is the cornea which is responsible for most of the refraction, whilst the function of the lens is to 'fine tune' the focusing by altering its shape and hence its power. This 'fine tuning' technique is called *accommodation*. To understand accommodation it is simplest to consider the optical system of the cornea and lens together as a single lens system, which can change its power from approximately +60 D to +75 D (a focal length changing from approximately +17 mm to +13 mm).

If the eye is used to observe a distant object, the rays of light from the object will be almost parallel when they enter the eye. Hence the ciliary muscles can relax since the cornea in combination with the eye lens (when thin) has focal length of around 17 mm (the approximate distance from lens to retina). This relaxed eye is therefore effectively focusing parallel rays on to the retina as illustrated in Fig. 3.8(a).

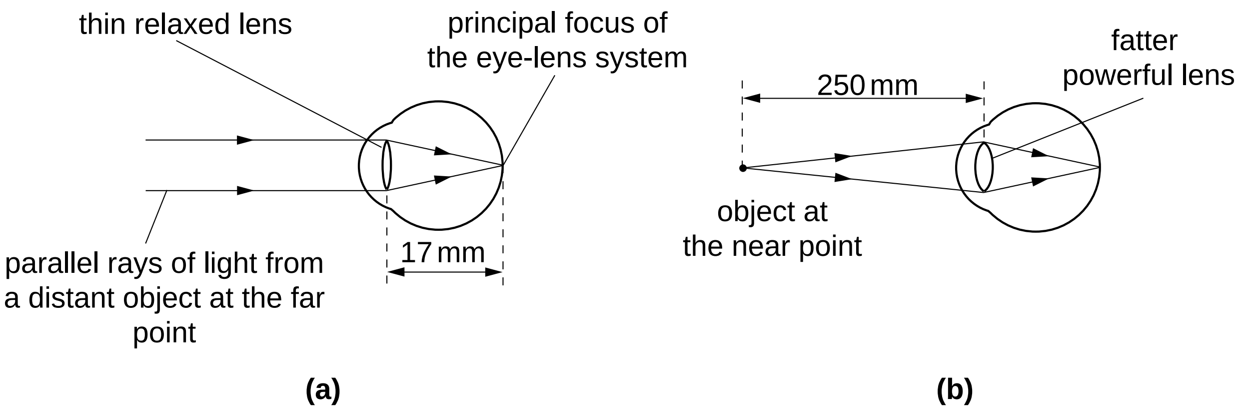


Fig. 3.8

When the eye looks at near objects the rays entering the eye will be diverging as in Fig. 3.8(b). Thus the eye lens must be made stronger to focus light on to the retina. This is achieved by the ciliary muscles squashing the lens around its edges in order to increase its power. For normal adult eyes, the shortest object distance for which the eye lens can still produce a focused image is about 250 mm. The closest position to the eye at which the eye can focus the image on to the retina is referred to as the *near point* of the eye. Even here, eye strain can be great over a period of time due to the effort required by the ciliary muscles. The *far point* of the normal eye is taken to be at infinity, i.e. very distant objects may be focused with the normal eye.

Fig. 3.9(a) and Fig. 3.9(b) show two instances of incoming light rays that are impossible for the eye to focus sharply on the retina.

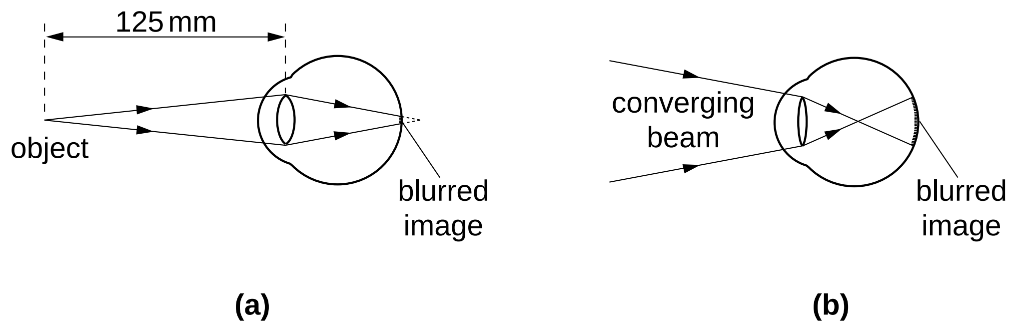


Fig. 3.9

In (a), the object is between the near point and the cornea. The lens cannot be made powerful enough to converge the rays on to the retina.

In (b), the converging beam incident on the eye converges to a point in front of the retina, even if the eye lens is at its most relaxed or least powerful. In both cases, an image is produced on the retina but it is blurred.

It may appear that, for a given lens power, an object can only be at a certain point in order to produce a focused image on the retina. This is not true. If a person views a distant object, he observes that there is a range of positions around that object, both closer to the eye and further away from it, where other objects are also in focus. The same applies to objects close to the near point although the range of focused positions is far less in this case. However, it has already been shown that different object positions give different image positions when the same power of lens is used.

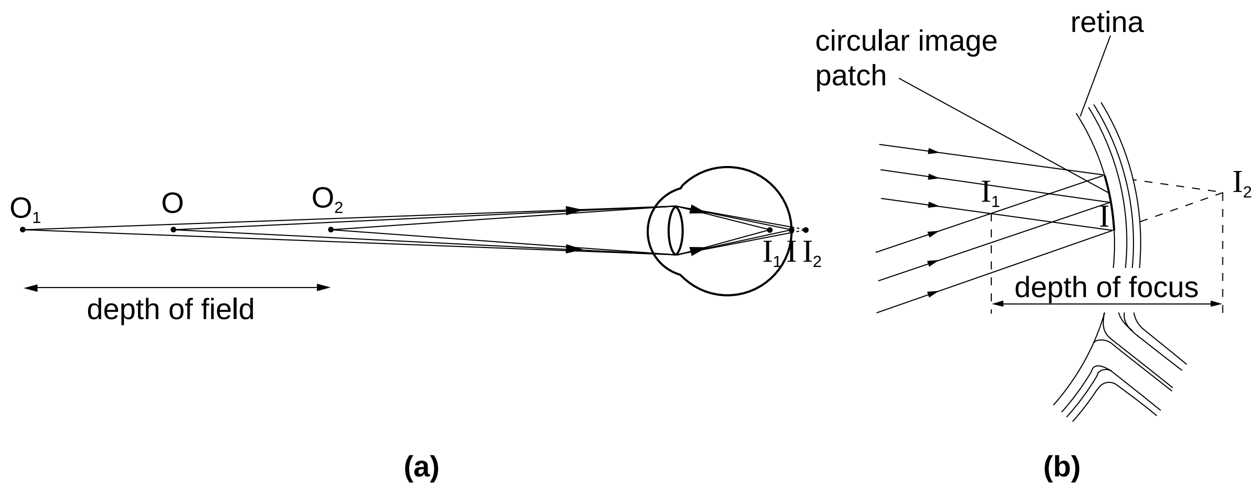
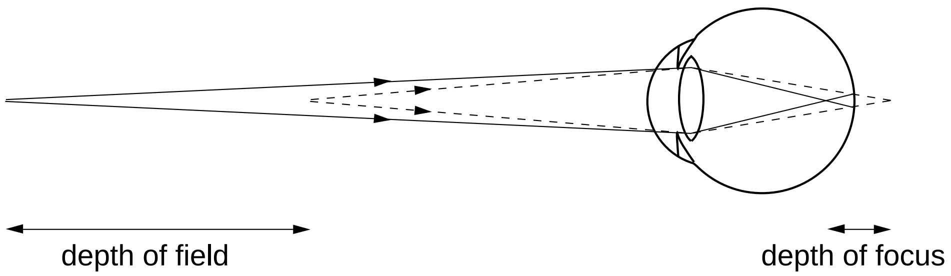


Fig. 3.10

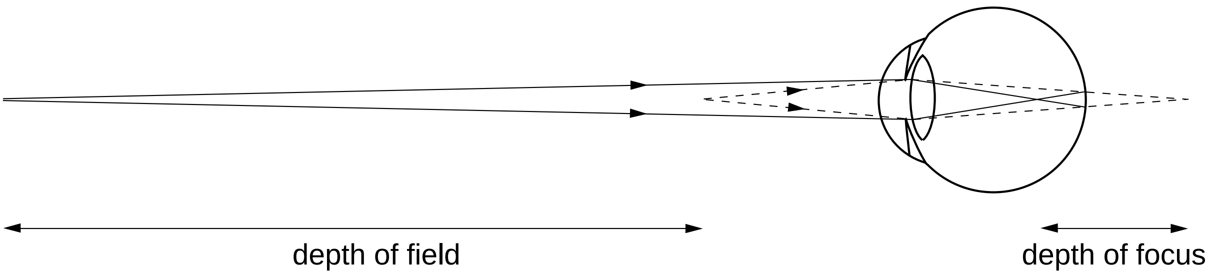
Fig. 3.10(a) shows the limits of position O_1 and O_2 around object position O where the eye lens system will produce acceptably focused images on the retina. In fact, as can be seen in Fig. 3.10(b), the images I_1 and I_2 of object positions O_1 and O_2 respectively fall slightly in front of or behind the retina. The important feature of the rays forming I_1 and I_2 is the circular image patch which they produce on the retina. If this patch is small enough to cover no more than a few nerve cells – the light receptors known as rods and cones – then all object positions O_1, O and O_2 are in acceptable focus. The distance O_1O_2 is therefore known as the *depth of field* and the corresponding distance I_1I_2 is known as the *depth of focus*.

The angle at which rays of light entering the eye from distant objects varies little over a wide range of object distances. Hence, for a given depth of focus, the range of object distances over which an acceptably focused image is obtained is large. When viewing objects close to the eye, this is not so. If a second object is only a small distance from the viewed object, its image distance from the retina will be relatively large, thus producing a circular patch on the retina too large for clarity.

The depth of field and the depth of focus are both dependent on the diameter of the aperture of the iris. This is illustrated in Fig. 3.11.



(a) large aperture iris



(b) small aperture iris

Fig. 3.11

It can be seen that, for a particular diameter of image patch on the retina (i.e. tolerable focusing), both the depth of field and the depth of focus increase as the aperture of the iris decreases in diameter.

3 (c) Candidates should be able to distinguish between short sight, long sight and astigmatism.

3 (e) Candidates should be able to explain how short sight, long sight and astigmatism can be corrected by using spectacle lenses or contact lenses.

The ability of the eye to focus images of objects situated at different distances from the eye has been discussed in section *3(a)*. When an object is placed as close as possible to the eye for a focused image to be formed on the retina, the object is said to be at the *near point* of the eye. For normal vision, the near point is usually taken as being 25 cm from the eye. The *far point* is the position of the furthest object from the eye which gives rise to a focused image on the retina. For normal vision, the far point is at infinity.

Short sight (Myopia)

When a person is said to be short-sighted, it means that only objects which are close to the eye can be focused. This is illustrated in Fig. 3.12.

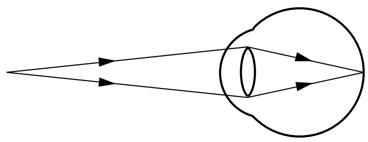


image of an object close to the eye is sharp

(a)

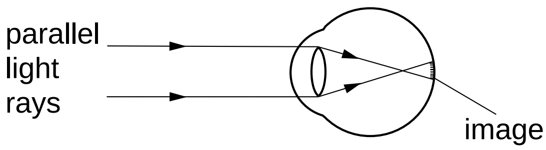


image of a distant object is blurred

(b)

Fig. 3.12

Short sight is caused either by the inability of the eye to allow the lens to relax enough to become sufficiently low powered, or by the eyeball being too long so that the distance from the lens to the retina is too large. Both these effects cause light from a distant object to focus in front of the retina, as shown in Fig. 3.12(b).

The eye can focus on to the retina rays of light coming from near objects, so that it can cope with divergent light. By finding a method of changing the parallel rays of light from a distant object into a divergent beam when they reach the cornea, the problem of short sight can be solved.

Spectacles using diverging lenses are therefore used. By placing a diverging lens of the correct power in front of the short-sighted person's eye, the image of a distant object will be focused on to the retina, as illustrated in Fig. 3.13(a).

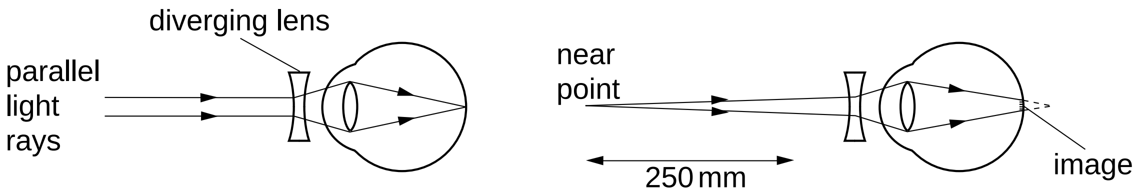


image now sharp with the help of a diverging lens

(a)

image of an object at the near point now blurred

(b)

Fig. 3.13

This has solved the person's problem as regards focusing on distant objects but if an object at the near point is to be viewed, this might be impossible to see clearly without removing the lens. This is because the diverging lens has made the light rays from the object close to the eye even more divergent, as seen in Fig. 3.13(b). Whether a short-sighted person has to remove the lens to see near objects depends on the power of the diverging lens which is required to view distant objects. Another factor to take into account is the position of that person's near point without the spectacles.

Long sight (Hypermetropia)

A person having long sight is a person who can only focus clearly on objects which are a long distance away from the eye, as shown in Fig. 3.14(a).

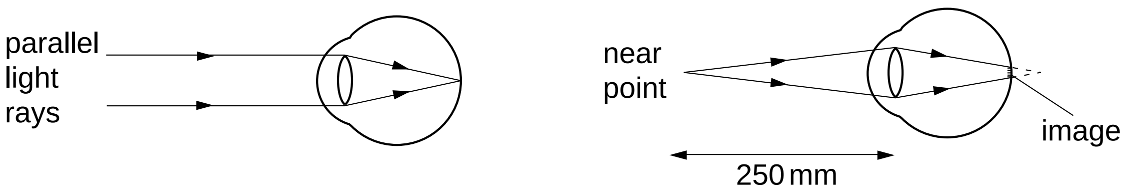


image of a distant object is sharp

(a)

image of an object at the near point is blurred

(b)

Fig. 3.14

The optical system of this eye is not powerful enough to focus light from objects close to the eye. Long sight is caused either by the inability of the eye to make the lens thicker and hence more powerful, or by the eye being too short such that the distance from the lens to the retina is too short. Both of these effects cause light rays from objects close to the eye to focus at points behind the retina, as shown in Fig. 3.14(b).

Long-sighted people tend to have difficulty in reading due to the blurred images their eyes produce when books are held closer than at arm's length. Also, long sight can be a problem for older people since the eye lens may harden and become too stiff to accommodate. Since long-sighted people cannot make the eye lens powerful enough to bend more divergent light rays to form an image on the retina, converging lenses are used. These converge the light slightly inwards before reaching the eye cornea-lens system, as shown in Fig. 3.15.

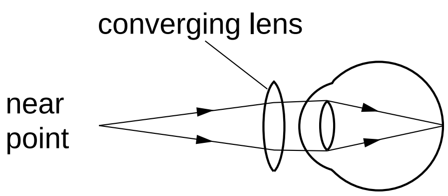


image of an object at the near point now
sharp with the help of a converging lens

Fig. 3.15

This eye is then able to focus an image of an object at the near point for a normal eye on to the retina.

Astigmatism

A person with astigmatism has difficulty in focusing light rays from objects in different planes at the same time. This means that if confronted with a grid similar to the pattern shown in Fig. 3.16, the person would be unable to focus at any one time all four sets of parallel lines.

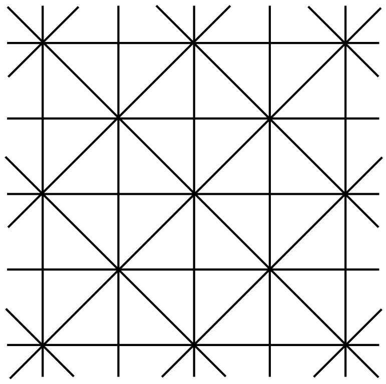


Fig. 3.16

The cause of astigmatism is normally an uneven surface to the cornea such that different planes of the surface have different curvatures. In fact, the cornea shape is commonly more cylindrical than spherical in cross-section. Consequently, the simplest way of correcting astigmatism is to use a cylindrical lens (see Fig. 3.17), adjusted so that its axis is perpendicular to the axis in which the eye cornea-lens system is cylindrical.

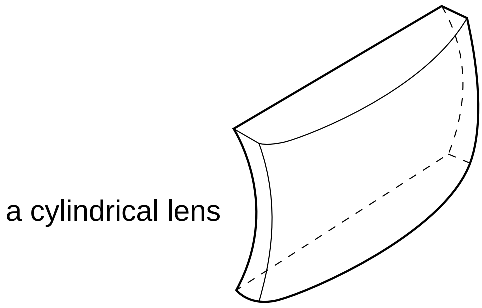


Fig. 3.17

The cylindrical lens refracts incoming light rays into a single plane rather than a point, before the perpendicular cylindrical eye lens system completes the focusing to form a sharp image.

In general, the focal lengths of lenses required to correct eye defects are comparatively long. As a result, the lenses used in spectacles are often meniscus lenses. Such lenses are illustrated in Fig. 3.18.

Fig. 3.18

Contact lenses have become an increasingly common alternative to spectacles for the correction of defects of vision. These lenses are about the size of the eye’s iris and are floated on a thin film of fluid which covers the eye. The radius of curvature of the two surfaces of the contact lens are quite different. The inner surface is made to fit the shape of the cornea and surrounding eyeball, whilst the curvature of the outer surface determines the nature and power of the lens.

All three forms of eye defect can be treated using contact lenses. In cases of astigmatism, where there can be only one correct orientation of the contact lens, the lens may be made to be heavier at the bottom so that the lens automatically aligns itself by gravity.

Some contact lens shapes are shown in Fig. 3.19.

Fig. 3.19

3 (f) *Candidates should be able to recall and apply the lens formula to calculate the focal length of the auxiliary lenses required to correct short sight and to correct long sight.*

A person with short sight cannot see distant objects clearly. The far point has moved from infinity to somewhere closer to the eye, and the image of an object at infinity would be formed in front of the retina. This is illustrated in Fig. 3.20.

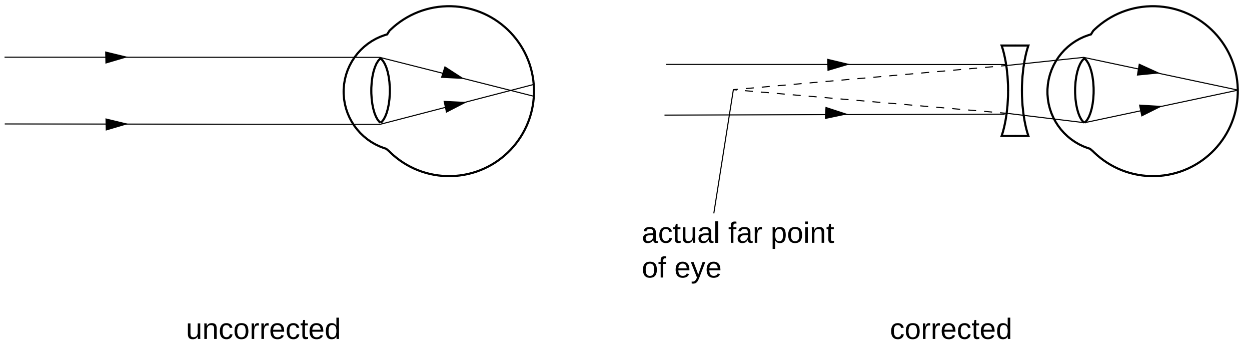


Fig. 3.20

To correct this defect, a diverging lens is used. The effect of the lens is to make parallel light (i.e. light from infinity) appear to diverge from the actual far point of the eye.

For example, a person has short sight and cannot focus on objects further than 200 cm from the eye. The focal length of the lens required can be obtained using the lens formula, i.e.

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} .$$

Since an object situated at infinity must appear to the eye to be at the far point,

$$u = \infty$$
$$v = -200\,\text{cm} \qquad \text{(virtual image)}$$
$$\text{so} \quad \frac{1}{\infty} - \frac{1}{200} = \frac{1}{f} ,$$
$$f = -200\,\text{cm, i.e. diverging lens.}$$

Note that it is assumed that the correcting lens is close to the eye.

Consider the following example.

A person cannot obtain focused images of objects nearer than 500 mm.
Identify this person's defect of vision and suggest the type and power of the lens which would be required to correct it.
What would be the effect of this lens on that person's far point? You may assume that the normal near point is 250 mm from the eye.

Since the person can focus objects from infinity to 500 mm, he must be long-sighted. The eye cannot make its lens thick enough to form on the retina a focused image of near objects and consequently requires assistance to converge the light, i.e. a converging lens is required.

The arrangement is shown in Fig. 3.21.

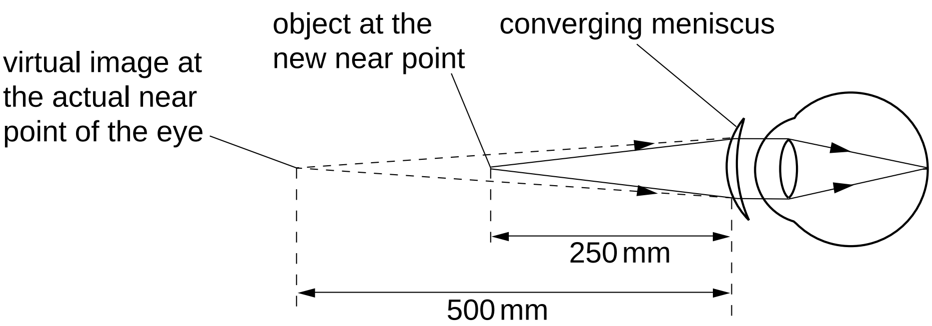


Fig. 3.21

This shows a lens using a near point object 250 mm from the eye in order to form an image at 500 mm from the eye, because this is the closest point at which the eye can form an image on the retina. It is assumed that the correcting lens is very close to the eye.

Since the image at 500 mm is on the same side of the lens as the object, the image is virtual, so applying the sign convention, v is negative.

Using $\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$ where $u = +250$ mm and $v = -500$ mm,

$$\frac{1}{f} = \frac{1}{+250} + \frac{1}{-500} = \frac{2}{500} - \frac{1}{500} = \frac{1}{500},$$

therefore $f = +500$ mm

This means that the focal length of the converging meniscus lens is +500 mm, i.e. +0.50 m and the lens power is $+\frac{1}{0.50}\text{ D} = 2.0\text{ D}$.

Since the person's near point has now been brought closer to the eye by the use of the corrective lens, the same must happen to the far point. In order to calculate the new far point position: *f*, the focal length of the lens chosen is +500 mm;
v, the image distance of the new far point which is to be infinity (∞).

Now $$\frac{1}{f} = \frac{1}{u} + \frac{1}{v},$$

so $$\frac{1}{+500} = \frac{1}{u} + \frac{1}{\infty}$$

$$= \frac{1}{u} + \text{zero}.$$

Therefore, $$\frac{1}{+500} = \frac{1}{u}$$

and $$u = +500\,\text{mm}$$

so the new far point using the converging meniscus is now 0.50 m in front of the lens (see Fig. 3.22).

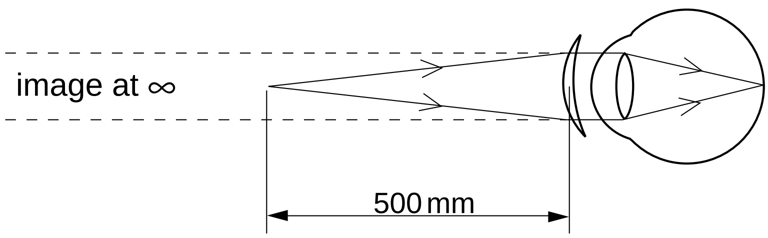


Fig. 3.22

M4. The Physics of Hearing

4 (a) *Candidates should be able to explain how the ear responds to an incoming sound wave.*

The ear may be divided into three sections, the outer ear, the middle ear and the inner ear (see Fig. 4.1).

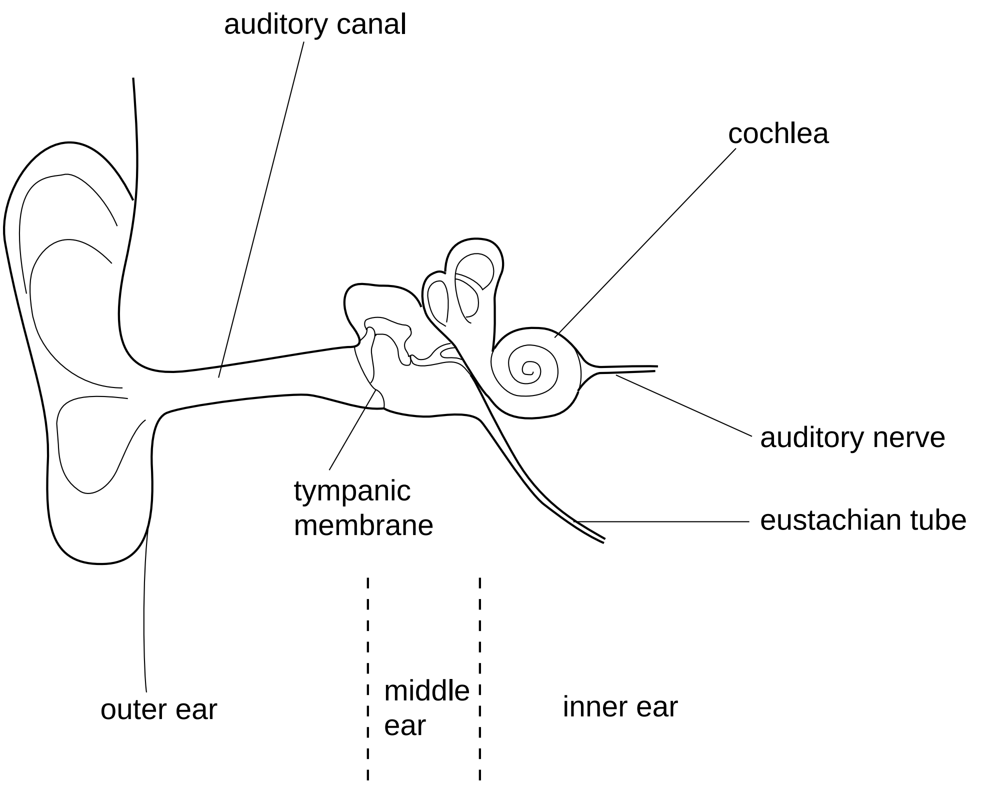


Fig. 4.1

The function of the outer ear is to collect and direct sound energy on to the tympanic membrane (eardrum). This is situated at the closed end of the auditory canal.

The middle ear consists of an air-filled cavity which houses a number of bones, acting as a lever system. The lever system consists of three bones, sometimes referred to as the hammer, anvil and stirrup, as shown enlarged in Fig. 4.2.

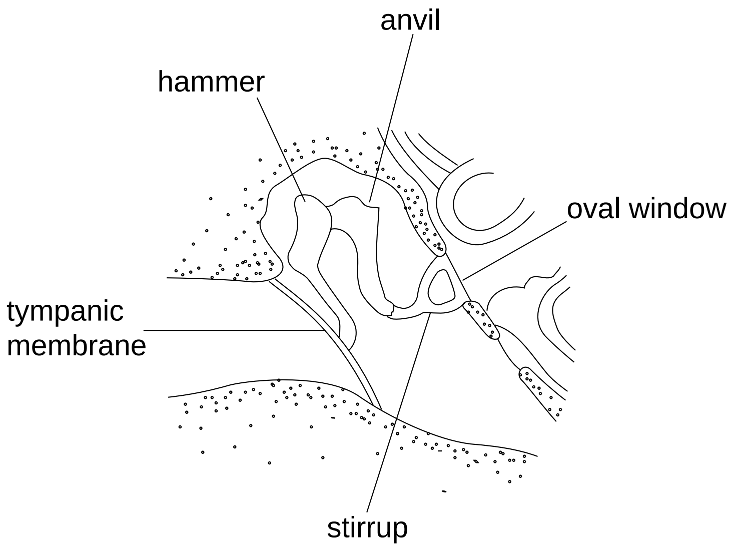


Fig. 4.2

The purpose of this lever system is to reduce the amplitude of vibration produced in the tympanic membrane and, at the same time, increase the vibrational pressure on the oval window. The mechanism may be represented by a lever system as shown in Fig. 4.3.

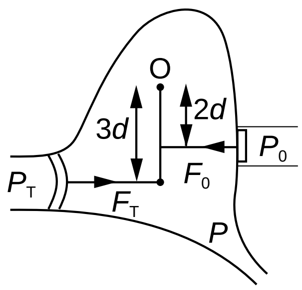


Fig. 4.3

F_T is the instantaneous force produced in the hammer by the tympanic membrane and F_0 is the resultant force in the stirrup. The lever oscillates about O. Using the principle of moments,

$$F_T \times 3d = F_0 \times 2d$$

$$\frac{F_0}{F_T} = \frac{3}{2}$$

These forces F_0 and F_T are related to the relevant pressures as follows.

The instantaneous pressure difference $(P_0 - P)$ is that between the inner ear and the middle ear. Similarly, the pressure difference between the outer ear and the middle ear is $(P_T - P)$. P_T, P and P_0 represent the instantaneous pressures in the outer ear, the middle ear and the inner ear, respectively (see Fig. 4.3). If A_T is the area of the tympanic membrane, then the force F_T on the membrane is

$$F_T = A_T (P_T - P).$$

Similarly, the force F_0 on the oval window of area A_0 is

$$F_0 = A_0 (P_0 - P).$$

Hence $$\frac{F_0}{F_T} = \frac{A_0}{A_T} \times \frac{(P_0 - P)}{(P_T - P)}.$$

The pressure changes on the oval window are greater than those on the tympanic membrane by a factor equal to

$$\frac{F_0}{F_T} \times \frac{A_T}{A_0} = \frac{3}{2} \times \frac{A_T}{A_0}.$$

Now, for a typical ear, A_T = 55 mm^2 and A_0 = 3.2 mm^2. Thus, the pressure difference is magnified by a factor of

$$\frac{3}{2} \times \frac{55}{3.2} \simeq 25.$$

The middle ear is connected to the back of the throat by the eustachion tube so that, in sound-free conditions, atmospheric pressure may act on both sides of the tympanic membrane. If these pressures are not equal, as in for example, during the rapid ascent or descent in an aircraft, discomfort may be experienced and the ears may “pop” as the pressures are equalised.

The inner ear is filled with liquid and houses the cochlea. Inside the cochlea is a membrane consisting of hairs which runs the length of the cochlea. These hairs vary in length, diameter and stiffness such that each resonates at a different frequency from that of other hairs. When a pressure wave passes through the liquid in the cochlea, hairs resonate according to the frequencies in the pressure wave. These resonating hairs stimulate nerves which transmit messages to the brain via the auditory nerve.

4 (b) Candidates should be able to show an understanding of the significance of the terms sensitivity and frequency response.

The range of frequencies which can be heard by different people varies, but most can hear sounds in the range 20 Hz to 15 kHz. If the frequency is below 20 Hz, the pressure wave does not disturb the membranes or hairs in the inner ear. Therefore, such a low frequency is not heard. The high frequency limit for hearing is about 20 kHz but this limit does decrease with age. Fig. 4.4 gives the frequency range for some musical instruments.

	frequency range/Hz
piano violin trumpet	27 – 4186 230 – 2217 174 – 932

Fig. 4.4

The human ear can distinguish between differences of 2 Hz or 3 Hz in the frequency range 60 Hz – 1000 Hz. Beyond 1000 Hz, it becomes progressively more difficult to distinguish between close frequencies.

The smallest sound intensity which can be heard by the ear is known as the *threshold of hearing* and is generally taken as $1.0 \times 10^{-12}\,\mathrm{W\,m^{-2}}$ at 3 kHz. The threshold of hearing varies with frequency. For low frequencies, much below 1 kHz, higher intensities are required. if this were not the case, it would be possible to hear our own blood flow which would then cause interference for the reception of external sounds. Fig. 4.5 shows the average threshold intensity levels at which particular frequencies are heard.

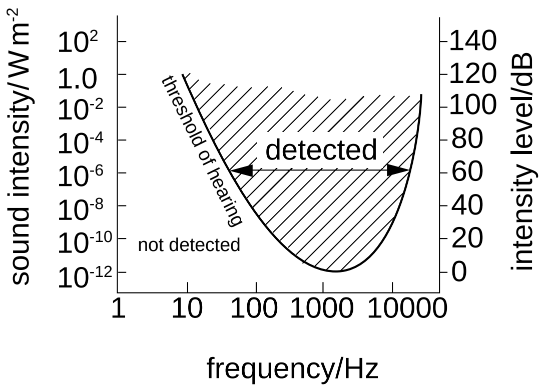


Fig. 4.5

Above these threshold intensities, sound can be detected but the sensitivity of the ear to changes in sound intensity is not constant. The sensitivity of the ear is its ability to detect the smallest fractional change ΔI of the intensity I. Thus, sensitivity depends on the ratio $\frac{\Delta I}{I}$ and a smaller discernible relative change in intensity implies a greater sensitivity. Also, for the same sensitivity, a smaller change in intensity can be detected when the intensity is low than when the intensity is high.

4 (c) *Candidates should be able to show an appreciation of the very wide range of intensities which can be detected by the ear and recall the orders of magnitude of the threshold of hearing and the intensity at which discomfort is experienced.*

The ear can detect a very wide range of intensities from a threshold of approximately $1.0 \times 10^{-12}\,\mathrm{W\,m^{-2}}$, corresponding to a pressure change of about $2 \times 10^{-5}\,\mathrm{Pa}$, to an upper limit of about $100\,\mathrm{W\,m^{-2}}$ in the frequency range 1 kHz – 6 kHz (see Fig. 4.5). At the upper limit, pain and temporary deafness may be experienced. The range of intensities over which a sound can be heard reduces as the frequency of the sound moves out of this frequency range. Serious damage can be caused to the hearing ability of the ear if sounds of high intensity are experienced for more than a short time. Fig. 4.6 indicates the intensity for a number of different sounds.

	intensity/$\mathrm{W\,m^{-2}}$
threshold of pain	1×10^{2}
discotheque	1
thunder overhead	1×10^{-1}
jet overhead	1×10^{-2}
road drill	1×10^{-3}
road traffic	1×10^{-4}
conversation	1×10^{-5}
rustling leaves	1×10^{-8}

Fig. 4.6

4 (d) *Candidates should be able to show an understanding of the significance of the logarithmic response of the ear to intensity.*

4 (e) *Candidates should be able to recall and solve problems using the equation intensity level = 10 lg (I/I_0), giving intensity level in dB in terms of the intensity I and the threshold intensity I_0.*

The response of the ear to intensity of sound at different levels is not linear. At low intensities, small changes in intensity may be noticed. At high intensities, the same actual change in intensity may not be detected because the ratio of the additional stimulus to the original stimulus already being received is too small. Equal changes in intensity across the range of audible intensities are not perceived by the ear as equal changes in loudness of the sound. If the intensity at a given frequency was increased from $1 \times 10^{-7}\,\mathrm{W\,m^{-2}}$ to $2 \times 10^{-7}\,\mathrm{W\,m^{-2}}$ and then from $2 \times 10^{-7}\,\mathrm{W\,m^{-2}}$ to $3 \times 10^{-7}\,\mathrm{W\,m^{-2}}$, the ear would judge that the loudness had increased by a smaller amount for the second change (see section *4(f)*).

For the ear to judge that the loudness has increased in equal steps, the *intensity* would have to increase by the *same ratio* on each occasion, i.e. the intensities would have to be in the ratios 1:2:4:8 etc. It follows that loudness change depends on

$$\frac{\text{intensity change}}{\text{initial intensity}}\ .$$

Consider the intensities shown in Fig. 4.7.

intensity I / W m^{-2}	1 x 10^{-12}	2 x 10^{-12}	4 x 10^{-12}	8 x 10^{-12}	16 x 10^{-12}
$\frac{I}{1 \times 10^{-12}}$	1	2	4	8	16
loudness	0	1*L*	2*L*	3*L*	4*L*
$\lg \frac{I}{1 \times 10^{-12}}$	0	0.301	0.602	0.903	1.204

Fig. 4.7

It can be seen that if the ratio

$$\frac{I}{1 \times 10^{-12}}$$

is calculated, the ratio shows a doubling on moving from one intensity to the next. Thus, the loudness has increased in equal steps of loudness *L*. Referring to Fig. 4.7, the logarithm of

$$\frac{I}{1 \times 10^{-12}}$$

also increases by equal steps. Hence, the logarithm of

$$\frac{I}{1 \times 10^{-12}}$$

relates directly to loudness and thus the ear displays a logarithmic response to intensity of sound.

The lowest sound intensity which the average human ear can detect is approximately $1 \times 10^{-12}\,\mathrm{W\,m^{-2}}$ at about 3000 Hz. This intensity is known as the threshold intensity I_0 and any increase from this intensity is perceived as an increase in the loudness of the sound. However, as explained above, the perceived loudness does not increase linearly with intensity, but does so on a logarithmic scale. Thus the *intensity level* of a sound, which gives an indication of the loudness of that sound, is given by

$$\text{intensity level} = 10\,\lg \frac{I}{I_0},$$

where I is the intensity of sound incident on the ear. Intensity level is measured in decibel (dB).

A sound of intensity $2.0 \times 10^{-4}\,\mathrm{Wm^{-2}}$ has an intensity level of 83 dB, because

$$10\,\lg\left(\frac{2 \times 10^{-4}}{1 \times 10^{-12}}\right) = 83.$$

The minimum change in intensity level which an average human ear can detect is approximately 3 dB.

Fig. 4.8 shows the intensity levels corresponding to the intensities listed in Fig. 4.6.

	intensity / $\mathrm{W\,m^{-2}}$	*intensity level* / dB
threshold of pain	1×10^{2}	140
discotheque close to the speaker	1	120
thunder overhead	1×10^{-1}	110
jet overhead	1×10^{-2}	100
road drill	1×10^{-3}	90
road traffic	1×10^{-4}	80
conversation	1×10^{-5}	70
rustling leaves	1×10^{-8}	40

Fig. 4.8

90 dB is considered to be at the upper limit of acceptable noise and can be due to – for example – a heavy lorry some 7 m away.

The intensity of sound received by the ear depends on the power of the source and the distance of the ear from the source. The earphone of a personal stereo may give sound intensity levels of as much as 100 dB, when fitted to the ear, and yet it is hardly audible to another person standing nearby.

4 (f) *Candidates should be able to show an understanding that loudness is the subjective response of an individual to an intensity level.*

From previous discussions, it appears that loudness and intensity level are directly related. This is not strictly true. Intensity level is defined in terms of the intensity of the sound and the threshold intensity at one specific frequency (3 kHz). It is therefore a precise physical quantity. On the other hand, loudness depends on both the observer and the frequency. Loudness is the subjective response of a person to sound of a particular intensity.

In order to define a consistent unit of loudness, the frequency 1 kHz is chosen as standard and all sounds, regardless of frequency, are compared with the intensity of sound of frequency 1 kHz.

The intensity of the standard 1 kHz source is adjusted until it is perceived as being as loud as the sound being evaluated. If the intensity level of the 1 kHz standard source is found to be, for example, 60 dB when it is as loud as the second source, this second source is said to have a loudness of 60 phons.

Fig. 4.9 shows how a person with normal hearing perceives intensity levels at different frequencies.

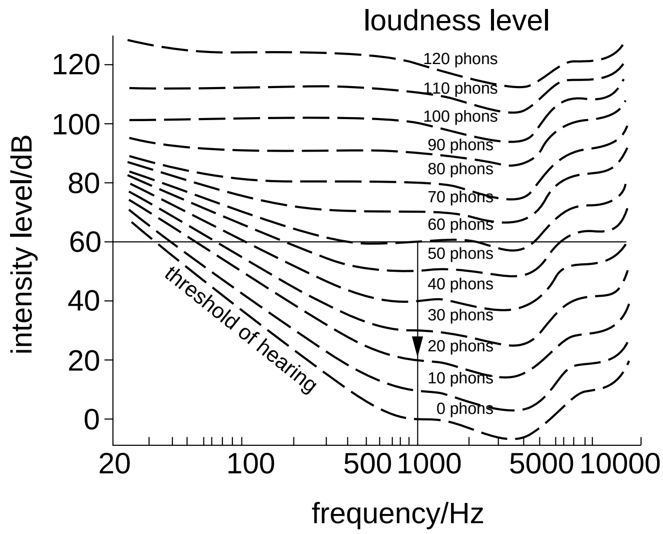


Fig. 4.9

Each broken line represents a line of equal loudness and the loudness, measured in phon, is numerically equal to the intensity level of the sound, measured in dB, at 1 kHz.

It should be remembered that loudness is a subjective response, as is the nuisance value of sound. Two different sounds may have the same phon level but they may be perceived differently, e.g. a rock-concert may be a pleasure to some people but a nuisance to others.