ESAT Guide

Physics

June 2025

Contents

INTRODUCTION

This is worth reading before you use this guide.

We have three aims in writing this guide:

First, we want to set out what we expect you to know for the ESAT. We do this by basing each part of the guide on the relevant part of the specification.

Second, we want to encourage you to think deeply and carefully about science and mathematics and to develop a good understanding of the topics in the specification. To help with this, we have added a lot of discussion and examples as well as some exercises throughout the guide.

Third, we want to make sure that all candidates have access to a free resource to help them prepare for the ESAT.

How to use this guide

You do not need to work through all this guide as you will find that you know many of the topics in the specification very well already. Use this guide as a resource to help you clarify and review topics that you are less familiar with. We have broken down our discussion to fit exactly with the specification to make things as simple to navigate as possible.

What this guide is not

This guide is not a comprehensive textbook: we do not cover every topic to the same level of detail, and we do not develop every topic from scratch. It is also not a substitute for sustained hard work and preparation. It is a resource to help you and to guide you in the right direction.

Should I take an ESAT course?

We do not recommend that you take a course, and we do NOT endorse any courses. No one from the ESAT development team teaches on any courses. All the resources you need to prepare are available from the UAT-UK website and are entirely free.

A final note

We have used boxes throughout the guide to help you navigate.

The relevant part of the specification is found in these sorts of boxes:

Specification

examples in these sorts of boxes:

Examples

and exercises [answers are at the end of each section] in these sorts of boxes:

Exercises

We hope to be able to update and, if necessary, correct the guide now and again. Look at the date on the front page to see when the guide was last edited.

P1. Electricity

P1.1

Electrostatics:

a. Know and understand that insulators can be charged by friction.
b. Know and understand that charging is caused by gain or loss of electrons.
c. Know and understand that like charges repel and unlike charges attract.
d. Understand applications and hazards associated with electrostatics, including the role of earthing.

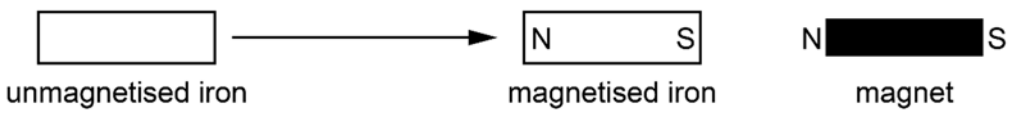
Know and understand that insulators can be charged by friction

- When two insulators move relative to each other, friction between the two can result in both objects becoming electrically charged.
- Examples of this are: plastic being rubbed by a duster; hair being combed; fuel moving through a pipe; an aeroplane moving through the air.
- Conductors, such as metals, can also become charged, but will only retain that charge if they are insulated from their surroundings. If they are not insulated, then any charge that builds up will leak away.

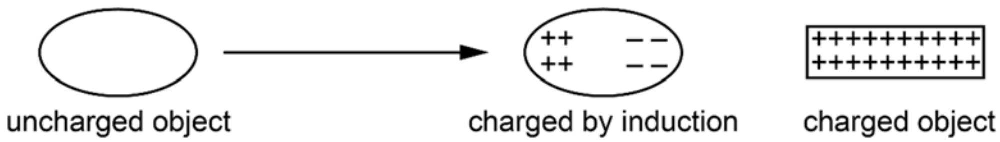
Charging by induction

It is also possible for objects to become charged by a process called induction. Rather like when an unmagnetised piece of iron becomes magnetised when placed near a magnet, a neutral object placed near to a charged object can become charged:

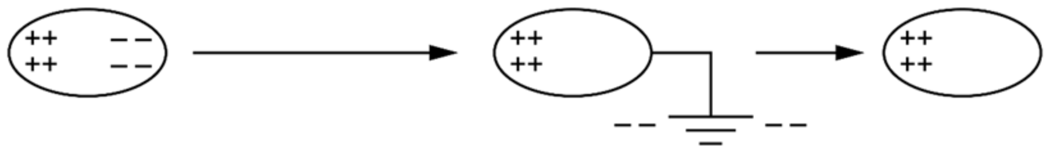
Magnetic induction:



Electrostatic induction:



If one end of an object charged by induction is momentarily earthed, allowing the charge that has accumulated at that end to leak away, then the object becomes permanently charged:



This is why uncharged objects (such as pieces of paper) can get attracted to charged objects (such as pieces of plastic) even though they are electrically neutral.

Know and understand that charging is caused by gain or loss of electrons

- We know that atoms are made from protons and neutrons (which are tightly bound together deep inside the atom in a small region called the nucleus) and electrons (which are found in certain shells around the nucleus). The electrons are the particles that form the outer regions of an atom, and are much less tightly bound than the particles in the nucleus.
- Protons are positively charged and electrons are negatively charged. Neutral atoms contain the same number of each. Neutrons are uncharged and have no effect on the overall charge of the atom.
- When an atom loses an electron, it has more protons than electrons and so becomes positively charged. When an atom gains an electron, it has more electrons than protons and so becomes negatively charged.
- When two objects are rubbed together, friction can result in electrons being transferred from one to the other. This results in the two objects becoming electrostatically charged. One object (the one that loses electrons) becomes charged positive, and the other (the one that gains electrons) becomes charged negative.

It is always electrons that are responsible for electrostatic charge.

A common misconception is that positive electrostatic charge is caused by movement of protons. But protons are so tightly bound, deep inside the nucleus, that they cannot move from one object to another. Positive charge is caused by electrons moving away (leaving more protons than electrons), just as negative charge is caused by electrons moving in (resulting in more electrons than protons). It is always electrons that move between objects.

The total charge always remains the same. One of the fundamental principles of physics is that charge is always conserved. So, for every object that becomes positively charged, another must become negatively charged.

Which way the electrons move depends on which atoms attract the electrons more strongly. When two objects are rubbed together, electrons move away from the atoms that attract them less strongly, and towards the atoms that attract them more strongly. So, whether a particular material becomes positively or negatively charged, when charged by friction, depends on the nature of the two materials involved.

Charging by induction

It is electrons moving that are responsible for charging by induction. In the example above, the initially uncharged object has become positively charged because electrons that were in it have been repelled to Earth. Similarly, an object can become negatively charged by induction if electrons are attracted from Earth.

When an uncharged rubber balloon is rubbed with a piece of cloth, it becomes negatively charged and remains charged for some time.

Explain how it becomes charged and why it retains this charge.

The balloon is initially neutral. Electrons are transferred from the cloth to the balloon when they are rubbed together.

Electrons are negatively charged. The balloon ends up with excess of electrons giving it an overall negative charge.

(This is called 'charging by friction'.)

The charge remains on the balloon because it is made from an insulating material (rubber) so charge cannot flow away, even if part of the balloon is connected to earth (e.g. held in the hand).

Know and understand that like charges repel and unlike charges attract

- Two charged objects placed near each other exert an electrostatic force on each other.
- The size of this force depends on the size of the two charges and the distance between them.
- The larger the charges, the larger the force, and the larger the distance of separation the smaller the force.
- The direction of the force depends on how the signs of the charges compare.
- Like charges (+ and +, or – and –) repel each other.
- Unlike (opposite) charges (+ and –) attract each other.

There are several applications and consequences of this, for example:

- When two objects become electrostatically charged by friction between them (with one becoming positively charged and the other becoming negatively charged), they will then attract each other as a result. This is why, for example, hair that is combed often 'sticks' to the comb.
- In theory, it is possible for two conductors to charge each other by friction between them. However, the fact that electrons in them are free to move means that any charge transfer between them will instantly return, given that the objects must be in contact for friction to occur between them.
- It is attraction of electrons to positive charge, or repulsion of electrons from negative charge, that is responsible for the movement of charge during electrostatic induction.

A Van der Graaf generator can be used to demonstrate electrostatic effects. The top dome becomes strongly positively charged. When a person stands on an insulated stool and touches the dome their hair stands on end as shown. Explain why this occurs.

The dome is positively charged so electrons flow from the person to the dome, making the person, and their hair, positive. Like charges repel, so every hair repels every other hair and the scalp. These repulsive forces make the hairs stand on end. The insulated stool prevents electrons being drawn up from earth (otherwise no charge would build up on the person).

Photo: Science Photo Library

Understand applications and hazards associated with electrostatics, including the role of earthing

- Electrostatic charging by friction has some useful applications, such as photocopying, laser printing and electrostatic air cleaning. No knowledge of particular applications is expected, but an understanding of the principles of electrostatics in any application that may be presented is required. Some examples are given below, not to define required knowledge but to illustrate how the principles can be applied.
- Electrostatic charging by friction is often a nuisance. Hair sticking to the comb is an example. Thunderstorms are another. And sparking, causing mini electric shocks when getting out of the car or after walking on nylon carpets, is another.
- Sparking is what occurs when the air between two objects becomes ionised by a large voltage and therefore starts conducting. Two charged objects, that have air between them, can discharge by a spark between them. This will happen either when the charge becomes large enough or when the distance between the objects becomes small enough. Small sparks are a nuisance; large sparks can be dangerous.
- The risk of fires being caused by sparking is an aspect of electrostatic charging by friction that can make it very dangerous. An example of this is the risk of explosion when aircraft are refuelled.
- The risk of sparking can be eliminated by earthing. If two objects, that would otherwise cause each other to become charged by friction, are connected together by a wire (or if one of them is connected to Earth by a wire) then electrostatic charging cannot take place and the risk is averted.

Examples

Photocopying and printing

The scanning process results in charge being placed on the paper at the locations where the image is to be printed. The paper is then exposed to toner powder, which 'sticks' to the paper at those locations as a result of electrostatic induction. The paper is then heated so that the toner powder melts and then re-solidifies on the paper. Attraction between like and unlike charges is crucial for the operation of these devices.

Aircraft refuelling

When aircraft are refuelled, large volumes of fuel flow through the pipe very quickly. This creates large amounts of friction, resulting in the fuel and the pipe becoming electrostatically charged. Any sparking presents a significant risk of explosion of the fuel in the fuel tank. For this reason, the fuel tank and the refuelling pipe are always earthed before refuelling takes place. This earthing prevents the build-up of charge and so eliminates the risk of explosion.

Electrostatics is used when paint spraying the panels of a car. The tiny paint droplets are positively charged as they leave the spray gun. The panels of the car are earthed.

Which of the following statements about electrostatic paint spraying is/are correct?

1. Droplets of paint are attracted to one another.

2. Droplets of paint are attracted to the panels.

3. Droplets of paint can reach parts of the panels that are not directly in line with the spray gun.

4. The panels of the car body gradually become positively charged.

All the droplets have the same (positive) charge, so they repel one another. This prevents them clumping together and leads to a smooth coat of paint on the car, so statement 1) is incorrect.

As the droplets approach the car, they induce a negative charge on the surface and this attracts them to it, so statement 2) is correct.

Droplets that pass behind the car body can be attracted back to it by induction, so statement 3) is correct.

The car is earthed so, as positive charge is added by the droplets, electrons flow from earth to neutralise them, and the car body remains neutral. Statement 4) is incorrect.

P1.2

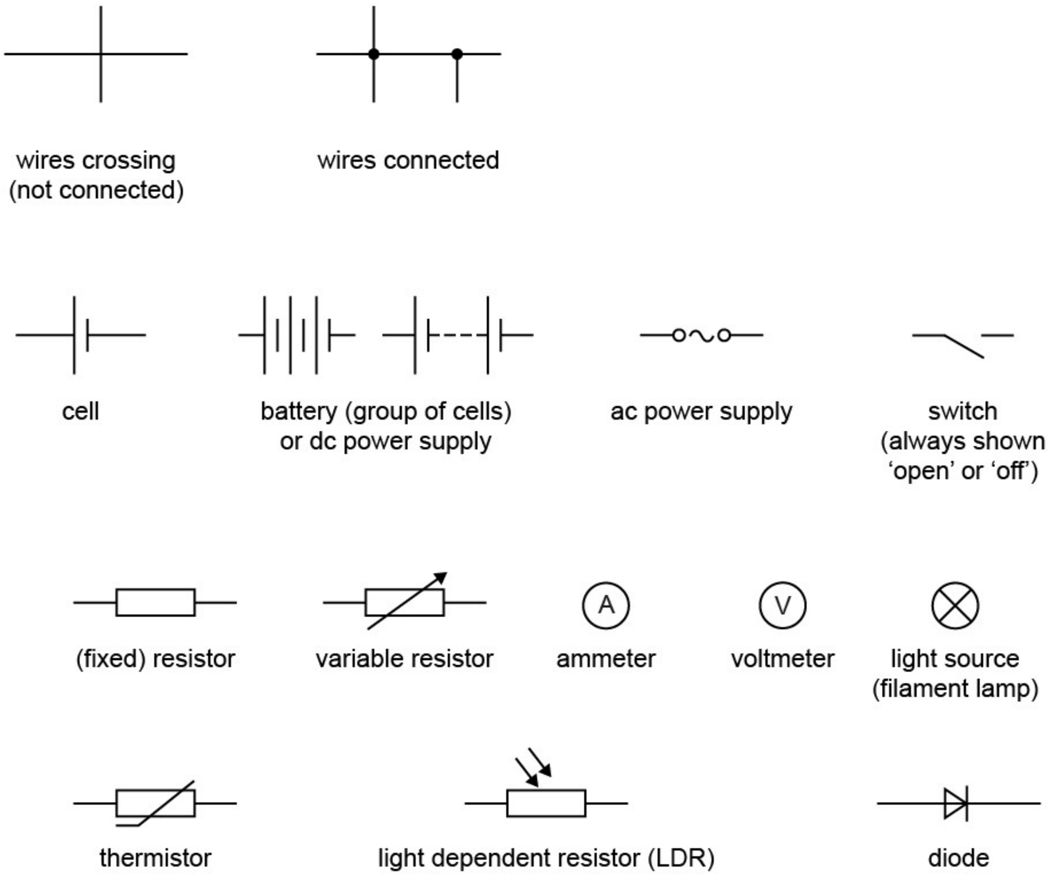
Electric circuits:

a. Know and recognise the basic circuit symbols and diagrams, including: cell, battery, light source, resistor, variable resistor, ammeter, voltmeter, switch, diode.
b. Understand the difference between alternating current (ac) and direct current (dc).
c. Understand the difference between conductors and insulators, and recall examples of each type.
d. Know and be able to apply: $current = \frac{charge}{time}$, $I = \frac{Q}{t}$
e. Know and understand the use of voltmeters and ammeters.
f. Know and be able to apply: $resistance = \frac{volatge}{current}$, $R = \frac{V}{I}$
g. Recall and interpret *V*–*I* graphs for a fixed resistor and a filament lamp.
h. Know the properties of NTC (negative temperature coefficient) thermistors, LDRs (light-dependent resistors) and ideal diodes.
i. Know and understand the current and voltage rules for series and parallel circuits.
j. Calculate the total resistance for resistor combinations in series.
k. Understand that the total resistance of a parallel combination is less than that of any individual resistor.
l. Know and be able to apply: $voltage = \frac{energy}{charge}$, $V = \frac{E}{Q}$
m. Know and be able to apply: power = current × voltage, $P = IV = I^2R$
n. Know and be able to apply: energy transfer = power × time, $E = VIt$

Know and recognise the basic circuit symbols and diagrams, including: cell, battery, light source, resistor, variable resistor, ammeter, voltmeter, switch, diode

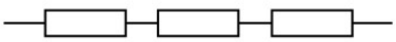
Components in electrical circuits are represented by standard symbols.

Some very common symbols

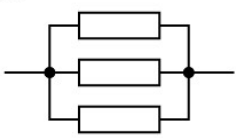


Series and parallel connections

Components can be connected in series:



or in parallel:



Understand the difference between alternating current (ac) and direct current (dc)

- Direct current (dc) is current that is always in the same direction.
- Alternating current (ac) is current that repeatedly changes direction, usually very rapidly.

Direct current

Cells or batteries (groups of cells) are sources of dc.

The output from a power supply from mains electricity can be converted from ac to dc using diodes as a 'rectifier'. A diode only allows current in one direction, in the direction of the arrow on the symbol.

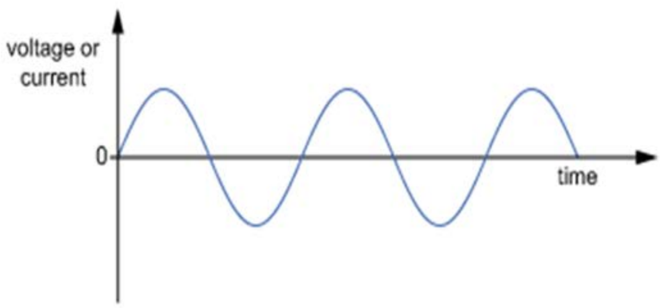
Alternating current

Generators in power stations produce ac.

Although any current that changes direction is ac, usually the change is repeated regularly to produce a 'waveform' that can be seen on an oscilloscope or computer screen.

In the UK and the rest of Europe the frequency of the mains supply is 50 Hz (i.e. the current changes direction 100 times each second, producing 50 complete 'to and fro' cycles in one second).

A very common waveform is the sine wave or 'sinusoidal' wave (a graph of $y = \sin x$ *)*

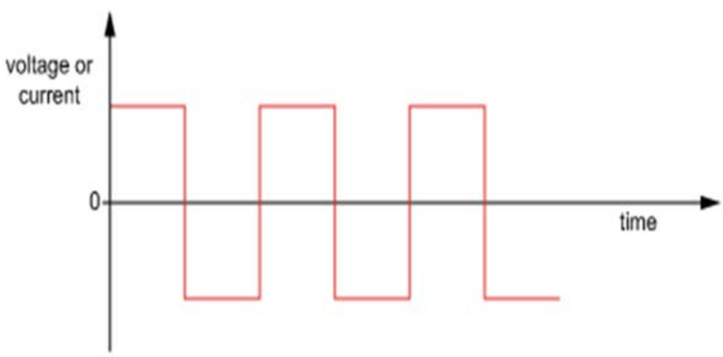


Positive values represent current in one direction and negative values represent current in the opposite direction.

Symbol for an ac supply:

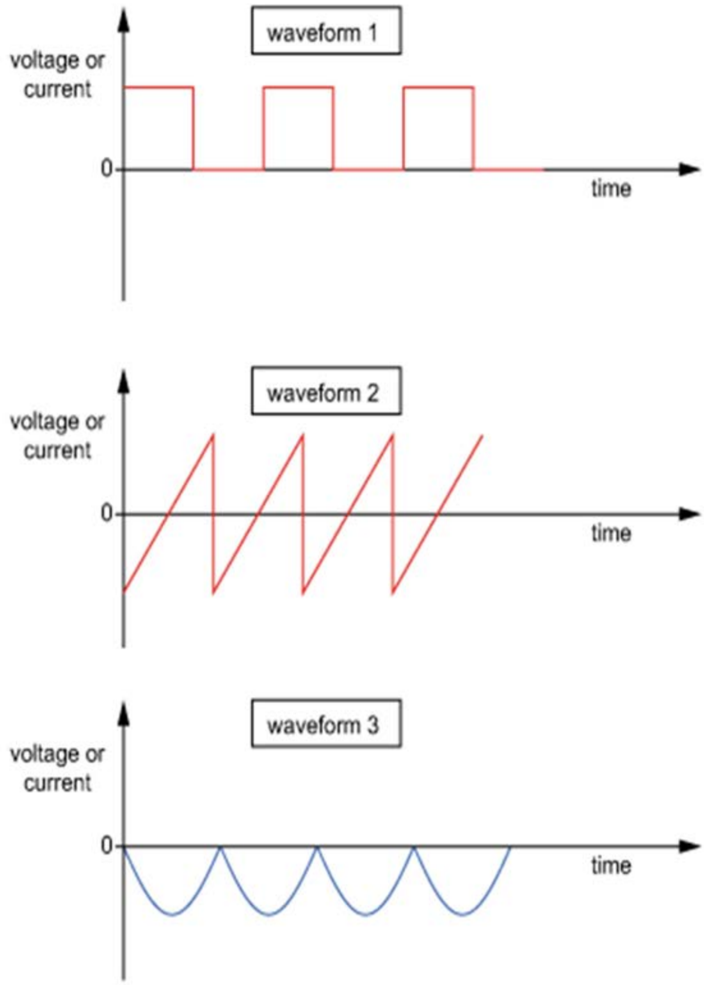
Any current that changes direction regularly is an ac; the waveform can be any shape.

A 'square wave'



Essentially, the voltage switches from positive to negative and back.

Which of these waveforms represents an alternating current?



A current is only alternating if it keeps changing direction. In other words, its waveform must include a graph line that is sometimes above the axis and sometimes below it. Only waveform 2 shows this.

Understand the difference between conductors and insulators, and recall examples of each type

Examples of good conductors:

all metals, particularly copper, gold and silver

carbon (in the form of graphite) - ionic solutions.

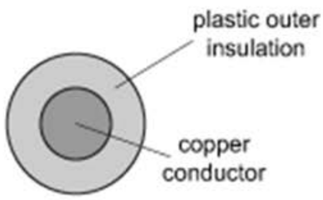
Examples of good insulators (poor conductors):

most non-metals, particularly plastics, rubber, dry wood, air, vacuum.

Water, unless extremely pure, is a conductor, so wet or damp materials are not good insulators.

Almost all materials allow electric charge to move through them to some extent, but some offer very little resistance to this flow of charge (conductors), while others have a very large resistance (insulators). An electric cable usually uses an insulator and a conductor.

Cross-section of a cable



A bathroom light is usually fitted with a long nylon pull-cord to operate a switch on the ceiling.

Why is this necessary?

In a steamy atmosphere such as in a bathroom, water will condense on to cold surfaces. Water conducts electricity, so a person touching a standard light switch could receive an electric shock through the water that condenses on the outside and the inside of the switch.

Nylon is an insulator and does not absorb water readily, so will not conduct a current from the switch to the person pulling the cord.

Know and be able to apply:

$$\text{current} = \frac{\text{charge}}{\text{time}}$$

$$I = \frac{Q}{t}$$

Electric charge Q is measured in coulombs (C). (The letter Q is used for 'quantity' of charge.)

The current I in a circuit is the rate of flow of electric charge.

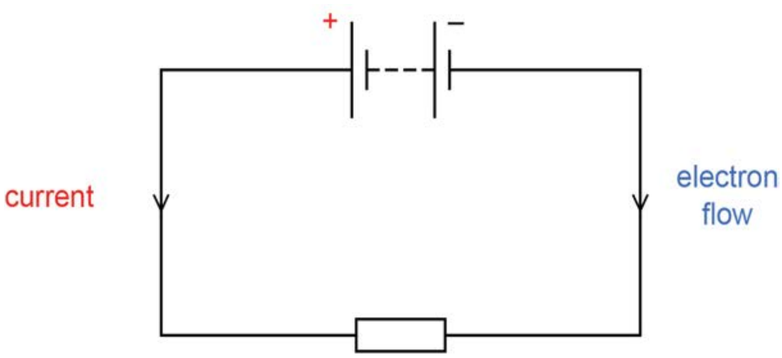
So $current = \frac{charge}{time} = \frac{Q}{t}$ and has the unit coulomb per second (C s^{-1}) or ampere (amp) (A).

1 A is a large unit of current, so other commonly used units are:

- milliamp (mA) = 1 $\times 10^{-3}$ A
- microamp (μA) = 1 $\times 10^{-6}$ A

The reason why metals are good conductors is that they contain free electrons that can move about through the metal and carry their (negative) charge with them. If a voltage is connected across a metal, the positive end of the metal attracts electrons and the negative end repels electrons. In this way the electrons can move along the metal and cause a current (a flow of charge).

Current is from the positive end of a conductor to the negative end, but the flow of electrons is in the opposite direction:



Conduction in electrolytes (liquids other than liquid metals)

Many liquids conduct quite well because they contain positive and negative ions. If two electrodes are placed in the liquid and a voltage is applied across them, the positive ions move towards the negative metal plate (the cathode), and the negative ions move towards the positive metal plate (the anode).

There is a current of 30 mA in a resistor for 20 s.

How much charge passes through the resistor in this time?

$$\text{current} = \frac{\text{charge}}{\text{time}}$$

So: charge = current × time = (30×10^{-3}) A × 20 s = 0.60 C

A charge of 6000 μC passes through a component in 50 minutes.

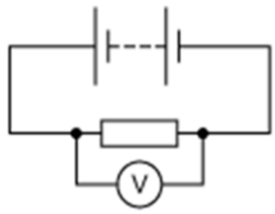
What is the average current in the component?

$$\text{current} = \frac{\text{charge}}{\text{time}} = \frac{6000 \times 10^{-6}}{(50 \times 60)}$$

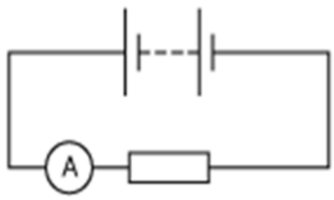
= 2.0×10^{-6} A = 2.0 μA

Know and understand the use of voltmeters and ammeters

A voltmeter is connected in parallel with a component and measures voltage.



An ammeter is connected in series with a component and measures current.

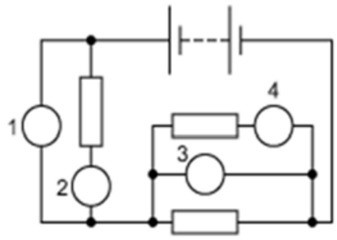


A voltmeter has very high resistance, otherwise it would tend to 'short circuit' the component across which it was connected (because there would be a significant amount of current in the voltmeter instead of in the component).

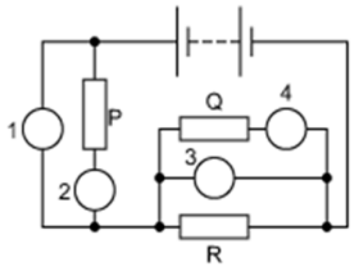
An ammeter has very low resistance, otherwise it would tend to reduce the amount of current that it was being used to measure.

The circuit shown contains several correctly connected meters labelled 1, 2, 3 and 4.

Which type of meter is each?



If the three resistors are labelled P, Q and R:



Meter 1 is connected in parallel with resistor P, and has no other component in series with it, so it is a voltmeter.

Meter 2 is connected in series with resistor P, so it is an ammeter.

Meter 3 is connected in parallel with resistor R, and has no other component in series with it, so it is a voltmeter.

Meter 4 is connected in series with resistor Q, so it is an ammeter. Although at first glance it may appear to be connected in parallel with resistor R, the fact that it is in series with resistor Q in the upper branch of the parallel arrangement means that it must be an ammeter.

Know and be able to apply:

$$\text{resistance} = \frac{\text{voltage}}{\text{current}}$$

$$R = \frac{V}{I}$$

For some components, such as a metallic conductor at constant temperature, resistance R is related to voltage V and current I by the following equation: $R = \frac{V}{I}$

For some components, such as a metallic conductor at constant temperature, the current is directly proportional to the voltage causing it. This is known as Ohm's law:

current I $\propto$ voltage V or: V $\propto$ I or: V = I × constant

The constant is the resistance R of the component. Unit: ohm (Ω)

so: V = I × R

or: $R = \frac{V}{I}$

1 Ω turns out to be quite a small unit of resistance, so other commonly used units are:

- kilohm (kΩ) = 1 ×10^3 Ω
- megohm (MΩ) = 1 ×10^6 Ω

There is a current of 60 mA in a resistor when it is connected to a 12 V battery.

What is the resistance of the resistor?

$$\text{resistance} = \frac{\text{voltage}}{\text{current}}$$

$= \frac{12}{60\times10^{-3}} = \frac{1}{5\times10^{-3}}$

= 0.2 × 10^3 = 2.0 × 10^2 = 200 Ω

A 72 MΩ resistor is connected across a 1.8 kV power supply.

What is the current in the resistor?

$I = \frac{V}{R}$

so: current =

$$\frac{1.8\times10^3}{72\times10^6} = \frac{18\times10^{-3}}{720}$$

= 0.025 × 10^{-3} = 2.5 × 10^{-5} = 25 × 10^{-6} A = 25 μA

Recall and interpret V –I graphs for a fixed resistor and a filament lamp

Fixed resistor

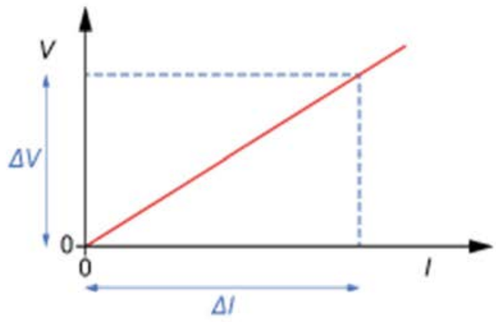
A fixed resistor at constant temperature has a constant resistance.

It is an ohmic conductor (that is, it obeys Ohm's law), so its

$$\text{resistance} = \frac{\text{voltage}}{\text{current}}$$

or $R = \frac{V}{I}$

This means that a graph of V against I will be a straight line passing through the origin:

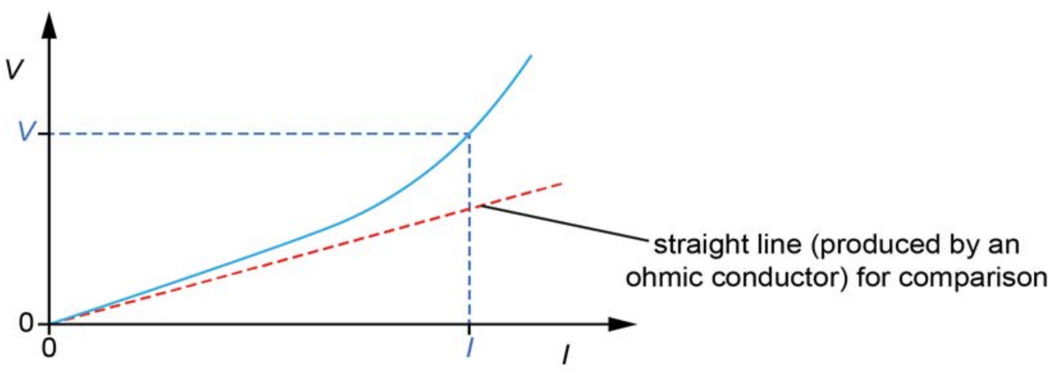


The resistance of the resistor is given by: $\frac{\Delta V}{\Delta I}$

Filament lamp

A filament lamp emits light because its filament becomes very hot. Although the filament is nearly always a metal, because its temperature changes as the current in it changes, the resistance of the filament is not constant.

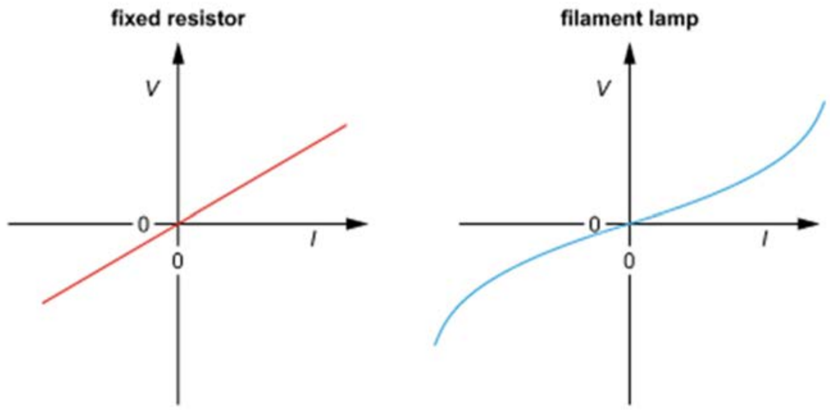
As the temperature of the filament increases, its resistance also increases and therefore the ratio V : I increases and the graph is a curve.



In this case the gradient of this curve is not the resistance. To determine the resistance at any point, the voltage at that point is divided by the current at that point. The resistance of the resistor is given by dividing one value of V by the corresponding value of I: resistance = $\frac{V}{I}$

Some textbooks show I − V graphs instead of V − I graphs; in this case the graph for a filament lamp is a curve with a decreasing gradient, rather than an increasing one.

Both a resistor and a filament lamp behave exactly the same regardless of the direction of the current in them, so full V − I graphs (showing current in both directions) look like this:



Know the properties of NTC thermistors, LDRs and ideal diodes

Thermistor

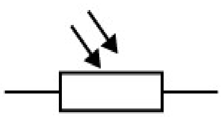
A thermistor is a resistor with a resistance that depends on its temperature. For the common type of thermistor (NTC or negative temperature coefficient type), as its temperature increases, its resistance decreases.

Symbol for a thermistor:

Light dependent resistor (LDR)

An LDR is a resistor with a resistance that depends on the intensity (effectively the brightness) of light that falls on it. As the light intensity increases, the resistance of the LDR decreases.

Symbol for an LDR:

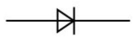


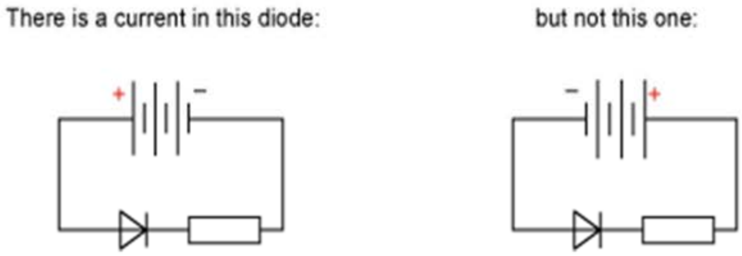
Ideal diode

A diode is a component that only allows a current in one direction. The direction of current allowed is shown by the arrowhead in the circuit symbol for a diode.

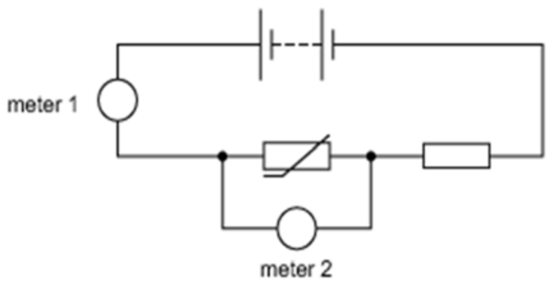
An ideal diode is simply one that behaves exactly as described above; real diodes aren't quite that perfect in practice.

Symbol for a diode:





A circuit includes a thermistor and two meters 1 and 2, connected correctly as shown in the diagram.



The temperature of the thermistor decreases.

What happens to the readings on meters 1 and 2?

(The thermistor is the NTC type.)

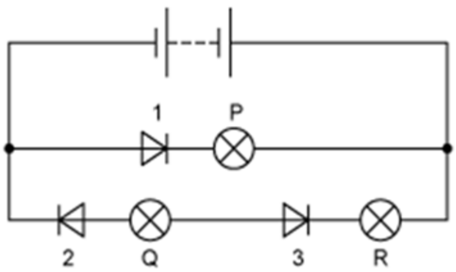
Meter 1 is connected in series with the components and is an ammeter.

Meter 2 is connected in parallel with the thermistor and is a voltmeter.

As the temperature of the thermistor decreases, its resistance increases. This increases the overall resistance of the circuit, reducing the current and therefore decreasing the reading on meter 1 (the ammeter).

The reduced current in the circuit results in there being a smaller voltage across the fixed resistor (from $V = I \times R$, where R is constant). This causes the voltage across the thermistor to increase, because the total voltage supplied by the battery has not changed, so the thermistor now has a larger 'share' of the battery's voltage. Note that this means that the effect of the increased resistance of the thermistor is greater than the effect of the decreased current in it.

A circuit includes three lamps P, Q and R and three diodes 1, 2 and 3, connected as shown in the diagram.



Which lamps are lit?

Lamp P is connected across the battery, but the right-hand side of the battery is positive. This means that diode 1 is reverse biased, which means that it is facing the wrong way to conduct. Its arrowhead is not pointing from the positive terminal of the battery to the negative terminal. Therefore diode 1 does not conduct, and lamp P is not lit.

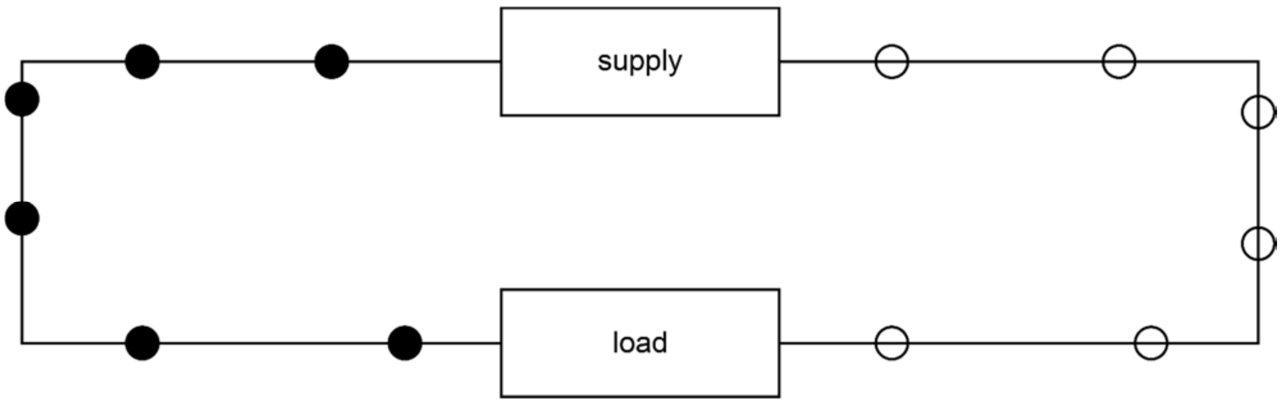
Considering lamp Q, diode 2 is forward biased (facing the right way to conduct). However, diode 3 is facing the wrong way to conduct. As diodes 2 and 3, and lamps Q and R, are all in series, diode 3 prevents any current in this branch of the circuit. Lamps Q and R are not lit.

Know and understand the current and voltage rules for series and parallel circuits

- For components connected in series, the current in each component is the same.
- For components connected in series, the total voltage across the components is the sum of the voltages across each individual component.
- For components connected in parallel, the voltage across each component is the same.
- At each branch in a parallel circuit, the total current moving into the branch is equal to the total current moving out of it.

Basic concepts

An electric circuit, at its simplest, consists of a supply and a load (often a resistor) connected together in a complete loop:



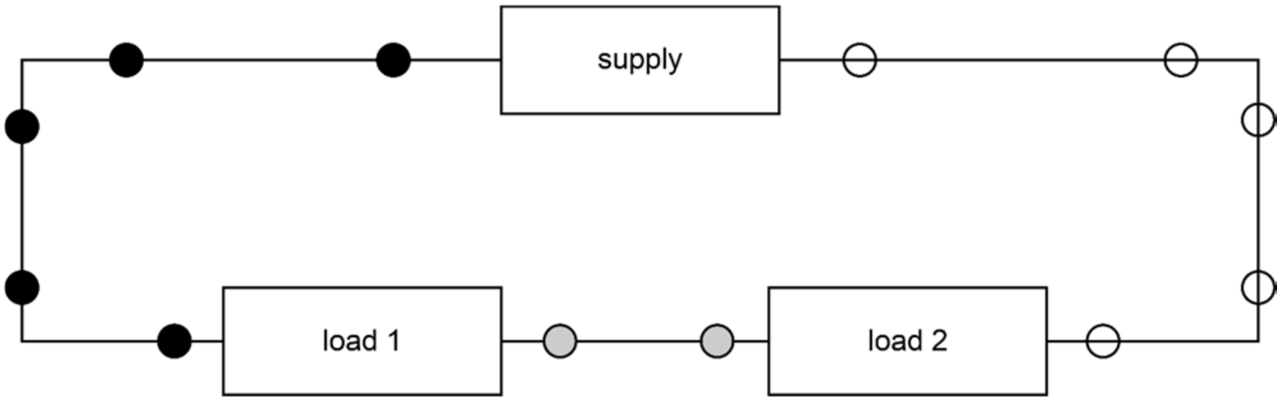
Charged particles (usually electrons) travel around the loop, picking up energy at the supply, carrying out around one side of the loop to the load and then returning to the supply around the other side of the loop.

The rate at which charged particles pass around the circuit (the charge passing a point in the circuit per unit time) is the current.

The difference in energy carried by each unit of charge either side of a circuit component (the energy lost or gained per unit charge) is voltage.

Series circuits

Consider two loads connected in series to the supply:

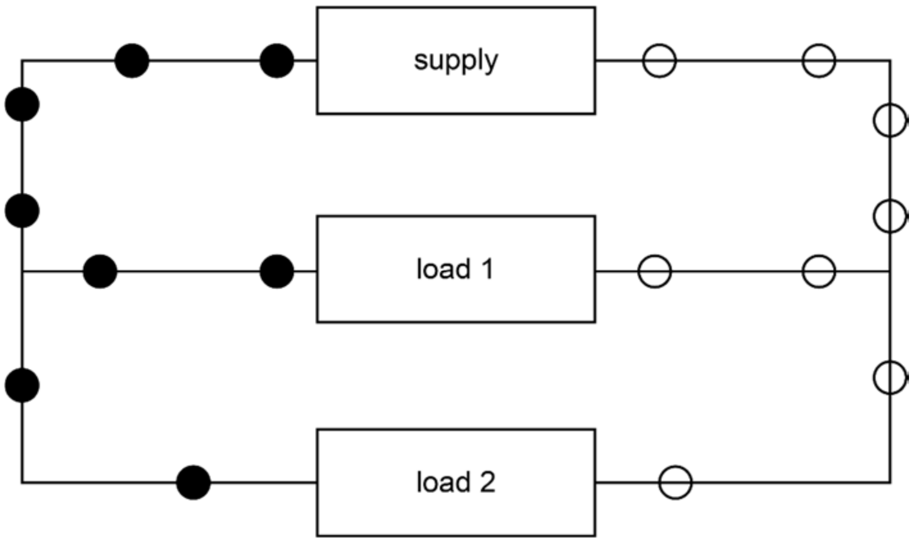


There is only one way around the circuit, so the charge passing any point per unit time (irrespective of how much energy that charge is carrying) will be the same at all points in the circuit. In other words, the current is the same in all parts of the circuit.

However, we can see that, of the total amount of energy picked up by each charge at the supply, some is transferred to load 1 and some to load 2. This means that the energy transferred per unit charge to load 1 plus the energy transferred per unit charge to load 2 is equal to the energy picked up per unit charge at the load. In other words, the voltages across the two loads add up to the supply voltage.

Parallel circuits

Consider two loads connected in parallel to the supply:



When they come to the junction, each charged particle can proceed one way or the other – either through load 1 or through load 2. This means that the total amount of charge passing through load 1 per unit time plus the total charge passing through load 2 per unit time equals the charge passing the supply per unit time. In other words, the currents in loads 1 and 2 add up to the supply current.

However, whichever way each unit of charge proceeds, we can see that it transfers all of its transported energy to the load through which it passes. So, the energy transferred to load 1 by each unit of charge that passes through it is the same as that transferred to load 2 per unit charge and is the same as the energy picked up by each unit of charge at the supply. In other words, the voltages across the supply and the two loads are equal.

Calculate the total resistance for resistor combinations in series

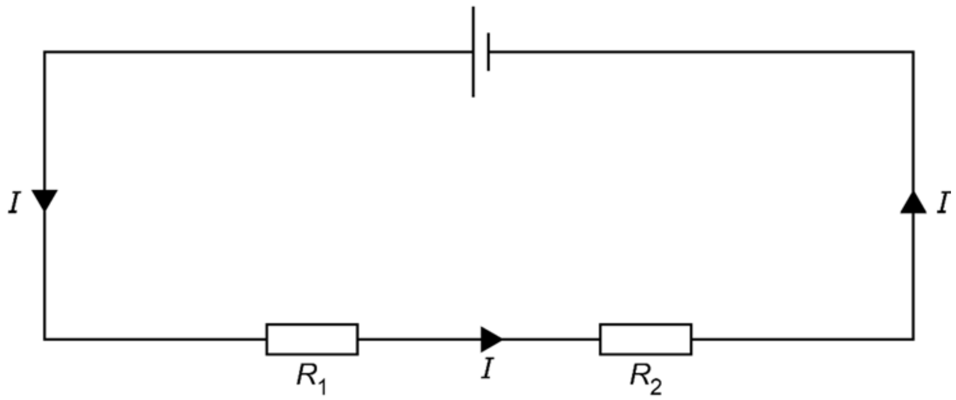
- The combined resistance R_T of two (or more) resistors of resistances R_1, R_2 etc. connected in series is the sum of the individual resistances:

$$R_T = R_1 + R_2 + \ldots\ldots$$

- This means that the combined resistance of two (or more) resistors connected in series is always greater than the resistance of any of the individual resistors.
- For two resistors of resistance R connected in series, the combined resistance will be 2R.

Proof

Consider two resistors, of resistances R_1 and R_2, connected in series to a supply.



We know the current through each resistor is the same, and that it is also the same as the current drawn from the supply. Let's call this current I.

The voltage across R_1 is therefore IR_1, and the voltage across R_2 is IR_2.

We also know that voltages in series add up to the total voltage. So, the supply voltage is $IR_1 + IR_2$.

The combined resistance is therefore supply voltage / supply current = $(IR_1 + IR_2) / I = R_1 + R_2$.

Example

Two resistors, of resistances 2Ω and 3Ω, are connected in series.

Combined resistance = 2 + 3 = 5Ω.

Note that 5Ω is greater than both 2Ω and 3Ω.

If these resistors are connected in series to a 10V supply, then the supply current = 10 / 5 = 2A.

The voltages across the two resistors are therefore 2 × 2 = 4V and 2 × 3 = 6V respectively.

Note that 4V and 6V add up to the supply voltage, 10V.

Understand that the total resistance of a parallel combination is less than that of any individual resistor

- The combined resistance R_T of two (or more) resistors of resistances R_1, R_2 etc. connected in parallel is always less than the resistance of any of the individual resistors.
- For two resistors of resistance R connected in series, the combined resistance is R/2.
- This makes sense if we think of two resistors connected in parallel to a supply. We know that the voltage across both resistors is equal to the supply voltage, and each resistor carries a current given by (I = V /R) caused by that voltage. The total current from the supply is the sum of the currents to each resistor, and so the presence of the two resistors enables a larger current to flow from the supply. Thus, the combined resistance of the circuit is lower than that of either resistor.

Example

Two resistors, of resistances 2Ω and 3Ω, are connected in parallel.

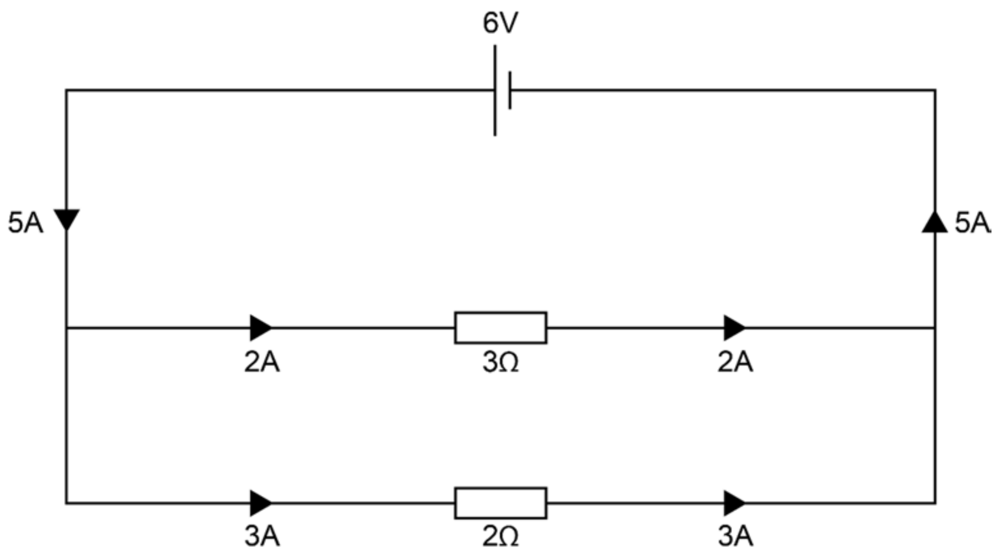
The combined resistance is less than both 2Ω and 3Ω (which means it is less than 2Ω).

(It is actually 1.2Ω).

If these resistors are connected in series to a 6V supply, then the current in the 2Ω resistor is 6 / 2 = 3A.

The current in the 3Ω resistor is 6 / 3 = 2A.

Therefore, total current from the supply = 3 + 2 = 5A.



The current from the supply is greater than that which flows through either resistor, hence the combined resistance is less than the resistance of either resistor.

The combined resistance can be found from supply voltage / supply current = 6 / 5 = 1.2Ω.

Know and be able to apply: $\boldsymbol{voltage = \frac{energy}{charge}}$ $\boldsymbol{V = \frac{E}{Q}}$

- When current flows in a circuit, charged particles carry energy from the supply to the components in the circuit. This means that, when charge passes through a component, some energy is transferred between the charged particles and the component, and the particles have different energy before passing the component compared with after.
- The voltage across the component is the energy transferred per unit charge.
- Mathematically, voltage = energy / charge.
- In symbols, V = E/Q.
- The unit of voltage is therefore the joule per coulomb. This has a special name, the volt (V). $1V = 1JC^{-1}$.
- Just as energy is often measured in units such as kJ, voltages can also be measured in units such as mV and kV. When using the equation V = E/Q, it is important to make sure that the units used for the different quantities are compatible. If in doubt, convert to joules and volts.

Voltage and current are often confused with each other.

Whilst current is the rate at which the charged particles move around the circuit, voltage is to do with the energy that each one transfers when passing through a component. It is this difference that accounts for the way that current and voltages relate to each other in series and parallel circuits (see above).

In a series circuit, each charged particle passes through each component and hence the current is the same in each one. However, each charged particle transfers some energy to each component.

In a parallel circuit, some charged particles move through one component and some move through another. They all move through the supply, and so the supply current is the total of the currents in each branch. However, each charged particle only transfers its energy to the component(s) in the branch through which it moves, and so the energy per unit charge (voltage) is the same for all the components connected in parallel.

Know and be able to apply: power = current × voltage, $P = IV = I^2R$

- We know that current = charge / time. We also know that voltage = energy / charge. Therefore, current × voltage = (charge / time) × (energy / charge) = energy / time.
- However, we also know that energy / time (the rate of transfer of energy) is called power.
- Therefore, electrical power transfer is given by power = current × voltage.
- In symbols, $P = V I$.
- A current of 1A is 1 coulomb per second and a voltage of 1V is 1 joule per coulomb. The power dissipated by 1A at 1V is 1 (coulomb / second) × 1 (joule / coulomb) = 1 joule per second. We know that 1 joule per second is a power of 1 watt. Therefore, in the above equation, when current is in A and voltage is in V, power is in W.
- When using the equation $P = V I$, it is important to make sure that the units used for the different quantities are compatible. If in doubt, convert to amps, volts and watts.
- We also know that resistance = voltage / current ($R = V/I$). From this, $V = IR$. We can substitute $V = IR$ for the 'V' in $P = V I$ to get an expression for power in terms of I and R: $P = (IR)I = I^2R$.
- We therefore have two expressions for electrical power: $P = V I = I^2R$. Both of these are useful, depending on which pieces of information about the circuit are known.

There is a third expression for electrical power that can also be useful.

From $R = V/I$, we know that $I = V/R$.

Substituting this for I in $P = V I$, we get $P = V (V/R) = V^2/R$.

So, the complete set of equations is $P = V I = I^2R = V^2/R$.

Know and be able to apply: energy transfer = power × time, $E = V It$

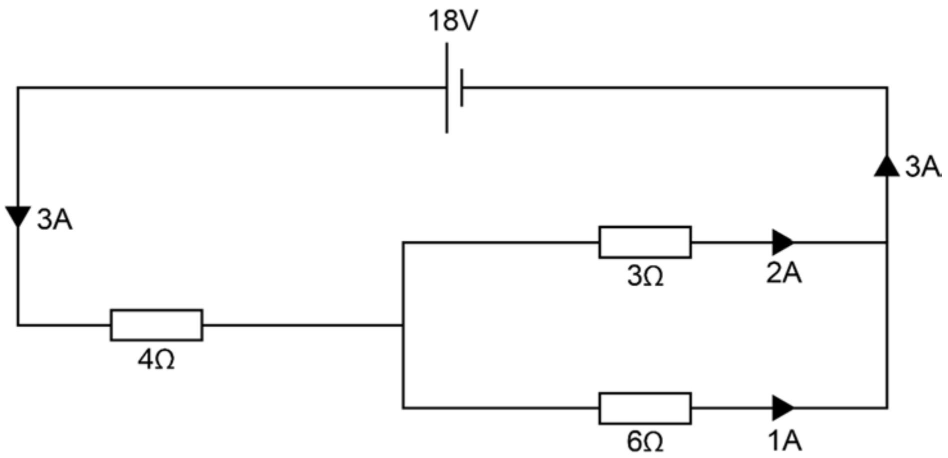
- We know that electrical power is given by $P = V I$. We also know that power is defined as energy per unit time.
- Therefore, energy transferred = power × time = ($V I$) × time.
- It therefore follows that the energy transferred in time t in a component carrying current I and with a voltage V across it is given by energy = $V It$.
- Care should be taken to ensure, when using this equation, that units are compatible. When V is in volts, I is in amps and t is in seconds, energy will be in joules.

Two resistors, of resistances 3Ω and 6Ω, are connected in parallel. The parallel combination is connected in series with a third resistor, of resistance 4Ω, to a supply voltage of 18V. The 4Ω resistor dissipates a power of 36W.

How much energy is dissipated in the 6Ω resistor in 1 minute?

Current in 4Ω resistor = √(power / resistance) = √(36 / 4) = √9 = 3A.

At the junction where the current splits into the two branches, twice as much current flows in the 3Ω resistor as in the 6Ω resistor. Therefore, current in the 3Ω resistor = 2A and current in the 6Ω resistor = 1A.



so voltage across the 6Ω resistor = current × resistance = 1 × 6 = 6V.

(Note, this is the same as the voltage across the 3Ω resistor, 2 × 3 = 6V. This is to be expected, given that the two components are connected in parallel.) ∴ energy transferred in the 6Ω resistor = V It = 6 × 1 × 60 = 360J.

P2. Magnetism

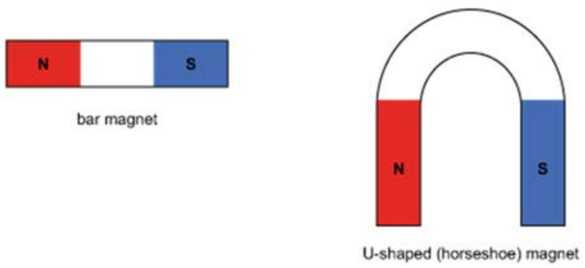
P2.1

Properties of magnets:

a. Know and be able to use the terms *north pole*, *south pole*, *attraction* and *repulsion*.
b. Know the magnetic field pattern around a bar magnet (including direction).
c. Understand the difference between soft and hard magnetic materials (e.g. iron and steel).
d. Qualitatively understand induced magnetism.

Know and be able to use the terms north pole, south pole, attraction and repulsion

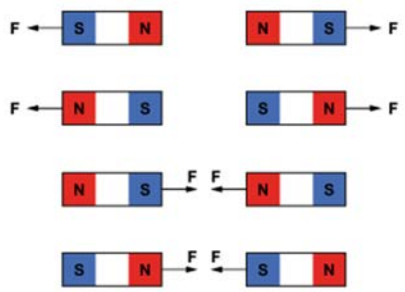
Permanent magnets have two poles, a north pole (N) at one end and south pole (S) at the other. Magnets come in different shapes and sizes, e.g.:



Magnets exert forces on one another when they are placed close together. The nature of these forces can be summarised as:

- Like poles repel (e.g. N and N or S and S).
- Unlike poles attract (e.g. N and S or S and N).

The magnetic forces are shown for pairs of bar magnets in the diagrams below:



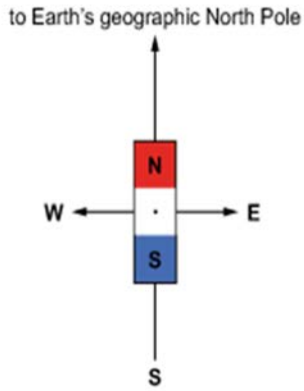
The forces are strongest when the magnetic poles are close together and get weaker as the distance between them increases.

The north or south pole of a magnet will also attract certain magnetic materials, such as iron, cobalt and nickel.

To demonstrate that something is a permanent magnet you need to demonstrate that it can repel another permanent magnet (attraction would only show that it is a magnetic material, not whether it is permanently magnetised).

North-seeking and south-seeking

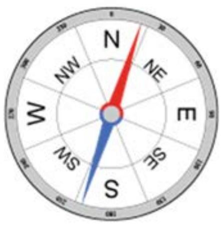
The ‘north’ pole on a bar magnet is really a ‘north-seeking’ pole. This means that it is attracted to the north geographic pole of the Earth. Similarly, a ‘south’ pole on a bar magnet is a ‘south-seeking’ pole and is attracted toward the geographic south pole of the Earth. If a bar magnet is pivoted at its centre so that it is free to rotate, it will align itself in a north-south direction as shown:



The Earth’s north geographic pole is therefore a south-seeking pole and the Earth’s south geographic pole is a north seeking pole!

The magnetic compass

A magnetic compass consists of a bar magnet pivoted about its centre inside a case:



This can be used for navigation. Compasses can also be used to identify the poles of a magnet and to trace the pattern of magnetic fields.

A student is given an unmarked bar magnet and asked to find out which end of the magnet is the north pole.

How could the student use a compass to identify the north pole?

How could the student use another bar magnet, whose poles are marked, to identify the north pole?

Why can't iron filings be used to find the north pole?

The needle in a magnetic compass will line up with the magnetic field of the bar magnet, so it will point away from the north pole and towards the south pole.

Like poles repel and unlike poles attract. The student could move the north pole of the marked bar magnet toward one end of the unmarked one and see if it is attracted or repelled. If it is repelled, then that is the north pole. If it is attracted, then the other end of the unmarked bar magnet is the north pole.

Iron filings line up with the magnetic field lines, but the student would not know which end of the filings was north or south. This means that it would be possible to see the magnetic field pattern, but it would not be possible to distinguish the north and south poles of the bar magnet.

Know the magnetic field pattern around a bar magnet (including direction)

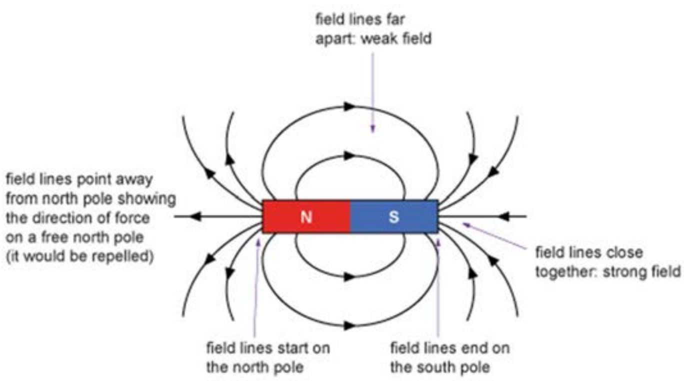
A magnetic field is a region of space in which magnetic forces act on magnets or magnetic materials.

The magnetic field can be represented by drawing magnetic field lines.

Magnetic field lines:

- start on north poles and end on south poles, or form closed loops cannot start or end in space
- cannot cross one another
- point in the direction of force that would be exerted on a free north pole* (north to south)
- are closer together where the field is stronger.

A bar magnet creates a magnetic field in the space that surrounds it:

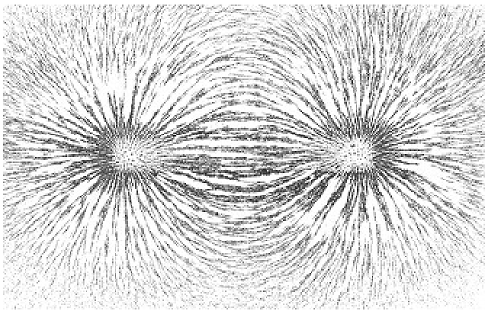


*Free north poles do not actually exist but are useful for this definition.

Using iron filings to show magnetic field patterns

Iron is a magnetic material that is easily magnetised and demagnetised. Iron filings consist of a large number of tiny pieces of iron, each of which becomes magnetised like a tiny bar magnet when placed in a magnetic field. Once magnetised, they line up with the field, so if iron filings are scattered near a magnet, they will tend to line up along the field lines (each tiny bar magnet acting like a compass needle).

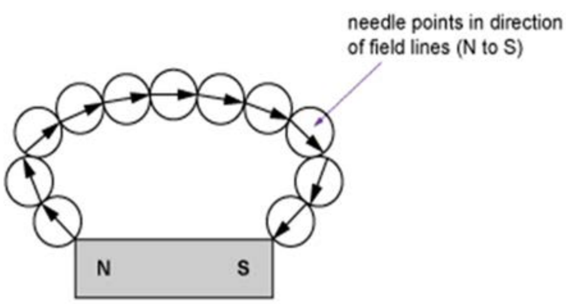
A good way to demonstrate the magnetic field of a bar magnet is to place a piece of white card on top of the magnet and sprinkle iron filings evenly across the card. Then tap the card and the filings will reveal the pattern. Tapping allows them to turn but then they stay in place because of friction with the card. The result looks like this:



Note that this method shows the magnetic field pattern but does not indicate which end of the magnet is north or south.

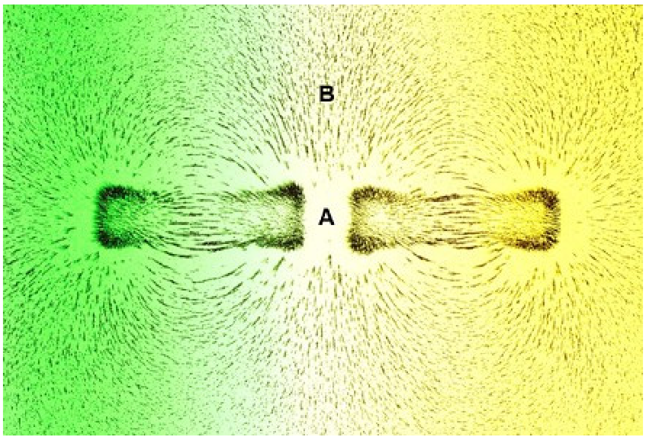
Using a plotting compass

A small magnetic compass can be used to trace field lines in a magnetic field. The compass needle aligns with the magnetic field lines, so it shows the direction of the field at each place. If it is carefully moved through the field, its path traces out the shape of the field lines.



Note that this method shows the direction of the field lines but not the strength of the field.

The image below shows the pattern of iron filings close to two bar magnets.



State whether the force between the two magnets is an attraction or a repulsion. Explain your answer by referring to the magnetic field at points A and B.

For repulsion, the two poles on either side of A must be the same (either both north or both south). We can see that this is the case because the field at A is very weak (no filings), showing that the field lines from each pole are in opposite directions, cancelling out and making the field weak or zero there. We can also see that the field lines from both poles reinforce one another at B, so they must either both point away from the poles (two north poles) or both point toward them (two south poles). The force between the magnets is a repulsion.

Understand the difference between soft and hard magnetic materials (e.g. iron and steel)

Some materials, e.g. iron, steel, cobalt and nickel, are affected by magnetic forces and can become magnetised when they are placed in a magnetic field. The ease with which they become magnetised and lose their magnetisation varies from material to material:

- Soft magnetic materials – are easy to magnetise but also easily lose their magnetisation.
- Iron is an example of a 'soft' magnetic material.
- Hard magnetic materials – are difficult to magnetise but once they are magnetised, they are difficult to demagnetise.
- Steel is an example of a 'hard' magnetic material.

Note that being a soft or hard magnetic material has nothing to do with mechanical hardness!

Using soft and hard magnetic materials

Electromagnets usually consist of a coil of many turns wound around a soft iron core. A soft magnetic material is used because it gains and loses its magnetisation very quickly. When there is an electric current in the coil, it creates a magnetic field that rapidly magnetises the core. The combined effect of the magnetic field from the current and the much stronger magnetic field from the magnetised core add together to create the field of the electromagnet. When the current is switched off, the core rapidly loses its magnetisation and the magnetic field of the electromagnet switches off.

Permanent magnets are made from hard magnetic materials (like steel). The advantage of using a hard magnetic material is that once magnetised, it will remain magnetised for a long time. However, even a hard magnetic material will eventually become demagnetised. Heating the magnet or hitting it can also cause demagnetisation.

Hard and soft magnetic materials have different uses.

Which type of material is suitable for each of the uses below and why?

The needle of a magnetic compass.

The core of an electromagnet.

The magnet inside a loudspeaker.

A refrigerator magnet.

Hard. The needle needs to retain its magnetism for a long time, otherwise it could not be used for navigation.

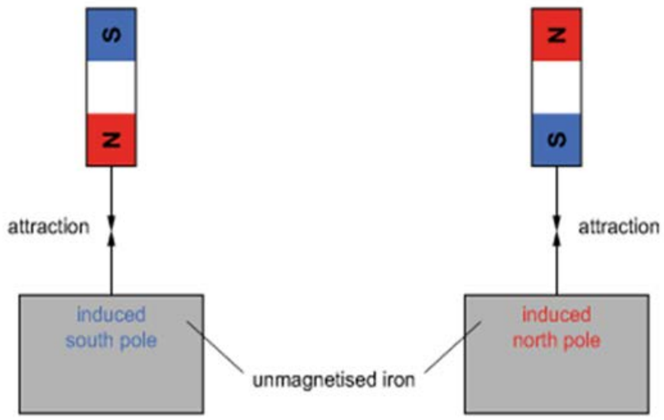
Soft. The core is magnetised when there is current in the coil surrounding it. If it retained its magnetisation, it could not be switched off.

Hard. The loudspeaker magnet must retain its magnetism, otherwise the forces on the loudspeaker coil will gradually get weaker and the speaker will become less effective.

Hard. The magnet must retain its magnetisation, otherwise the force of attraction to the refrigerator door will weaken and it will fall off.

Qualitatively understand induced magnetism

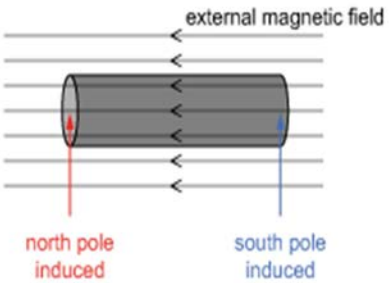
When either pole of a bar magnet is held close to an unmagnetised magnetic material (e.g. iron, steel, nickel or cobalt), there is a force of attraction between the magnet and the material. There is never a force of repulsion.



The magnetic field of the bar magnet has induced magnetism in the piece of iron.

Opposite magnetic poles attract, so the effect of bringing a magnetic pole close to the iron is to induce an opposite magnetic pole on the nearest part of the iron. If a north pole approaches the iron, then the induced pole is a south pole, and vice versa.

Induced magnetism is used to make permanent magnets. If an unmagnetised sample of a hard magnetic material is placed in a strong magnetic field (e.g. inside a solenoid), then magnetic poles are induced onto the ends of the sample:



Once magnetised, the hard magnetic material retains its magnetisation and can be used as a bar magnet.

You are given two metal cylinders, X and Y. When any pair of ends of these two cylinders are brought close together, they attract one another.

Which of the following statements could be true?

1. Both X and Y are bar magnets.
2. X is a bar magnet and Y is unmagnetised iron.
3. Y is a bar magnet and X is unmagnetised iron.
4. Both X and Y are unmagnetised iron.

When a bar magnet is brought close to a piece of unmagnetised iron, it induces an opposite magnetic pole on the surface of the iron and there is a force of attraction between the two opposite magnetic poles. This could happen if X is a magnet and Y a piece of unmagnetised iron, or if Y is a magnet and X a piece of unmagnetised iron. Therefore statements 2) and 3) could be true.

If both X and Y were bar magnets, then they would repel whenever their two like poles were brought together. This does not happen, so statement 1) cannot be true.

If both X and Y were unmagnetised iron, then there would be no magnetic forces. Since the cylinders do exert magnetic forces on one another, statement 4) cannot be true.

Photos: Science Photo Library

P2.2

Magnetic field due to an electric current:

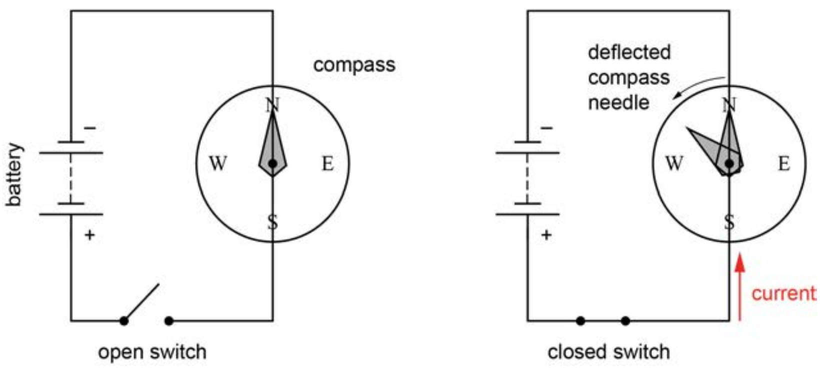
a. Know and understand the magnetic effect of a current.
b. Know the magnetic field patterns around current-carrying wires (including direction) for straight wires and coils/solenoids.
c. Know and understand the factors affecting magnetic field strength around a wire.
d. Understand the difference between permanent magnets and electromagnets.

Know and understand the magnetic effect of a current

Electric currents create magnetic fields in the surrounding space.

This can be demonstrated by placing a small magnetic compass close to a current-carrying conductor and then switching the current on and off. The compass needle will point north when the current is off and will deflect from north when the current is on, showing that the current has created a magnetic field around the conductor.

This is shown in the two diagrams below. The compass has been placed under the conducting wire and the circuit has been placed so that the wire runs north-south.



Factors affecting the magnetic field created by an electric current

- Reversing the direction of the current reverses the deflection of the compass needle, showing that the direction of the magnetic field depends on the direction of the current.
- Increasing the current increases the deflection of the compass needle, showing that the strength of the magnetic field depends on the size of the electric current.

Magnetic fields and moving charges

Electric current consists of moving electric charges. The magnetic field is created by these moving charges and not by the material through which they are moving (e.g. a copper conductor). A beam of charged particles (e.g. electrons or ions) moving through a vacuum will also create a magnetic field, just like an electric current in a wire.

A teacher demonstrates the magnetic effect of an electric current using iron filings. To do this, he places a piece of white card perpendicular to a long straight wire such that the wire passes through the centre of the card.

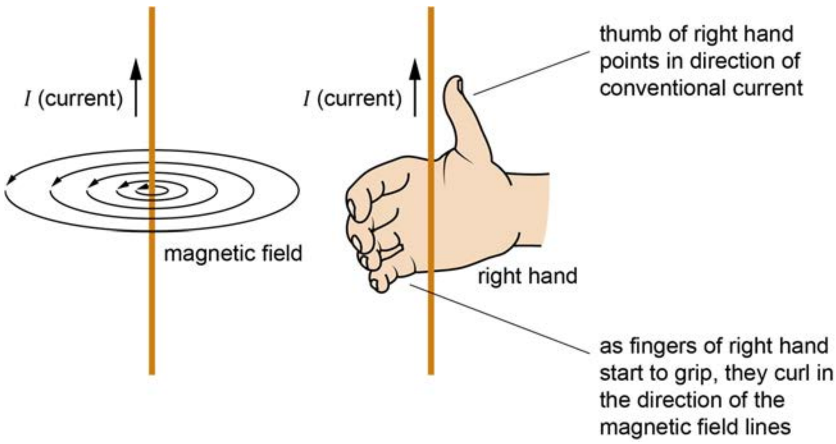
Describe the pattern that forms when there is a current in the wire.

The iron filings form concentric rings around the wire. The rings are closest together near the wire (where the magnetic field is strongest).

Know the magnetic field patterns around current-carrying wires (including direction) for straight wires and coils/solenoids

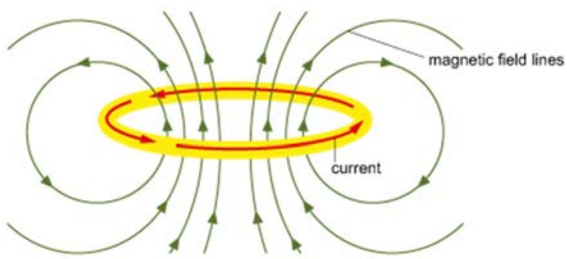
The magnetic field pattern around a long, straight current-carrying wire:

- Consists of concentric circles
- Circles are further apart the greater the distance from the wire
- Direction of field can be predicted using the right-hand grip rule



Magnetic field pattern created by a narrow coil

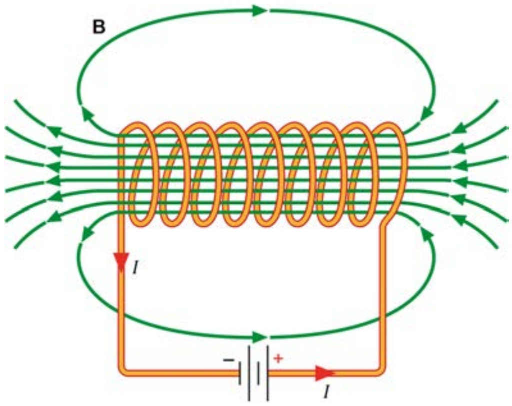
When current-carrying wires are wound into a coil, the magnetic field created by each part of the coil adds together creating a strong field through the centre of the coil. The pattern for a narrow coil consisting of just one turn looks like this:



The direction of the magnetic field through the coil can be predicted by using the right-hand grip rule at any point on the coil.

Magnetic field pattern created by a long coil or solenoid

A long coil, or solenoid, consists of many narrow coils wound close together. The magnetic field created by each of the narrow coils adds together to create a very uniform field through the centre of the solenoid. The resultant pattern is shown below:



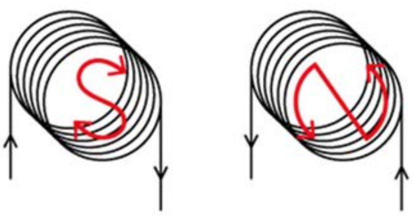
Key points:

- The direction of the field can be predicted using the right-hand grip rule.
- The magnetic field pattern is similar to that of a bar magnet.
- One end of the solenoid (where field lines emerge) acts as a magnetic north pole and the other end (where the field lines enter the solenoid) acts like a magnetic south pole.
- The field at the sides of the solenoid is weak and in the opposite direction to the field inside the solenoid.
- The field inside the solenoid is uniform (constant strength and direction) in the centre and through most of the coil, decreasing at the ends.
- The field can be controlled by controlling the current.

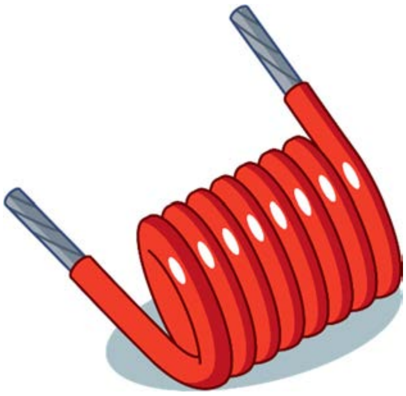
Another way to identify which end of a solenoid is the north pole is to note whether the current circulates clockwise or anticlockwise when you look at the end of the coil:

- clockwise = south pole
- anti-clockwise = north pole.

These directions can also be linked to the letters S and N to give an easy way to identify the poles:



The coil shown is made from insulated wire. The ends of the wire are connected to a d.c. power supply so that A is connected to the positive terminal and B to the negative terminal.



Which end of the coil acts like a north pole?

Current goes from positive to negative, so the current in the coil is from A to B. The right-hand grip rule can be used to find the direction of the magnetic field close to the end of the coil. At the end closest to A, the field lines are coming out of the coil, so this end acts like a north pole.

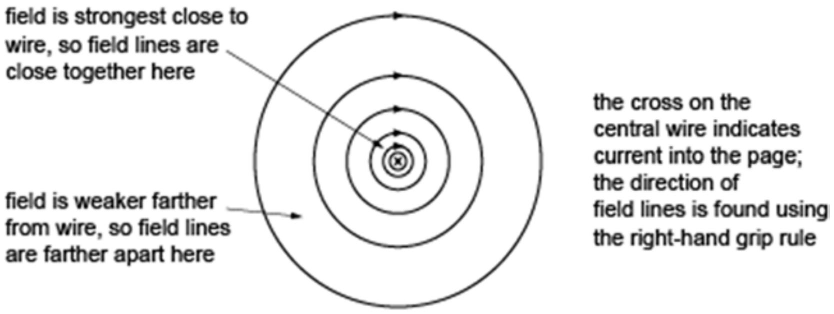
An alternative method is to imagine looking at the end near A and observing the direction of rotation of current in the coil. It is anticlockwise, so that end acts like a north pole.

Know and understand the factors affecting magnetic field strength around a wire

The strength of the magnetic field around a wire depends on:

- the current in the wire: increasing current increases magnetic field strength
- the distance from the wire: farther from the wire the field is weaker
- the medium surrounding the wire: magnetic media such as iron can increase the field strength.

The diagram below shows the magnetic field created by a long, straight wire carrying current into the page.



Effect of iron on magnetic field strength

Iron is a ferromagnetic material. Each iron atom acts like a tiny bar magnet being north at one end and south at the other. When an external magnetic field from a current-carrying wire or coil passes through a ferromagnetic material, the atomic magnets can line up with the external field to create a much stronger resultant field. This is why iron cores are often used in electromagnetic devices such as motors, generators and transformers.

Effect of an iron core inside a coil

Electromagnets use long coils, called solenoids, to create a strong magnetic field. The strength of the magnetic field inside and at the ends of the solenoid is increased by:

- increasing the number of turns in the same length of solenoid (turns per unit length)
- using a soft iron core inside the coil increasing the current in the coil.

Overhead power cables carry high voltage a.c. currents. People living close to the cables are exposed to small magnetic fields.

Which of the following statements about these fields is/are correct?

1) People further from the cable experience a weaker magnetic field.

2) The strength of the magnetic field directly beneath the cable has constant magnitude.

3) The direction of the magnetic field created by the cable regularly changes its direction.

4) The magnetic field is created by moving electrons in the cable.

5) The strength of the magnetic field depends on the magnitude of current in the cable.

When there is an electric current in a wire, it creates a magnetic field around the wire. The greater the current, the stronger the magnetic field, so statement 5) is correct.

The current consists of a flow of electrons inside the wire, so statement 4) is correct.

Farther from the wire the field lines are farther apart so the magnetic field is weaker there. Statement 1) is correct.

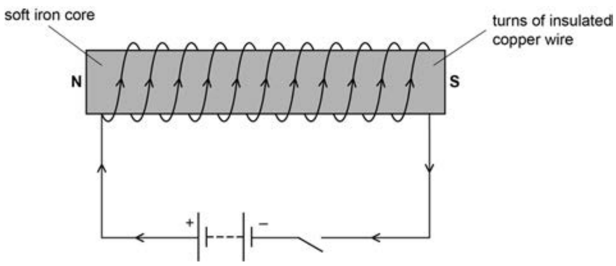
In this case the current alternates, so its direction continually changes. This creates a magnetic field which also continually changes its direction, so statement 3) is correct.

An alternating current varies in magnitude so the field it creates also varies in magnitude, so statement 2) is incorrect.

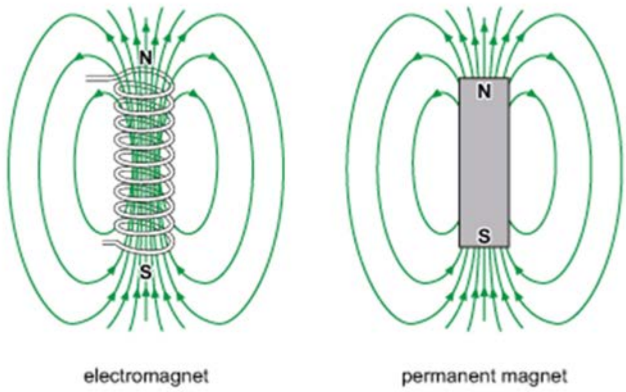
Therefore, all statements except statement 2) are correct.

Understand the difference between permanent magnets and electromagnets

An electromagnet consists of a long coil or solenoid wound around a core made from a soft magnetic material such as iron.



The magnetic field pattern created by an electromagnet is like that of a permanent bar magnet:



Differences between electromagnets and permanent magnets

Despite the similarities, there are also important differences between electromagnets and permanent magnets. These are summarised in the table below.

	Electromagnet	Permanent magnet
Action as a magnet	Can be switched on/off	Continuous
Strength of magnetic field	Can be varied by changing the current in the coil	Constant (but may decay with time)
Polarity of magnet	Can be reversed by reversing current direction	Constant
Materials used	Core made from soft magnetic material (e.g. iron) so that it can magnetise and demagnetise quickly	Magnet made from hard magnetic material (e.g. steel) so that once magnetised, it remains magnetised.

Permanent magnets

Hard magnetic materials are difficult to magnetise and demagnetise. A permanent magnet can be made by placing a hard magnetic material in a strong external magnetic field, usually from an electromagnet. Suitable materials for permanent magnets are alloys of iron, nickel or cobalt or rare earth metals and certain ceramics. Neodymium magnets are now quite common and can be used to make very strong permanent magnets.

Whilst these magnets do retain their strength for a long time, they can be weakened by sudden impacts and lose their magnetisation completely if heated above a certain temperature (the Curie temperature, which is specific to each material).

Electromagnets

The strength of an electromagnet can be increased by increasing the number of turns per unit length on the solenoid and by increasing the current in the wires. However, for normal conducting wires, this results in a strong heating effect that could melt the insulation, so there is a limit to how much current can be passed through the coil. The strongest electromagnets avoid this problem by using superconducting coils, that is coils of zero resistance. Such superconducting electromagnets are used in the Large Hadron Collider at CERN. The disadvantage is that the coils only become superconducting when cooled to extremely low temperatures using liquid helium.

The core of an electromagnet is made from a soft magnetic material. Why would a hard magnetic material be unsuitable for this purpose?

Hard magnetic materials are difficult to magnetise so they would not create a strong magnetic field when the current is switched on. Once magnetised, they retain their magnetisation so it would not be possible to switch the electromagnet off.

P2.3

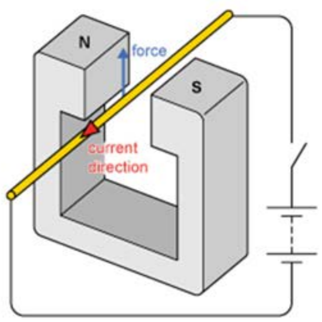
The motor effect:

a. Know that a wire carrying a current in a magnetic field can experience a force.
b. Know the factors affecting the direction of a force on a wire in a magnetic field (including the left-hand rule).
c. Know the factors affecting the magnitude of the force on a wire in a magnetic field.
d. Know and be able to apply $F = BIL$ for a straight wire at right angles to a uniform magnetic field.
e. Know and understand the construction and operation of a dc motor, including factors affecting the magnitude of the force produced.
f. Understand applications of electromagnets.

Know that a wire carrying a current in a magnetic field can experience a force

When a current carrying wire passes through a magnetic field such that its direction crosses the field lines, there is a force on the wire. This is called the motor effect.

The motor effect can be demonstrated by holding a straight wire between the poles of a magnet and momentarily switching the current in the wire on and off, as shown in the diagram.



When the current is switched on, the wire experiences a force and is pushed out of the field.

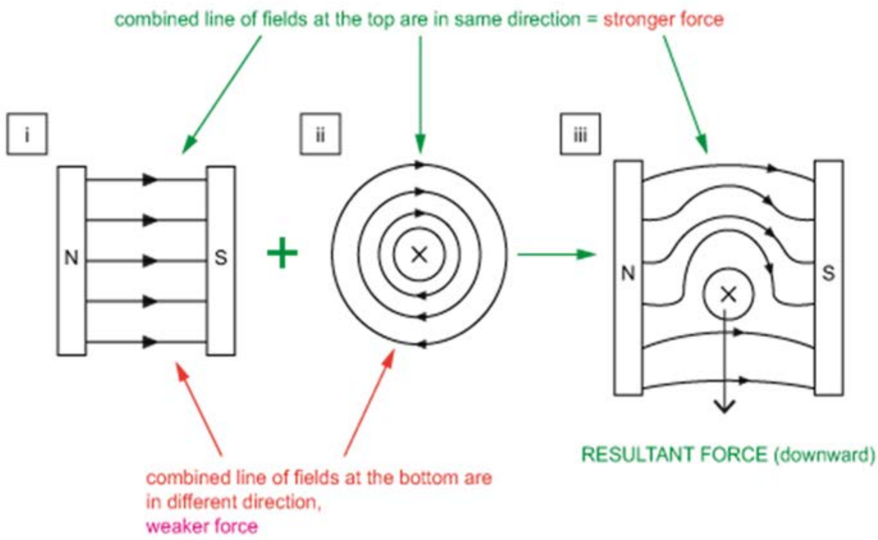
Note that the wire is *not* attracted directly toward or repelled directly away from the magnetic poles. The force acts in a direction perpendicular to both the current and the magnetic field.

If the direction of the current is parallel to the magnetic field, there is no motor effect force.

The force is a maximum when the current and magnetic field are at right angles to one another.

The origin of the motor effect force

Electric currents create magnetic fields around the wires that carry them. When the magnetic field from an electric current interacts with an external magnetic field, e.g. from a permanent magnet, the two magnetic fields interact. This creates a force on both the wire carrying the current and on the permanent magnet. The diagrams below show (i) a uniform horizontal magnetic field created by two permanent magnets, (ii) the magnetic field created by a wire carrying electric current into the page, and (iii) the combined magnetic field when the wire is placed between the two permanent magnets.



The resultant force on the wire is from the region of stronger field, above the wire, toward the region of weaker field, below the wire. Imagine that the field lines are like stretched elastic threads trying to contract. As they do so, they push the wire downwards. There is an equal and opposite upward force on the permanent magnets.

Moving charges

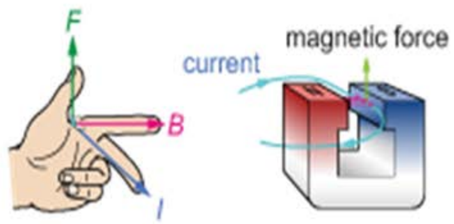
A beam of charged particles (or just one moving charged particle) is also an electric current, so the motor effect force also acts on moving charges and can be used to deflect them.

Know the factors affecting the direction of a force on a wire in a magnetic field (including the left-hand rule) Direction of the motor effect force

The motor effect force is always perpendicular to the directions of both the current and the magnetic field.

The direction of the motor effect force can be predicted using Fleming's left-hand rule:

Fleming's left-hand rule



- *F*: direction of motor effect force
- *B*: direction of magnetic field (N to S)
- *I*: direction of conventional current (+ to –)

The first finger points in the direction of the magnetic field (N to S).

The second finger points in the direction of conventional current (+ to −).

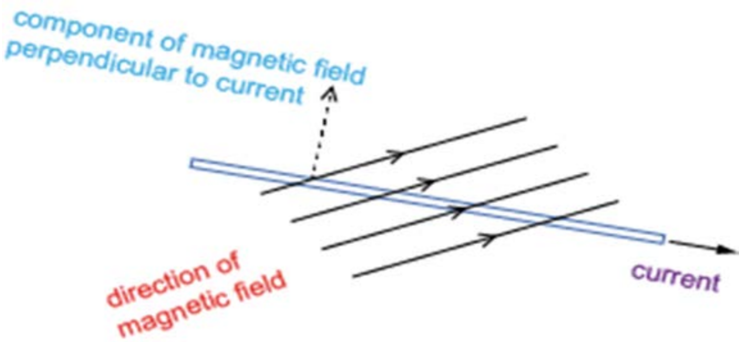
The thumb points in the direction of the motor effect force.

Reversing the direction of the current OR the magnetic field reverses the direction of the motor effect force.

Reversing the directions of *both* the current *and* the magnetic field results in no change in the direction of the motor effect force.

Getting the directions right

When using the left-hand rule, the direction of first finger is the same as the direction of the magnetic field acting perpendicular to the current. If the magnetic field and current are not at 90°, then the direction to use is the part (component) of the magnetic field that is perpendicular to the current, as shown in the diagram:

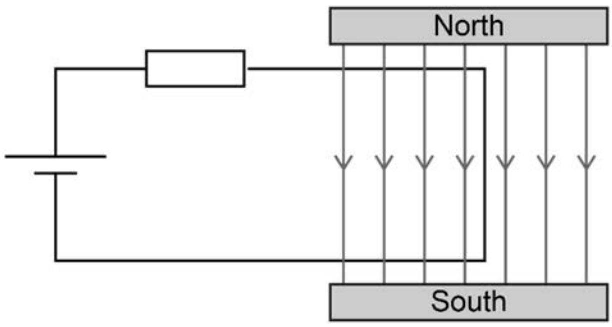


In the example above, the motor effect force is directly out of the page.

If the current is parallel to the magnetic field, then there is no part of the magnetic field perpendicular to the current and no motor effect force. When charged particles move parallel to the magnetic field lines they are not deflected.

Part of an electric circuit passes through a region of uniform magnetic field between the poles of a permanent magnet as shown.

a)



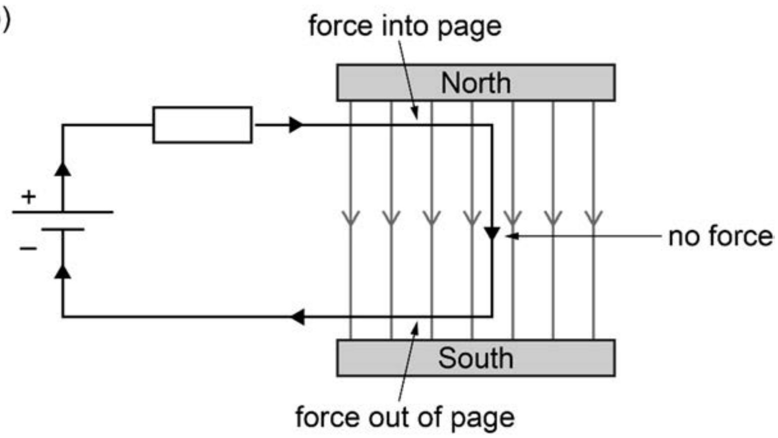
Which parts of the wire will experience a force from the magnetic field?

In which directions will these forces act?

The motor effect force acts on currents that are perpendicular to a magnetic field. There is a current in the circuit and the wires at the top and bottom are perpendicular to the field, so these both experience a motor effect force. The wire at the right-hand end of the circuit is parallel to the magnetic field lines so there is no force on this wire.

The motor effect force acts perpendicular to the current (positive to negative) and magnetic field (north to south) and its direction is given by Fleming's left-hand rule. The directions of the forces are shown:

b)



Know the factors affecting the magnitude of the force on a wire in a magnetic field

Magnitude of the force on a wire in a magnetic field

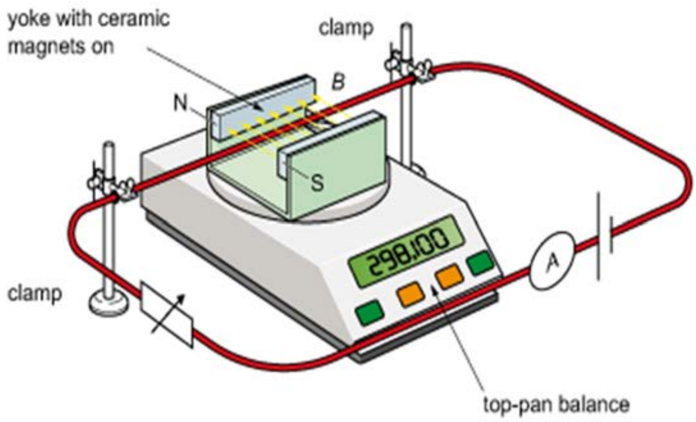
The force on a current-carrying wire in a magnetic field depends on:

- magnitude of the current: the greater the current the greater the force
- strength of the magnetic field: the greater the magnetic field strength the greater the force
- length of wire in the magnetic field: the greater the length in the field the greater the force
- angle between magnetic field and current: force is greatest at 90° and zero at 0°.

The direction of the force is given by Fleming's left-hand rule.

Investigating the strength of the motor effect force

The motor effect force can be investigated in an experiment like the one shown below. When there is a current in the wire there is an upward motor effect force on the wire. By Newton's third law, there is an equal downward force on the magnets. The magnets press down on the top pan balance and the reading increases. This change can be used to measure the motor effect force (using *W* = *mg*).



Increasing the current increases the force. Replacing the magnet with a stronger one of the same length increases the force. Reducing the length of wire in the magnetic field (e.g. by placing the ceramic magnets vertically in the yoke) reduces the force.

Know and be able to apply F = BIL for a straight wire at right angles to a uniform magnetic field

Equation for the motor effect force

The magnitude of the force F on a current-carrying wire at right angles to a uniform magnetic field depends on three factors:

- the magnitude of the current, I
- the strength of the magnetic field, B
- the length of wire at right angles to the field, L

and is given by the equation:

$$F = BIL$$

- F is the motor effect force in newtons (N).
- B is the magnetic field strength in tesla (T).
- I is the current in the wire in amps (A).
- L is the length of wire at 90° to the magnetic field in metres (m).

The direction of the force is given by the left-hand rule.

Unit of magnetic field strength: the tesla (T)

The formula for magnetic force can be rearranged to give:

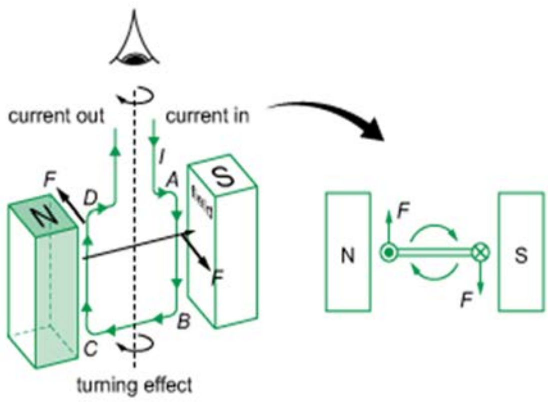
$$B = \frac{F}{IL}$$

This shows that magnetic field strength is 'force per unit current-length' and that a magnetic field of strength 1 T exerts a force of 1 N on a wire of length 1 m carrying a current of 1 A (when the wire is perpendicular to the field).

In terms of units: $1T = 1\ Nm^{-1}A^{-1}$

Turning effect on a rectangular coil in a magnetic field

One of the most important applications of the motor effect is in the dc motor. This uses the motor effect to create a turning effect (moment) on a current-carrying coil. The turning effect on the coil arises from a pair of motor effect forces acting in opposite directions on either side of the coil.



Wires AB and CD carry currents in opposite directions perpendicular to the same uniform magnetic field. The two forces are in opposite directions and are separated by the width of the coil (AD) so there is a resultant turning effect.

- Reversing the current reverses the direction of the turning effect.
- Increasing the current in the coil or the number of turns on the coil will increase the turning effect.

A straight wire carries a current I perpendicular to a uniform magnetic field of strength B and a motor effect force F is exerted on the wire. The length of wire perpendicular to the field is l. Complete the table to show corresponding values for F, B I and l.

F /N	B /T	I/A	l /m
(a)	0.20	4.0	0.050
0.10	0.40	(b)	0.20
0.050	0.10	20	(c)
0.50	(d)	10	0.050

F = BIl = 0.20 × 4.0 × 0.050 = 0.040 N

I = F/Bl = 0.10 / 0.40 × 0.20 = 1.25 A

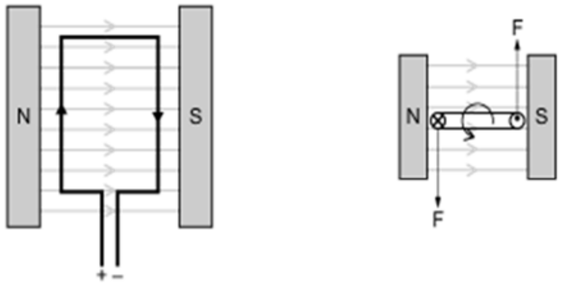
l = F/BI = 0.050 / 0.10 × 20 = 0.025 m

B = F/Il = 0.50 / 10 × 0.050 = 1.0 T

Know and understand the construction and operation of a dc motor, including factors affecting the magnitude of the force produced

Turning effect on a coil in a magnetic field

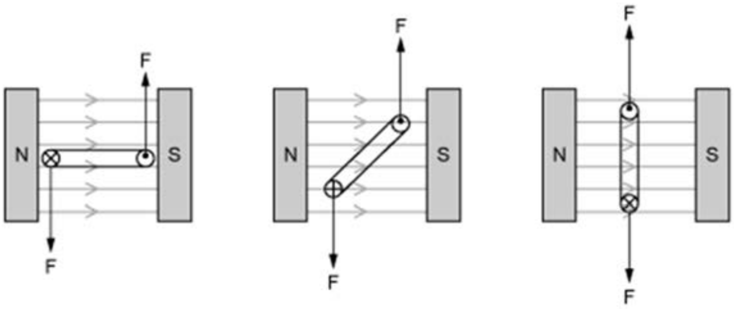
When a current-carrying rectangular coil is placed in a uniform magnetic field, the motor effect forces on either side of the coil can produce a turning effect on the coil.



view from above with coil in plane of magnetic field

view from end with coil in plane of magnetic field

As the coil turns inside a uniform magnetic field, the distance between the two motor effect forces changes and so the turning effect changes. It is at its maximum when the two forces are farthest apart (coil in plane of field) and zero when the two forces are in the same vertical plane (coil perpendicular to field).



maximum turning effect

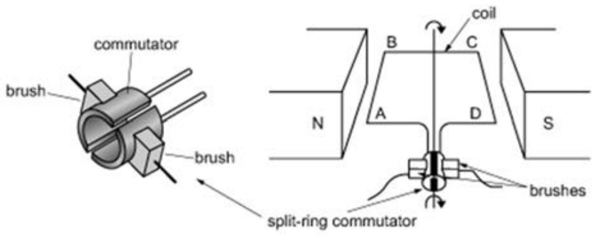
reduced turning effect

zero turning effect

If the coil continues to rotate past the vertical position, the turning effect changes direction, returning the coil to the vertical position.

A simple dc motor

To make the coil rotate continuously in the same direction, the current direction in the coil must be reversed every time the coil passes the vertical position. This is done by using a split-ring commutator. The commutator rotates with the coil and is connected to the dc power supply by two brushes. It acts as a rotating switch, reversing the connections to the coil every half rotation.

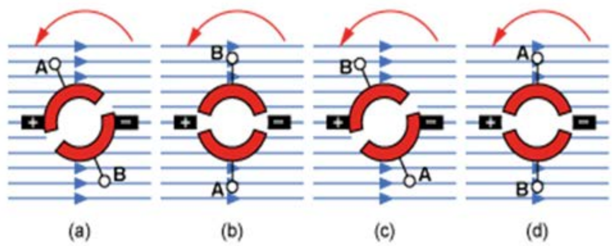


Factors affecting the turning effect on the coil in a dc motor

- Current in the coil: greater current creates a greater turning effect (achieved by increasing the voltage of the supply).
- Size of the coil: the larger the area of the coil, the greater the turning effect.
- Magnetic field strength: a stronger magnetic field creates a greater turning effect.
- Number of turns on the coil: the larger the number of turns, the greater the turning effect.
- Winding the coil onto a soft iron core: this increases the magnetic field strength and increases the turning effect.
- Angle of coil in field: maximum turning effect when the coil is in the plane of the field and zero when perpendicular to the field.

The split-ring commutator: how it works

The sequence of diagrams below shows how the rotation of the split-ring commutator switches the current direction in the coil every half-rotation. The two brushes are connected to the external dc power supply and make a sliding electrical contact with the surface of the commutator.



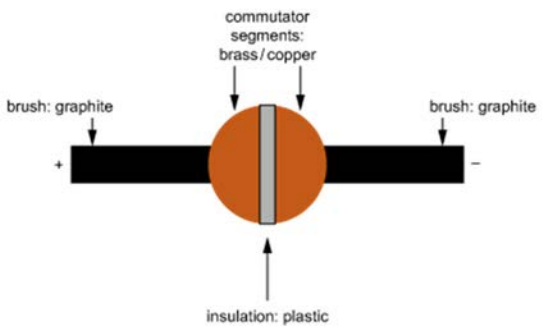
Momentum of rotating coil carries it past this position so that sides A and B switch polarity.

Side A of coil is connected to positive terminal and has downward force. Side B is connected to negative terminal and has upward force.

Momentum of rotating coil carries it past this position so that sides A and B switch polarity.

Side B of coil is connected to positive terminal and has downward force. Side A is connected to negative terminal and has upward force. Coil continues to rotate in same direction.

The commutator segments are usually made from brass or copper, and the brushes are usually made from graphite. The graphite brushes make a low friction sliding contact with the surface of the commutator but maintain good electrical contact, allowing current to enter and leave the coil. Over time the brushes wear down and must be replaced.



Which of the following statements about a simple dc motor is/are correct?

1) The current in the coil changes direction once per rotation.

2) The material used to make the brushes must be a conductor.

3) The forces on each side of the coil are caused by the motor effect.

4) The forces on each side of the coil act in opposite directions.

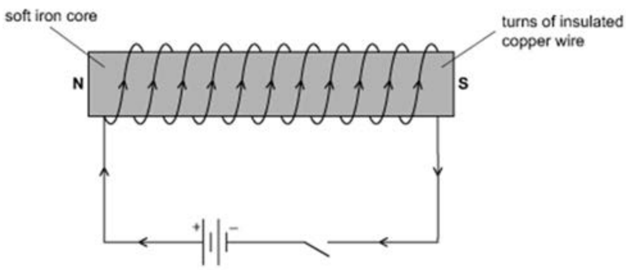
5) The motor will spin faster if the current or magnetic field is increased.

The current in the coil changes direction every half turn, so it changes twice per rotation, not once. Therefore statement 1) is incorrect. The other statements are all correct.

Understand applications of electromagnets

Structure of an electromagnet

An electromagnet consists of many turns of insulated wire wound onto a soft iron core. When there is current in the coil, the ends of the electromagnet act as north and south magnetic poles creating an external magnetic field like that of a bar magnet.

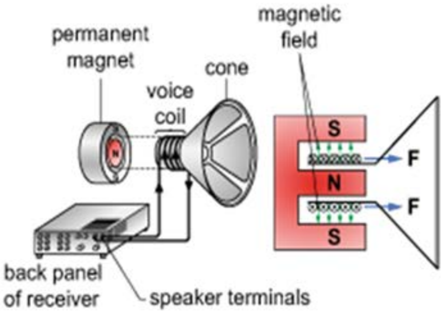


The key features of an electromagnet are:

- it can be switched on or off by switching the current on or off
- its strength can be varied by varying the magnitude of the current in the coil
- its polarity can be reversed by reversing the direction of the current.

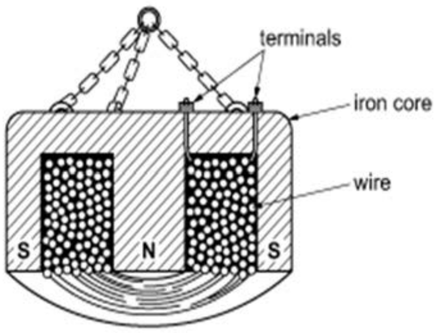
Example of application: loudspeaker

A voice coil is connected to a movable speaker cone and placed in the radial magnetic field created by a permanent magnet. When there is current in the voice coil, a motor effect force pushes it forwards or backwards and the cone creates compressions and rarefactions in the air, creating a sound.



Example of application: lifting magnets

Large electromagnets are used to lift heavy objects made of magnetic materials (e.g. cars). The coil consists of many turns of thick insulated copper wire and there is a large current in the coil when it is switched on. The soft iron core intensifies the field. Lifted objects can be released by switching off the current.



State and explain two advantages of an electromagnet compared with a permanent magnet when used to lift and move heavy objects.

Electromagnets can be turned on and off so that the lifted object can be picked up and released. It would be difficult to remove the object from a permanent magnet.

P2.4

Electromagnetic induction:

a. Know and understand that a voltage is induced when a wire cuts magnetic field lines, or when a magnetic field changes.
b. Know the factors affecting the magnitude of an induced voltage.
c. Know the factors affecting the direction of an induced voltage.
d. Understand the operation of an ac generator, including factors affecting the output voltage.
e. Interpret the graphical representation of the output voltage of a simple ac generator.
f. Understand applications of electromagnetic induction.

Know and understand that a voltage is induced when a wire cuts magnetic field lines, or when a magnetic field changes

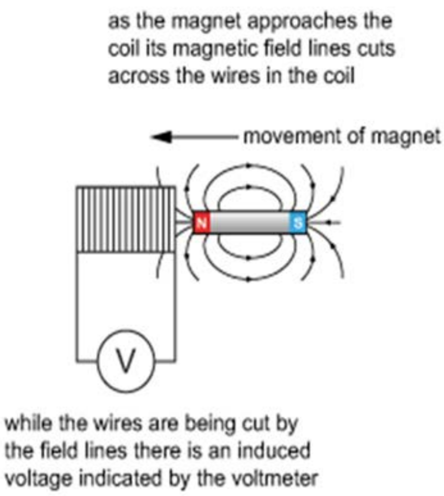
A voltage is induced in a conductor when:

- it cuts across the lines of a magnetic field
- the magnetic field passing through it changes.

This is called electromagnetic induction.

Electromagnetic induction can be shown by inducing voltages in a coil.

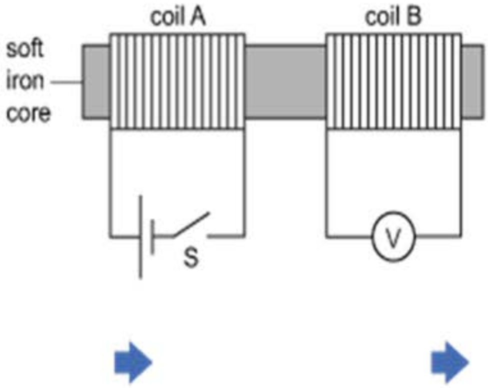
A voltage is induced in a coil when the wires in the coil cut the magnetic field lines



There is no induced voltage when there is no relative motion between the magnet and the coil.

It does not matter whether the coil or the magnet moves; if magnetic field lines cut the wires in the coil, then there will be an induced voltage in the coil.

- the magnetic field through the coil changes



As switch S is closed, there is an increasing magnetic field through coil A.

The soft iron coil carries the changing magnetic field through coil B.

There is an induced voltage in coil B while the magnetic field through it changes.

The induced voltage falls to zero when the magnetic field becomes constant. There is only an induced voltage while the magnetic field is changing.

If the switch is opened, the magnetic field changes again, suddenly falling to zero. This also induces a voltage in coil B.

Cutting field lines and changing magnetic field

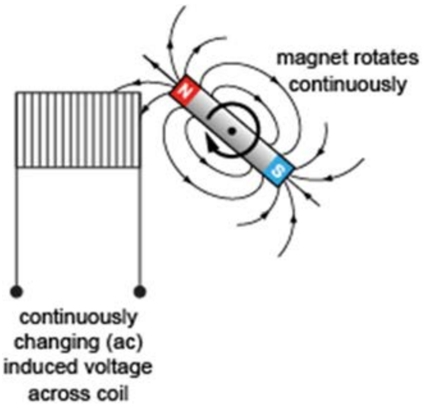
In the first example, a magnet is moved relative to a coil and the induced voltage was explained by the wires in the coil cutting through the magnetic field lines. However, this could also be explained by saying that the magnetic field through the coil has changed (it gets stronger when the magnet approaches and weaker when it moves away).

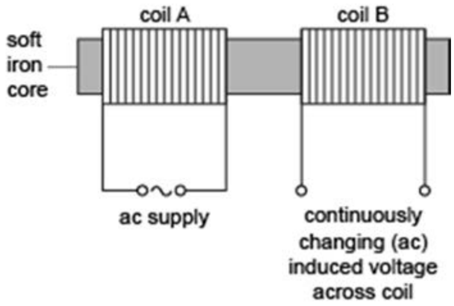
In the second example, the induced voltage was explained by a changing magnetic field through the coil. However, this could also be explained by saying that as the magnetic field changes its strength, field lines must move past the wires in the coil, which cut them (you can imagine a strengthening field pushing field lines outwards and a weakening field drawing them inwards).

Examples like these two show that cutting field lines and changing magnetic field strength can be two different ways to describe the same physical phenomenon.

Continuously changing magnetic fields

There is only an induced voltage when a change occurs (cutting across magnetic field lines or changing magnetic field strength). In both of the examples discussed above, the induced voltage only appears when i) the magnet is moving, or ii) the current in coil A is switched on or off. In order to generate a continuous voltage in the coil, the change must also be continuous. This could be achieved in the first example by rotating the magnet and in the second example by connecting coil A to an alternating current (ac) supply.





These principles are used to construct generators and transformers.

Voltage and current

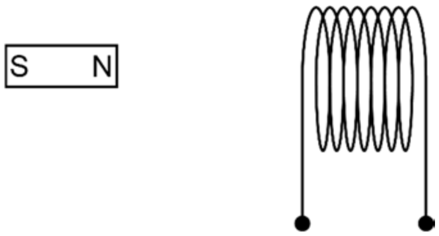
Electromagnetic induction always results in an induced voltage but will only produce a current if there is a closed circuit. In the two examples above the circuit is not closed so there is an induced voltage across the terminals of the coil but no current flows. If a resistor was connected across the terminals, then there would be a current (and energy would be transferred).

Know the factors affecting the magnitude of an induced voltage

The magnitude of an induced voltage is directly proportional to:

- the rate at which a wire cuts magnetic field lines, or
- the rate at which the magnetic field through a conductor (e.g. a coil) changes.

A bar magnet is held close to a coil as shown.



In which of the following situations would a voltage be induced across the terminals of the coil?

1) The magnet is moved toward the coil.

2) The magnet is moved away from the coil.

3) The coil is moved toward the magnet.

4) The coil is moved away from the magnet.

5) The coil and magnet are both moved to the right at the same speed.

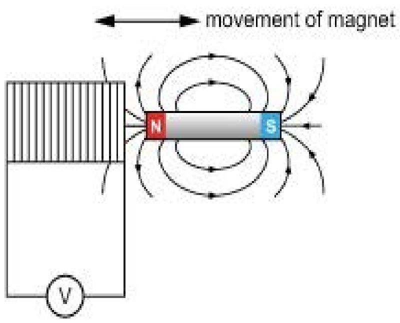
6) The coil and the magnet are both moved to the left at the same speed.

7) The magnet is placed at rest inside the coil.

A voltage will be induced in situations 1) – 4) but not in situations 5) – 7).

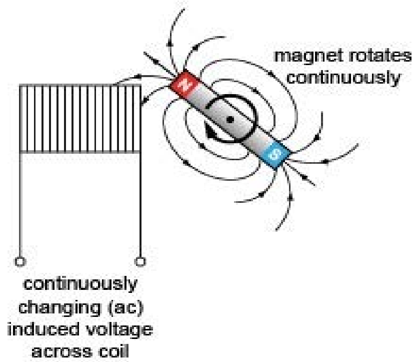
Electromagnetic induction occurs while the magnetic field inside a coil changes. For this to happen, there must be relative motion between the magnet and the coil, but it does not matter which one of them moves. This is the case in situations 1) – 4), but in situations 5) – 7) there is no relative motion, so the magnetic field inside the coil is not changing and there is no induced voltage.

Changing the magnitude of an induced voltage



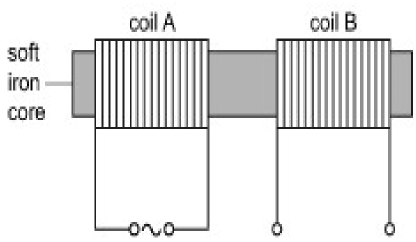
The induced voltage will increase if:

- the magnet is moved faster – more field lines are cut per second – the rate of cutting field lines increases
- a stronger magnet is used – there is a higher density of field lines, so more field lines are cut per second than with a weaker magnet at the same speed – the rate of cutting field lines increases*.



The induced voltage will increase if:

- the magnet spins at a greater rate – more field lines are cut per second – the rate of cutting field lines increases
- a stronger magnet is used – there is a higher density of field lines, so more field lines are cut per second than with aweaker magnet at the same speed – the rate of cutting field lines increases*.



The induced voltage will increase if:

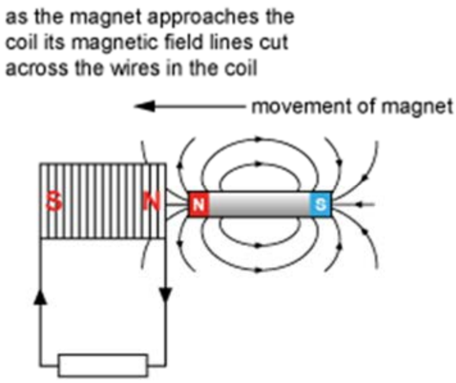
- the ac frequency in coil A is increased – the rate of change of the field through coil B increases
- the ac amplitude in coil A is increased – the rate of change of the field through coil B increases*.

*Increasing the number of turns on the coil will also increase the induced voltage in all of the examples above.

Know the factors affecting the direction of an induced voltage

An induced voltage is always in a direction that opposes the change that caused it.

To understand what this means, think about what happens when a bar magnet approaches a coil and the coil is connected to a resistor so that current can flow:



While the wires are cutting the field lines, there is an induced voltage in the coil and current in the circuit. The direction of current is such that a north pole is formed at the right hand end of the coil. This repels the approaching north pole, opposing the change that caused it.

- If the circuit was incomplete and there was no current, the induced voltage would still be in this direction.
- If the bar magnet is pulled away from the coil, the direction of the induced voltage reverses. This would cause a south pole at the right-hand end of the coil, attracting the bar magnet and again opposing the change that caused the induction.

The direction of an induced voltage reverses when:

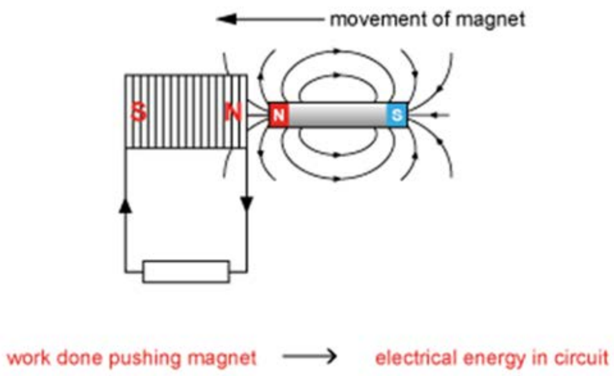
- the direction of the cutting of magnetic field lines reverses
- an increasing magnetic field in a coil changes to one that is decreasing
- a decreasing magnetic field in a coil changes to one that is increasing.

Conservation of energy

In order to conserve energy an induced voltage MUST oppose the change that caused it.

This can be seen by considering energy transfer when a magnet is pushed toward a coil.

The induced pole on the coil repels the bar magnet so work must be done by the person pushing the magnet. This work 'pays' for the electrical energy generated in the circuit when there is current in the resistor.



If, in the example above, the induced voltage was of the opposite sign, the end of the coil closest to the approaching bar magnet would have become a south pole. This would have attracted the bar magnet so that work would be done on the bar magnet and electrical energy would be generated in the circuit with no external work being done. This does not conserve energy and cannot happen.

When an aircraft flies due north, its aluminium wings cut through the vertical part of the Earth's magnetic field and a small voltage is induced across its wing tips. On the return journey, the aircraft is travelling due south at a higher speed.

How does the voltage induced across its wing tips on the return journey compare with that on the outward journey?

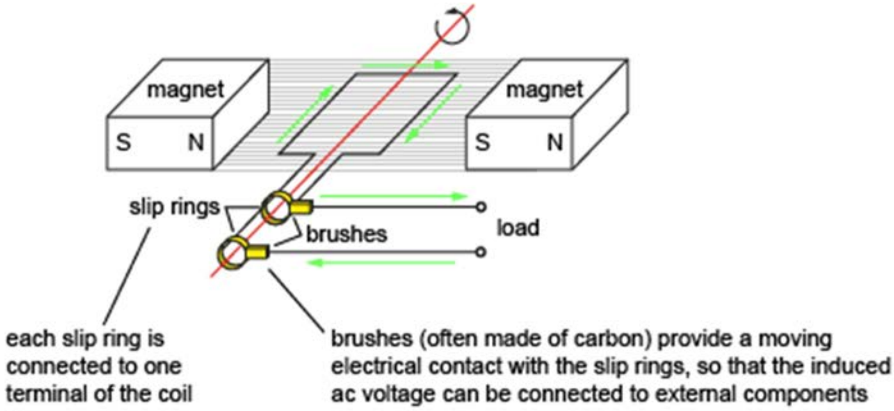
Explain your answer.

A higher voltage, but the polarity has reversed (i.e. east-west changes to west-east). The induced voltage is generated when a conductor (the aluminium wings) cuts through a magnetic field (the vertical part of the Earth's magnetic field). The magnitude of the induced voltage is proportional to the rate at which the magnetic field lines are cut; this has increased, so the voltage is greater. The polarity depends on the direction of cutting; this has reversed, so the polarity also reverses.

Understand the operation of an ac generator, including factors affecting the output voltage

A simple ac generator

A simple ac generator consists of a coil rotated in a magnetic field. As the coil rotates the magnetic field passing through it changes continuously, inducing a continuously changing voltage. The direction of the changing magnetic field in the coil changes every half rotation, so the sign of the induced voltage also changes every half rotation: the induced voltage is ac.



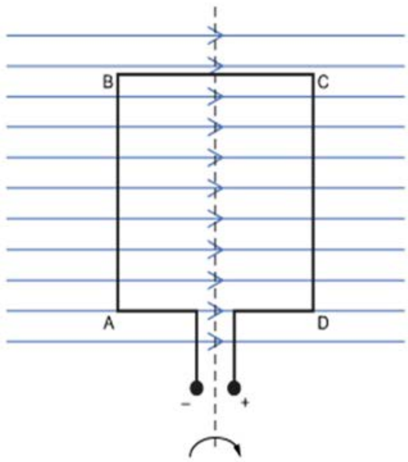
The amplitude of the output ac voltage increases if:

- the coil is rotated more rapidly – greater rate of change of magnetic field through the coil (or of cutting magnetic field lines)
- the magnetic field is stronger – greater rate of change of magnetic field through the coil (or of cutting magnetic field lines)
- the coil has greater area –greater rate of change of magnetic field through the coil (or of cutting magnetic field lines)
- there are more turns on the coil – each coil has the same induced voltage, and these voltages add together because the turns are in series.

The frequency of the output ac voltage is equal to the coil's rotation frequency.

Changing magnetic field strength or cutting magnetic field lines

The induced ac voltage can be explained either in terms of a changing magnetic field through the coil (as above) or by the rate at which the wires in the coil cut through the magnetic field lines. As the coil turns, one side moves up through the field while the other side moves down. This induces voltages in opposite directions on each side of the coil, producing a net voltage across the terminal of the coil:



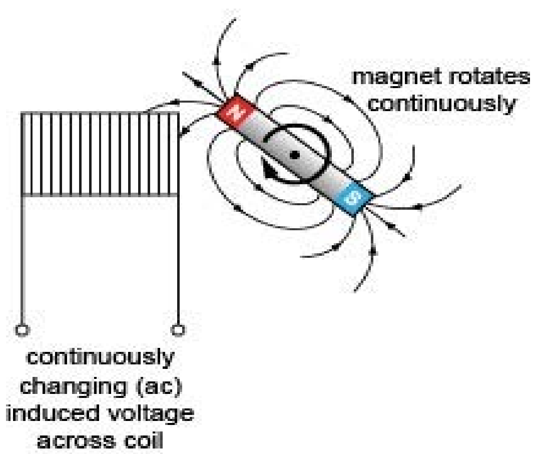
- Side AB moves out of the page: B becomes +ve and A becomes -ve *
- Side CD moves into the page: C becomes -ve and D becomes +ve *

These two induced voltages add together to make the output terminals positive and negative as shown.

*Knowing which end is positive and which is negative is not required at this level. The important thing to realise is that the induced voltages on each side will be in the opposite direction and so add together around the coil.

Moving magnets

It does not matter whether the coil moves through a stationary magnetic field or whether the magnetic field moves around a stationary coil. Both result in a changing magnetic field through the coil (or magnetic field lines being cut) so both induce a voltage. A simple design (used in some bicycle dynamos) uses a rotating magnet close to a coil:



Which of the following statements about a simple ac generator is/are correct?

1) If the generator is turned in the opposite direction, its output will still be ac.

2) Slip rings and brushes must be made from conducting materials.

3) If the time for one rotation of the coil is doubled, the peak voltage doubles and the frequency halves.

4) One of the slip rings is always positive and the other slip ring is always negative.

Statement 1): Whichever direction the coil rotates, the wires on one side of the coil will move down through the field for half of the rotation and up through the field for the other half of the rotation. So the direction of cutting changes and the polarity of the output changes, generating ac in both directions of rotation.

Statement 2): If the generator is to supply electric current to external components, current must be able to flow through the brushes and slip rings, so both must be made of a conducting material. For example, the brushes are often carbon and the slip rings brass.

Statement 3): When the time for one rotation doubles, the coil is rotating more slowly and it cuts the field at a lower rate inducing a smaller voltage, so this statement is incorrect. The frequency, however, (given by $f = 1/T$) does halve.

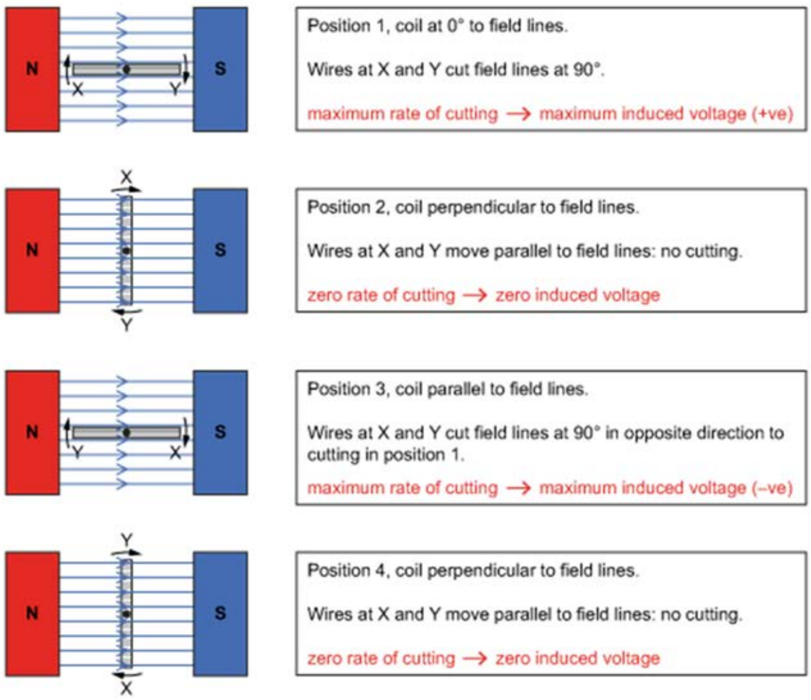
Statement 4): For half of the rotation, one slip ring is positive and the other negative but in the other half of the rotation, the wires are moving through the field in the opposite direction, so the slip rings both change polarity, becoming negative and positive respectively.

So, statements 1) and 2) are both correct but statements 3) and 4) are incorrect.

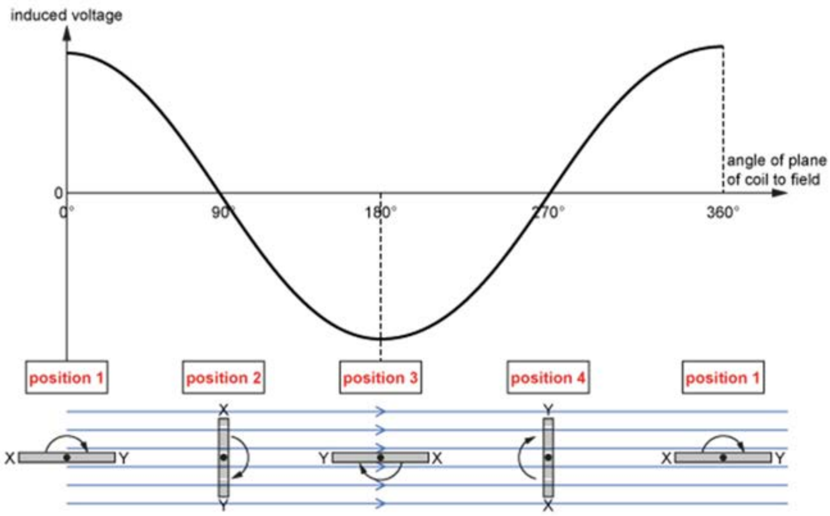
Interpret the graphical representation of the output voltage of a simple ac generator

The output of an ac generator is an induced voltage produced by electromagnetic induction.

The magnitude and sign of the induced voltage is directly proportional to the rate at which the wires in the coil cut through the magnetic field lines (or the rate at which the magnetic field through the coil changes). This causes the induced voltage to change in both magnitude and sign as the coil completes one revolution.



These positions are marked onto the graph of voltage output for one cycle shown below.

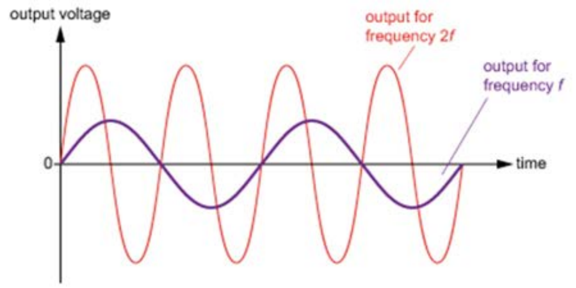


Amplitude and frequency

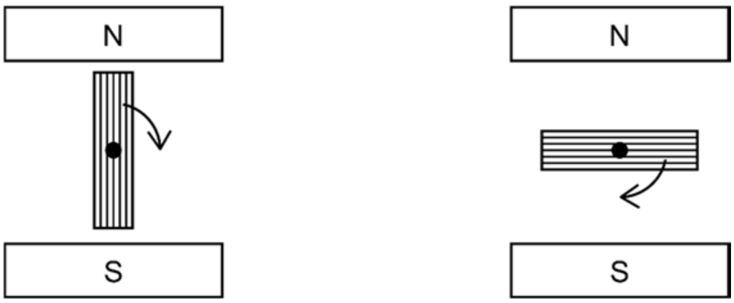
Increasing the frequency of rotation of the coil has two effects:

- it increases the frequency of the output ac voltage because the direction of cutting of the field lines changes more rapidly
- it increases the amplitude of the output ac voltage because the rate at which the field lines are cut increases.

The graph shows the effect of doubling the frequency of rotation from f to 2f:



The diagram shows two positions of the coil in a simple ac generator during a fraction of one revolution.



Which line in the table correctly describes the output voltage in these two positions? Explain your answer.

	Position 1	Position 2
a)	Zero voltage	Peak voltage
b)	Peak voltage	Zero voltage

The magnetic field goes from north to south, i.e. vertically downwards.

Peak voltage occurs when the wires at the ends of the coil cut across the field at the maximum rate, i.e. when they are moving horizontally. This occurs in position 1.

Zero voltage occurs when the wires at the ends of the coil move parallel to the magnetic field and do not cut across the field lines, i.e. when they move vertically. This occurs in position 2.

So b) is correct.

Understand applications of electromagnetic induction

The most important applications of electromagnetic induction are:

- generation of ac electrical energy – generators
- transmission of ac electrical energy – transformers.

Generators transfer mechanical work (to rotate the generator) to electrical energy in the form of ac electricity. The energy source for the mechanical work comes from:

- chemical energy – e.g. combustion of fossil fuels in coal, oil and gas fired power stations: water is heated to create high pressure steam that turns the generator
- kinetic energy – e.g. motion of the air through a wind turbine or water through a water turbine.

Other applications of generators include:

- car alternators – these recharge the car battery while the engine is running
- electromagnetic torches – a strong magnet can be shaken inside a coil to generate electricity and light an LED
- induction hobs on cookers – ac current in a coil underneath a glass plate creates an alternating magnetic field that passes through the base of a metal saucepan above it. The changing magnetic field induces an ac voltage in the base of the saucepan and the resulting current heats it up.

Transformers convert ac electricity at one voltage into ac electricity at a higher or lower voltage. This enables electrical energy to be transferred efficiently over long distances.

Other applications of transformers include:

- stepping down the ac mains supply to run low voltage devices such as laptop computers, and phone chargers etc. (this also usually requires the use of a device to convert ac to dc)
- stepping up currents for electrical welding equipment.

Generators are used in nuclear power stations to generate electricity for the national grid. Which of the following energy transfers is carried out by the generator?

1) nuclear energy → electrical energy

2) chemical energy → electrical energy

3) kinetic energy → electrical energy

4) thermal energy → electrical energy

Generators use electromagnetic induction to generate electricity when a magnet moves near a coil or vice versa. Other energy transfers result in fast moving steam (with large kinetic energy) being blasted at a turbine to make it spin. This drives the generator which transfers kinetic energy to electrical energy for output to transmission lines.

The answer is 3) kinetic energy → electrical energy

P2.5

Transformers:

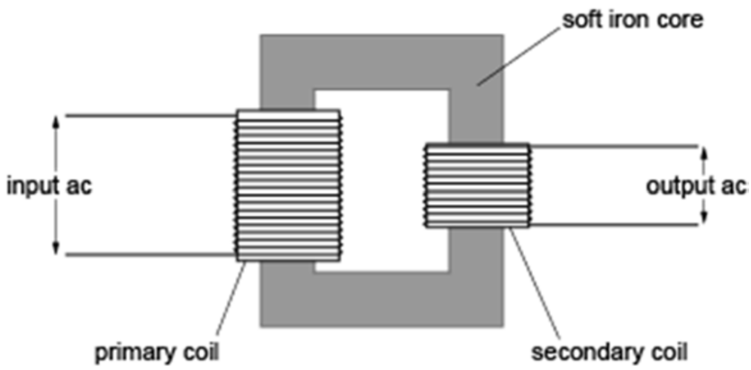
a. Know and understand the terms step-up transformer and step-down transformer.
b. Know and use the relationship between the number of turns on the primary and secondary coils, and the voltage ratio: $\frac{V_p}{V_s} = \frac{n_p}{n_s}$
c. Know that a consequence of 100% efficiency is total transfer of electrical power, and that this gives rise to the following relationship: $V_p I_p = V_s I_s$. Know and use this relationship to solve problems.
d. Understand power transmission, including calculating losses during transmission and the need for high voltage.

Know and understand the terms step-up transformer and step-down transformer

Transformers use electromagnetic induction to increase or decrease the voltage of an ac supply.

- A step-up transformer increases the voltage.
- A step-down transformer decreases the voltage.

Transformers consist of two coils wound onto a soft iron core. The coil on the input side of the transformer is called the primary coil and the coil on the output side of the transformer is called the secondary coil.



Transformers are used to step down mains ac voltages (230 V) down to the lower voltages used to power common electrical devices such as laptops and mobile phone chargers. They are also used to step up the ac output voltage from power stations to very high voltages for long distance transmission.

Why transformers need ac

When there is a current in the primary coil, it creates a magnetic field in the core and this passes through the secondary coil. Electromagnetic induction generates a voltage in the secondary coil when the magnetic field inside it changes. This only occurs if the current in the primary coil changes. A transformer uses ac because ac current in the primary coil is continuously changing, generating a continuously changing magnetic field and inducing a continually changing (ac) voltage in the secondary coil.

Current in primary coil changes

↓

Magnetic field in core changes

↓

Voltage is induced in the secondary coil

If the primary is connected to a constant dc supply there will still be a magnetic field in the core and this will still pass through the secondary. However, the magnetic field will be constant so there is no induced voltage in the secondary coil.

The soft iron core

Iron is a magnetically soft material. This means it can be magnetised and demagnetised easily and quickly. This allows the magnetic field in the core to change both in magnitude and direction as the ac current in the primary coil changes. The iron core loops through both coils linking them magnetically so that the changing magnetic field in the primary also passes through the secondary.

Which of the following statements about a step-up transformer is/are correct?

1) A step-up transformer has more turns on the secondary coil than on the primary coil.

2) A step-up transformer is used to increase current.

3) A step-up transformer is used to increase the voltage of a dc supply.

4) A step-up transformer is used to increase the energy of an ac supply.

Step-up transformers increase the voltage of an ac supply. The ratio of the secondary voltage to the primary voltage of a transformer is equal to the ratio of the number of turns on the secondary coil to the number of turns on the primary coil, so in order to step up the voltage, there must be more turns on the secondary than the primary, so statement 1) is correct.

Stepping up the voltage results in the current being reduced (or else energy would not be conserved) so statement 2) is incorrect.

All transformers work by electromagnetic induction, so they need a continually changing current in the primary, i.e. ac not dc, so statement 3) is incorrect.

Energy cannot be created so statement 4) is incorrect.

Know and use the relationship between the number of turns on the primary and secondary coils, and the voltage ratio: $\frac{V_p}{V_s} = \frac{n_p}{n_s}$

For an ideal transformer the voltage ratio is equal to the turns ratio.

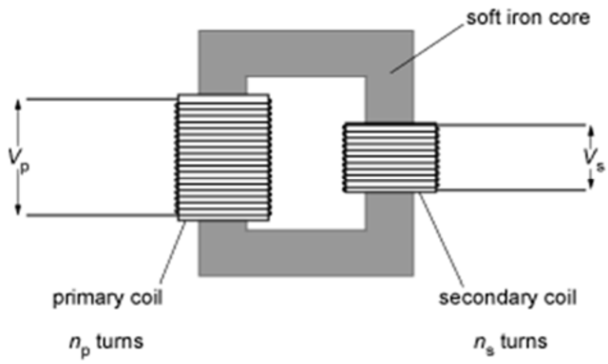
$$\frac{V_p}{V_s} = \frac{n_p}{n_s}$$

V_p = (ac) voltage across the primary coil

V_s = (ac) voltage across the secondary coil

n_p = number of turns on the primary coil

n_s = number of turns on the secondary coil



- For a step-up transformer $n_s > n_p$ so that $V_s > V_p$
- For a step-down transformer $n_s < n_p$ so that $V_s < V_p$

Ideal transformers

The equation:

$$\frac{V_p}{V_s} = \frac{n_p}{n_s}$$

is only valid for an ideal transformer, i.e. one that is 100% efficient. However, in practice many real transformers are almost 100% efficient and so this equation is still useful.

Conservation of energy

Whilst a transformer can be used to increase the voltage from a supply it cannot create energy. A consequence of increasing voltage is that the current in the secondary decreases and since energy transferred is equal to E = IVt, the decrease in current balances the increase in voltage so that energy is conserved.

A transformer is used to step down 240 V ac to 20 V ac.

Which of the following could be the numbers of turns on the primary and secondary coils?

a)	4800	400
b)	400	4800
c)	48000	400
d)	400	48000

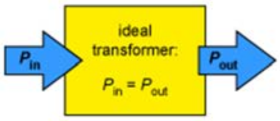
It is a step-down transformer, so there are fewer turns on the secondary coil than the primary coil. The voltage is stepped down to 20/240 = 1/12 of its original value so the ratio of turns must be 1:12. This is the case in a), since 400 = 4800/12. Therefore, the only possible combination is 4800 turns on the primary coil and 400 turns on the secondary coil.

Know that a consequence of 100% efficiency is total transfer of electrical power, and that this gives rise to the following relationship: $V_pI_p = V_sI_s$. Know and use this relationship to solve problems

Electrical power transfer through a transformer

An ideal transformer is 100% efficient.

This means that the electrical power P_{in} transferred to the primary coil is equal to the electrical power P_{out} transferred from the secondary coil to external components:



Electrical power is equal to voltage × current so:

$P_{in} = V_pI_p$

$P_{out} = V_sI_s$ where

I_p is the current in the primary coil and I_s is the current in the secondary coil

V_p is the voltage across the primary coil and V_s is the voltage across the secondary coil.

For an ideal transformer $P_{in} = P_{out}$ so:

$$V_pI_p = V_sI_s$$

Current ratio

The current ratio for an ideal transformer can be obtained by rearranging the equation $V_pI_p = V_sI_s$ to give:

$$\frac{V_p}{V_s} = \frac{I_s}{I_p}$$

In other words, the current ratio is the inverse of the voltage ratio. If an ideal transformer steps up* the voltage, then it must also step down the current by the same factor.

Since voltage ratio is equal to turns ratio, the current ratio is equal to the inverse ratio of the turns:

$$\frac{V_p}{V_s} = \frac{I_s}{I_p} = \frac{n_p}{n_s}$$

*Note that, when we refer to a step-up or step-down transformer, we are always referring to what it does to the voltage, not the current.

Real transformers

Whilst large transformers do come close to 100% efficiency, all real transformers fall short of this figure. This is because some of the input power is transferred to heat as a result of:

- the resistance in the wires on the coils
- heating effects in the core as it magnetises and demagnetises
- currents induced in the core (eddy currents) by the changing magnetic field.

A transformer is used to step up the voltage from a power station from 25 000 V to 250 000 V. The input current to the transformer is 4.0 A.

What is the output current? (Assume the transformer is 100% efficient.)

The transformer is 100% efficient, so the power input must equal the power output. Electrical power is given by the equation P = IV so $V_pI_p = V_sI_s$:

$250\,000 \times 4.0 = 25\,000 \times I_s$

$I_s = 25\,000 / (4.0 \times 250\,000) = 0.40$ A.

The voltage has stepped up by a factor of 10, so the current is stepped down by the same factor.

Understand power transmission, including calculating losses during transmission and the need for high voltage

Electrical power is transmitted over long distances using transmission lines.

High voltages are used in order to keep the current in the transmission lines low and therefore reduce losses due to heating of the cables.

The power transmitted is given by:

$P = IV$

where I is the current in the transmission line and V is the voltage between the cables. Note that the same power can be transmitted by high current and low voltage as by low current and high voltage.

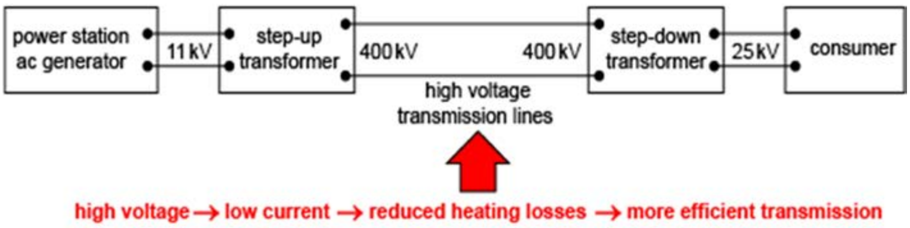
Transmission lines are made from conductors such as copper and aluminium that have electrical resistance. When there is a current in the transmission line, this resistance (R) causes some of the electrical power to be transferred to heat.

Power wasted in a transmission line = I^2R

To minimise wasted energy current must be made as small as possible.

This is done by using a transformer to step up the voltage to a very high level for transmission.

Here is a simple diagram showing how electrical energy is transferred from the generator in a power station to a consumer (the values given are as an example – in practice these will vary in different systems):



Voltage between cables and voltage along cables

When calculating the power output of a generator or the power delivered to a load resistor (e.g. to a consumer) we use the equation:

$P = IV$

The V in this equation refers to the potential difference between the terminals of the generator or between the cables connected to the load. This is the voltage that is stepped up at the start of the transmission line and stepped down at the consumer end.

However, the transmission lines themselves have some resistance. This is what causes heating in the cables and energy loss in transmission. When current flows in the transmission lines there is a voltage drop along the line equal to $V_{drop} = IR$. This should not be confused with the voltage between the cables referred to above and this is not the voltage that is stepped up or down by the transformers. ac power is transmitted at high V in order to reduce the current I in the transmission lines and therefore reduce the voltage drop V_{drop} along them.

Need for ac

Transformers are needed to step up and step down the voltage so that transmission losses are minimised.

Transformers work by electromagnetic induction which requires a changing magnetic field. ac provides a continually changing current and so generates a changing magnetic field inside the transformers. Transformers do not work with dc.

Typical voltages

The typical voltage output of a power station generator is around 11 kV or 33 kV.

Transmission voltages in the UK are usually 275 kV or 400 kV.

Consumer voltages vary depending on the type of consumer e.g. railways 25 kV / domestic 230 V.

Higher voltages are hard to insulate so consumers cannot use the same voltage as the high voltage transmission cables, hence the need for a step-down transformer at the consumer end.

Which of the following statements is/are correct about high voltage ac transmission?

1) Transmitting electricity at higher voltage reduces the current in the transmission lines.

2) Transmitting electricity at higher voltage increases the efficiency of the process.

3) Transmitting electricity at higher voltage reduces the energy losses due to heat transfer.

4) Transmitting electricity at higher voltage makes it easier to insulate the transmission lines.

5) Transmitting electricity at higher voltage allows more of the power generated to reach the consumer.

Energy cannot be created, so increasing the voltage must reduce the current. Statement 1) is correct.

If the current is lower, the cables will not heat up to the same extent, so less energy is lost and the process is more efficient. Statements 2) and 3) are correct.

This allows more of the generated power to reach the consumer. Statement 5) is correct.

The higher the voltage, the harder it is to insulate from other conductors. Statement 4) is incorrect.

So, all of the statements except statement 4) are correct.

P3. Mechanics

P3.1

Kinematics:

a. Know and understand the difference between scalar and vector quantities.
b. Know and understand the difference between distance and displacement and between speed and velocity.
c. Know and be able to apply: $\text{speed} = \frac{\text{distance}}{\text{time}}$, $\text{velocity} = \frac{\text{change in dispacement}}{\text{time}}$
d. Know and be able to apply: $\text{acceleration} = \frac{\text{change in velocity}}{\text{time}}$
e. Interpret distance–time, displacement–time, speed–time and velocity–time graphs.
f. Perform calculations using gradients and areas under graphs.
g. Know and be able to apply: $\text{average speed} = \frac{\text{total distance}}{\text{total time}}$
h. Know and be able to apply the equation of motion: $v^2 - u^2 = 2as$

Know and understand the difference between scalar and vector quantities

- A scalar quantity is one that has a *magnitude* (size), but not a *direction*.

Examples: mass, time, energy, temperature, power, density, pressure, speed, distance.

- A vector quantity is one that has both *magnitude* and *direction*.

Examples: force, velocity, acceleration, displacement, momentum.

Scalar quantities

Just because a scalar quantity has a magnitude only, that does not mean it can only have a positive value.

Many scalar quantities can have negative values, and the quantity can both increase and decrease. For example, a potential energy below the arbitrary zero reference point will be negative. Or a temperature below 0°C will be negative.

Other scalar quantities can only have positive values. For example, it would not make sense to talk about a mass being negative, but it is perfectly possible to talk about negative *changes* in mass.

Vector quantities

Vector quantities can also have positive or negative values, but always with reference to a direction being defined (arbitrarily) as the positive direction. For example, a force acting vertically upwards will be positive if upwards is defined as positive, and negative if downwards is defined as positive.

When vector quantities change, the change is itself a vector quantity, in other words, it has direction. For example, if an object has an initial velocity of $5ms^{-1}$ to the left and this increases to $8ms^{-1}$ to the left then the change in velocity is $3ms^{-1}$ to the left (or $-3ms^{-1}$ to the right).

Sometimes a question only deals with the magnitude of a vector quantity, and in such situations it is not necessary to consider directions. As an example, when the motion of an object in one direction in a straight line is being considered, the magnitude of the acceleration can be used in conjunction with the speed. In this situation the magnitude of the acceleration is effectively being used as a scalar quantity.

Consider the following narrative:

‘When I got in my car this morning to go to work, it was -5°C outside. It took me 10 minutes to defrost the windscreen and by then the traffic was really heavy. I was 25 minutes later than normal getting to work, which is only 750m away from my home in a direct line. I never got above the speed of 20 mph, and it would have been quicker to walk – the footpath is 3 miles shorter than the route by road, and I can walk at an average speed of 2mph.’ Which quantity in this narrative relates to the magnitude of a vector quantity?

The temperature and times are clearly scalar quantities.

The 20mph figure relates to the speed of the car, with no reference to direction, and is therefore a scalar quantity. The 2mph figure also relates to a speed.

The 3 mile figure relates to differences in actual distances travelled, irrespective of direction, and is therefore a scalar quantity.

The ‘as the crow flies’ figure of 750m is a displacement, the straight line displacement of my work from my house.

Displacement is a vector quantity, and hence 750m is the answer.

A person walks 5.0m due north, then 6.0m due east, then 7.0m due south, then 8.0m due west.

What is the final displacement of this person in the northerly direction from his starting point?

We can ignore the information about the distances travelled due east and due west because they have no effect on the displacement in the north-south direction.

In the north-south direction, the person travels 5.0m north and 7.0m south. This means that he ends up 2.0m south of the starting point.

But the question defines north as being the positive direction, so a displacement in the southerly direction is negative against this reference point. The answer is therefore -2.0m.

Know and understand the difference between distance and displacement and between speed and velocity

Distance is a scalar quantity, and displacement is a vector quantity.

Distance is therefore the magnitude of displacement.

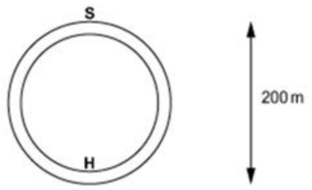
Displacement is distance in a particular direction.

Speed is a scalar quantity, and velocity is a vector quantity.

Speed is therefore the magnitude of velocity. Velocity is speed in a particular direction.

Distance and displacement

Consider a car moving, from starting point S, around a circular track of radius 100m (circumference 630m):



On the first time around the track, when the car has reached the half-way point H, the car has travelled a distance of 315m (half the circumference). But when it reaches H, its displacement from S is +200m.

By the time the car gets back to S, the distance travelled is 630m but its displacement from S is zero. Its displacement from H is -200m, with positive defined as the S→H direction.

When the car reaches point H on its second lap, the accumulated distance travelled becomes (315 + 630 =) 945m, but the displacement from S is once again +200m.

Speed and velocity

In the above example, suppose the car is moving at a constant speed of $20ms^{-1}$ in a clockwise direction.

The velocity at S is $+20ms^{-1}$ to the right ($-20ms^{-1}$ to the left). The velocity in the S→H direction is zero.

The velocity at H is $-20ms^{-1}$ to the right ($+20ms^{-1}$ to the left). The velocity in the S→H direction is zero.

A hotel has 10 floors, with each floor separated vertically by 5.0m from the adjacent floors. A lift starts at floor 2, moves up to floor 8, then to floor 6, then to floor 9, and finally to floor 1.

What is the distance travelled by the lift, and what is its final displacement from its starting point (taking upwards to be the positive direction)?

The lift moves up 6 floors (+30m), down 2 floors (–10m), up 3 floors (+15m) and down 8 floors (–40m). Distance travelled ignores direction and so distance = 30 + 10 + 15 + 40 = 95m. Displacement looks only at the relative position of the lift at the start and at the end. It starts on floor 2 and ends on floor 1. So the net displacement is (30 – 10 + 15 – 40 =) –5.0m.

A ball falls onto the ground vertically and bounces. The speed of impact is $15ms^{-1}$ and the speed of rebound is $10ms^{-1}$.

What is the change in speed of the ball as it bounces, and what is its change in velocity? (Take the upwards direction as positive.)

For the change in speed, we are not interested in direction. The only consideration is that the speed decreases from $15ms^{-1}$ to $10ms^{-1}$. A change of (15 – 10 =) $5ms^{-1}$.

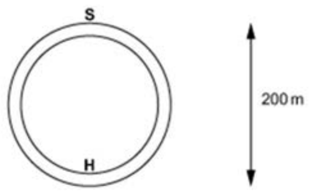
For the change in velocity, we have to take direction into account. The ball has a speed just before the impact of $15ms^{1}$ downwards, which is a velocity of $-15ms^{-1}$ in the upward direction. The velocity just after impact is $+10ms^{-1}$. So, the change in velocity is (10 – (–15)) = $+25ms^{-1}$.

Know and be able to apply: $speed = \frac{distance}{time}$ $velocity = \frac{change\ in\ dispacement}{time}$

Both speed and velocity have units of distance / time, so SI unit is ms^{-1}.

These formulae can only be used in situations where speed (or velocity) is constant.

Let us return to the car moving around a circular track of radius 100m from the previous section. The circumference of the track was 630m, and we have reference points marked for the start (S) and half-way-round (H) positions:



Suppose this car travels at a constant speed and takes 20 seconds to complete each lap.

From S, around the track, and back to S:

Speed = distance / time = 630 / 20 = 31.5 ms^{-1}.

Overall velocity = change in displacement / time = 0 / 20 = 0 ms^{-1}.

From S, half way around the track, to H:

Speed = distance / time = 315 / 10 = 31.5 ms^{-1}.

Overall velocity = change in displacement / time = 200 / 10 = 20 ms^{-1} in the S→H direction.

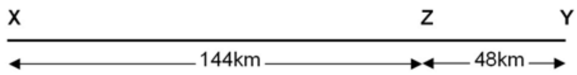
A train travels due east along a straight track at a speed of 80 ms^{-1}. It takes 40 minutes to get from station X to station Y. The train then travels (west), back along the track at constant speed to reach station Z (which is 144km from station

X). This journey takes 16 minutes.

What is the velocity of the train for the journey between Y and Z? (Take eastwards to be positive.)

Distance from X to Y = speed × time = 80 × 40 × 60 = 192 000m = 192km.

Distance X to Z = 144km and so distance from Y to Z is 192 – 144 = 48km.



Station Z is to the west of station Y so the displacement of the train from Y to Z = –48km.

Velocity = change in displacement / time = –48 000 / (16 ×60) = –50 ms^{-1}.

A bird flies at a constant speed of 20 ms^{-1}. It flies due south for 20 minutes, then due west for 10 minutes, then due north for 10 minutes, then due east for 20 minutes.

What is the displacement of the bird to the north of its starting point by the end of this journey?

Since we are interested only in the north-south displacement of the bird, we can ignore the east and west elements.

In the north-south direction, the bird flies south a distance = 20 × 20 × 60 = 24 000 m.

It then flies north a distance = 20 × 10 × 60 = 12 000 m.

It therefore ends up a distance of 24 000 – 12 000 = 12 000 m = 12 km south of its starting point. The displacement *north* of its starting point is therefore = –12 km.

Know and be able to apply: $\boldsymbol{acceleration = \frac{change\ in\ velocity}{time}}$

Unit of acceleration = unit of velocity / unit of time = (ms^{-1}) / s = ms^{-2}.

This means that velocity is changing at a rate of 1 ms^{-1} each second.

Acceleration is a vector quantity. Its direction is the same as the direction of the change in velocity.

This formula can only be used in situations where the acceleration is constant.

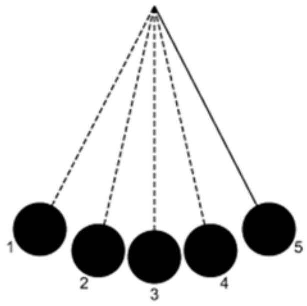
An object can have an acceleration even if its velocity is momentarily zero. Similarly, an object travelling at a constant velocity has zero acceleration.

For an object moving in a uniform direction, the magnitude of acceleration is equal to the rate of change of speed.

An object moving at constant speed can still have a non-zero acceleration if its direction of movement is changing (for example, an object moving in a circle).

An object that has an acceleration in the same direction as its direction of motion will speed up. An object that has an acceleration in the opposite direction to its direction of motion will slow down.

Where motion and accelerations in opposing directions are being considered, it is important to be aware of signs and the direction that is being considered the 'positive direction'. For example, consider the bob on the end of a pendulum swinging between points 1 and 5 (the points of maximum displacement) as shown:



If we consider left to right → as the positive direction, and consider one complete oscillation from position 1 to 5 and then back to 1 again, we can describe the motion of the bob as follows:

position	velocity	acceleration	description of motion
1	zero	positive	at rest but accelerating to the right
2	positive	positive	moving to the right and increasing speed
3	positive	zero	moving to the right at maximum speed
4	positive	negative	moving to the right and decreasing speed
5	zero	negative	at rest but accelerating to the left
4	negative	negative	moving to the left and increasing speed
3	negative	zero	moving to the left at maximum speed
2	negative	positive	moving to the left and decreasing speed
1	zero	positive	at rest but accelerating to the right

A car accelerates at a constant rate in a straight line from 20 ms^{-1} to 25 ms^{-1} in 4.0 seconds. The car moves in the same direction throughout. What is the magnitude of its acceleration?

Change in speed = 25 – 20 = 5.0 ms^{-1}.

The car does not change direction and so this is also the change in velocity. Therefore, acceleration = change in speed / time = 5.0 / 4.0 = 1.25 ms^{-2}.

A ball hits a bat at an initial speed of 25 ms^{-1} and rebounds in the opposite direction at 20 ms^{-1}. The time of impact is 0.30s. Taking the initial direction of motion of the ball as positive, what is the acceleration of the ball whilst in contact with the bat?

Initial velocity of ball = + 25 ms^{-1}.

Final velocity of ball = –20 ms^{-1}.

∴ Change in velocity of ball = –20 – 25 = –45 ms^{-1}.

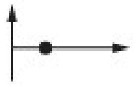
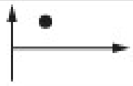
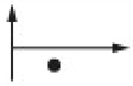
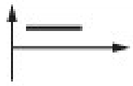
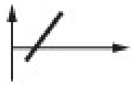
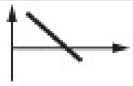
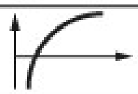
∴ Acceleration = change in velocity / time = –45 / 0.30 = –150 ms^{-2}.

Interpret distance–time, displacement–time, speed–time and velocity–time graphs

- A distance-time graph can only be positive and can only ever increase. (Distance is a scalar quantity, so only has a magnitude.) The distance travelled by an object can never decrease – even if it returns to a starting point, more distance is travelled in the process. (Be aware that some textbooks incorrectly show distance-time graphs decreasing, but these graphs should be labelled as displacement-time graphs.)
- A displacement-time graph shows how the distance of an object away from a fixed point varies with time. It can be positive or negative depending on where the object is in relation to the fixed point and which direction is defined as positive.
- A speed-time graph can only be positive and shows how the speed of an object varies with time irrespective of direction travelled.
- A velocity-time graph shows both the speed and direction of movement and can be positive or negative according to how the direction of movement relates to the direction defined as positive.

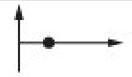
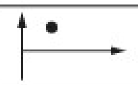
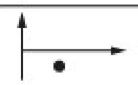
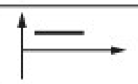
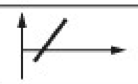
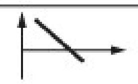
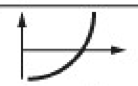
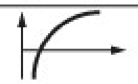
Displacement–time graphs

(shown with displacement on the y-axis and time on the x-axis)

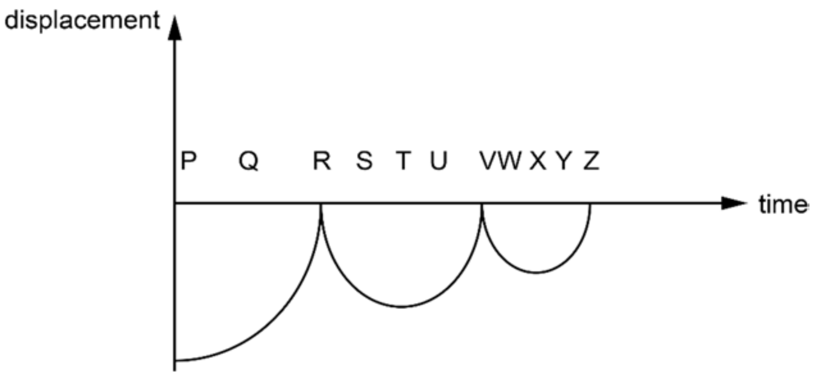
feature	appearance	meaning
zero value		located at starting point
positive value		located the positive side of the starting point
negative value		located on the negative side of the starting point
zero gradient		stationary
positive gradient		moving in the positive direction
negative gradient		moving in the negative direction
increasing gradient		increasing velocity
decreasing gradient		decreasing velocity

Velocity–time graphs

(shown with velocity on the y-axis and time on the x-axis)

feature	appearance	meaning
zero value		stationary
positive value		moving in the positive direction
negative value		moving in the negative direction
zero gradient		constant velocity
positive gradient		accelerating in the positive direction
negative gradient		accelerating in the negative direction
increasing gradient		increasing acceleration
decreasing gradient		decreasing acceleration

Consider the following displacement-time graph for a ball falling from a height and bouncing on the floor:



At which time(s) is the ball moving downwards with an increasing speed?

The graph as presented is unusual because it is clearly taking the upwards direction as negative (and the downward direction as positive).

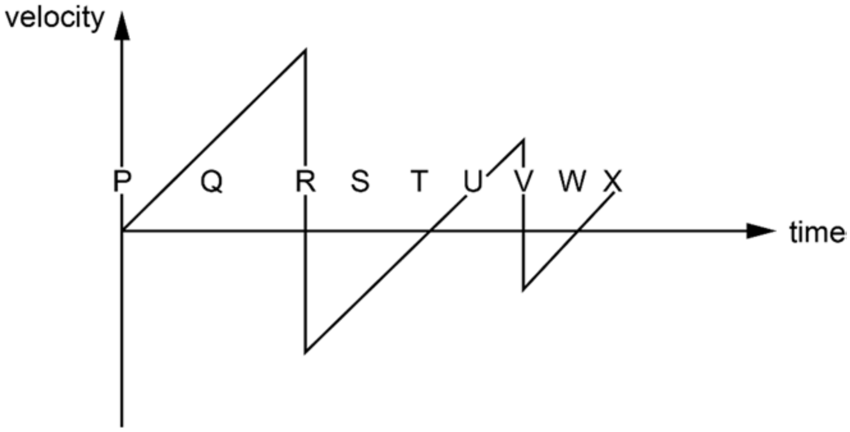
An increasing downward speed is therefore indicated on the graph by a positive gradient that is getting steeper. This happens at times Q, U and Y.

At times S and W, the ball is travelling upwards with a decreasing speed.

At times P, T and X, the ball is momentarily at rest at the top of each bounce.

At times R, V and Z, the ball is hitting the ground and the velocity goes quickly from a large downward value to a large negative value.

Consider the following velocity-time graph for a ball released from a height, then falling and bouncing on the floor:



At which time(s) is the ball at the top of a bounce?

As with the previous example, we can see here that the downward direction is being taken as positive and the upward direction as negative.

From P to R, the ball is falling from its release point P and accelerating downwards until it hits the floor.

At R, the ball bounces and its velocity quickly goes from a large downward (positive) value to a large upward (negative) value.

From R to T, the ball is moving upwards (a negative velocity) with decreasing speed until it reaches the top of the bounce at T.

At T, the velocity is momentarily zero as the movement of the ball changes direction from upwards (negative) to downwards (positive).

From T to V, the ball moves downwards (positive direction) with increasing speed until it bounces again at time V. The process is then repeated, with the top of the second bounce being reached at time W.

The answer to the question is therefore that the ball reaches the top of a bounce at times T and W.

It is worth stressing that the choice of direction defined as positive (downwards in these examples) is entirely arbitrary. The graphs could equally well have been presented with upwards defined as positive, in which case they would be vertically inverted. The important thing to bear in mind is the need for consistency of sign within a question.

Perform calculations using gradients and areas under graphs

Gradients

The gradient of a graph is found by calculating the change in the quantity on the y-axis divided by the change in the quantity on the x-axis. Therefore, where the y-axis quantity divided by the x-axis quantity has meaning, the gradient of the graph also has meaning.

The following relationships therefore follow from this:

- The gradient of a distance-time graph is the speed at that point.
- The gradient of a displacement-time graph is the velocity at that point.
- The gradient of a velocity-time graph is the acceleration at that point.

Areas

The area under a graph is found by calculating the average value of the quantity on the y-axis multiplied by the change in the quantity on the x-axis. Therefore, where the y-axis quantity multiplied by the x-axis quantity has meaning, the area under the graph also has meaning. The following relationships therefore follow from this:

- The area under a speed-time graph between two points is the distance moved between those points.
- The area under a velocity-time graph between two points is the change in displacement between those points.
- Areas below the axis in velocity-time graphs represent displacements in the negative direction.

Gradient of a displacement-time graph

- The steeper the gradient, the greater the speed.
- A positive gradient represents a velocity in the positive direction.
- A negative gradient represents a velocity in the negative direction.
- A gradient of zero represents zero velocity.
- The unit of velocity calculated from the gradient is the unit used for displacement on the y-axis divided by the unit used for time on the x-axis.

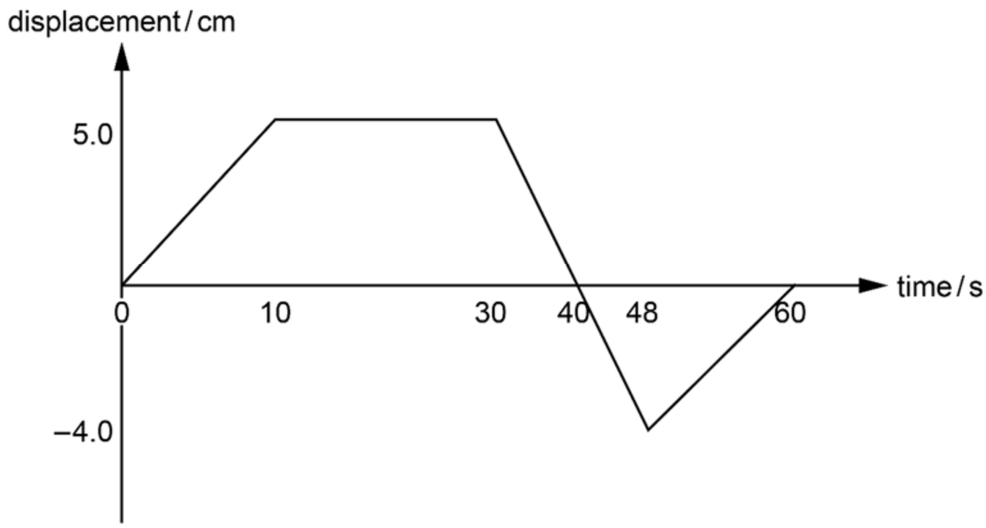
Gradient of a velocity-time graph

- The steeper the gradient, the greater the magnitude of the acceleration.
- A positive gradient represents an acceleration in the positive direction. For velocities in the positive direction, this corresponds to speeding up, whilst for velocities in the negative direction this corresponds to slowing down.
- A negative gradient represents a velocity in the negative direction. For velocities in the positive direction, this corresponds to slowing down, whilst for velocities in the negative direction this corresponds to speeding up.
- A gradient of zero represents acceleration of zero magnitude (constant velocity in the direction being considered).
- The unit of acceleration calculated from the gradient is the unit used for velocity on the y-axis divided by the unit used for time on the x-axis.

Area under a velocity-time graph

- A 'positive' area (i.e. an area bound by the time axis and velocities above it) gives a displacement in the positive direction.
- A 'negative' area (i.e. an area bound by the time axis and velocities below it) gives a displacement in the negative direction.
- When finding total displacements, individual displacements are added together (remembering to include a '-' sign for any negative areas).
- The unit of displacement calculated from the area under the graph is the unit used for velocity on the y-axis multiplied by the unit used for time on the x-axis.

Consider the following displacement-time graph for an object in motion:



What is the velocity of the object at time = 35s?

First, notice that at time = 35s, the line is a straight line with a negative gradient. This means that between 30s and 48s the object is travelling at constant speed in the negative direction.

There are three ways in which we can find the gradient of this line (30s to 40s, 30s to 48s, or 40s to 48s) and whichever method we choose will yield the same answer.

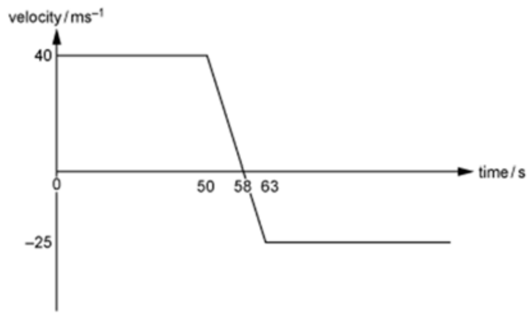
Let us consider the time interval 30s to 40s.

Change in displacement = (0.0 – 5.0) = –5.0 cm.

Time = (40 – 30) = 10s.

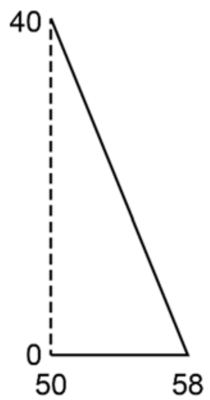
∴ Velocity = change in displacement / time = (–5.0) / 10 = –0.50 cm s^{-1}.

Consider the following velocity-time graph for a ball being hit by a bat:



What is the net displacement of the ball whilst it is in contact with the bat?

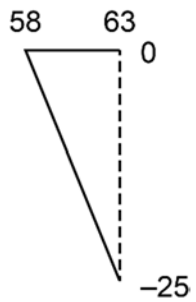
Consider the triangle above the time axis:



Area of triangle = $\frac{1}{2}$ × (58 – 50) × 10^{-3} × 40 = +0.16 m.

This is the displacement of the ball from its point of impact whilst it is still travelling in the positive (forward) direction until it becomes momentarily at rest.

Consider now the triangle below the time axis:



Area of triangle = $\frac{1}{2}$ × (63 – 58) × 10^{-3} × (– 25) = – 0.0625 m.

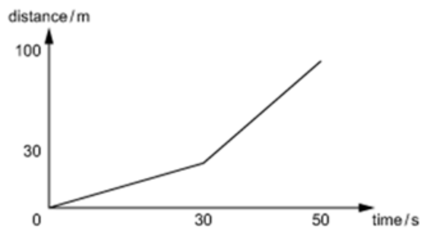
This is the displacement of the ball from the point where it is momentarily at rest to the point where it leaves contact with the bat whilst it is accelerating back the way it came (negative direction).

∴ Net displacement = 0.16 – 0.0625 = +0.0975m = +9.75 cm.

Know and be able to apply: $average\ speed = \frac{total\ distance}{total\ time}$

- Average speed = total distance / total time.
- Average velocity = net change in displacement / total time.
- Average acceleration = net change in velocity / total time.

Consider the motion of an object that is moving with different speeds as follows:



We can see that for the first 30 seconds, the distance travelled is 30m and so the speed (given by the gradient of the line) is 30 / 30 = 1.0 ms^{-1}.

During the next 20 seconds (from 30s to 50s), the distance travelled is 100 – 30 = 70m. Therefore, the speed (again given by the gradient) is 70 / 20 = 3.5 ms^{-1}.

How can the average speed be meaningfully calculated? Certainly not by just working out the average of the two separate speeds (which would be 2.25 ms^{-1}) because the object was at one speed for a longer period of time than the other.

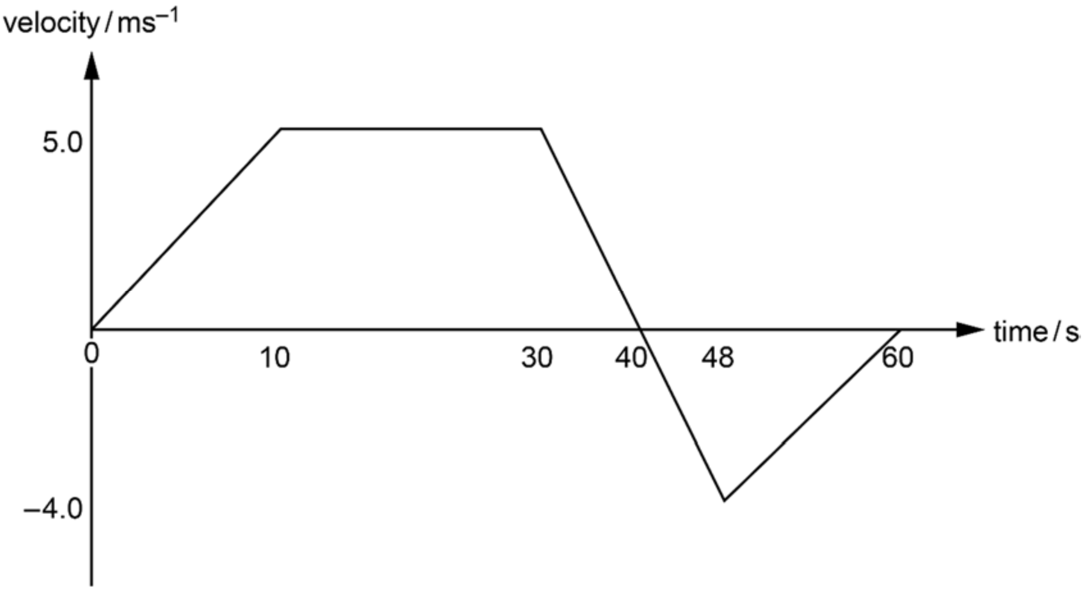
If the average speed is to be considered as the effective constant speed that, over the same period of time, would result in the object moving through the same distance, then it has to be calculated as total speed / total time.

Therefore, in this example, average speed = 100 / 50 = 2.0 ms^{-1}.

By this reasoning, we can see that average speed is related directly to the defining equation for speed. speed = distance / time, and so average speed = total distance / total time.

By the same reasoning, average velocity is related to the defining equation for velocity, and average acceleration is related to the defining equation for acceleration.

Consider the following velocity-time graph for an object in motion:



What is the average velocity of the object over the period of time shown in the graph?

First, we need to find the total displacement from the area under the graph:

Change in displacement from 0s to 10s = $\frac{1}{2}$ × 10 × 5.0 = +25m

Change in displacement from 10s to 30s = (30 – 10) × 5.0 = +100m

Change in displacement from 30s to 40s = $\frac{1}{2}$ × (40 – 30) × 5.0 = +25m

Change in displacement from 40s to 48s = $\frac{1}{2}$ × (48 – 40) × (–4.0) = –16m

Change in displacement from 48s to 60s = $\frac{1}{2}$ × (60 – 48) × (–4.0) = –24m

∴ Total change in displacement = +25 + 100 + 25 – 16 – 24 = +110m

Now we can find the average velocity:

average velocity = total change in displacement / total time = +110 / 60 = +1.83 ms^{-1}

In the above example, what is the average speed over the period of time shown in the graph?

We can use the magnitudes of the displacements from the previous example to find the distances travelled in each stage of the motion. From this:

Total distance = 25 + 100 + 25 + 16 + 24 = 190m

∴ Average speed = total distance / total time = 190 / 60 = 3.17 ms^{-1}

In the above example, what is the average acceleration over the period of time shown in the graph?

The initial velocity of the object (at time = 0s) is zero. The final velocity of the object (at time = 60s) is zero.

Therefore, whilst the velocity increases and decreases in between, the net change in velocity between 0 and 60s is zero.

Hence the average acceleration is zero.

Know and be able to apply the equation of motion: $v^2 - u^2 = 2as$

Equations of uniform motion

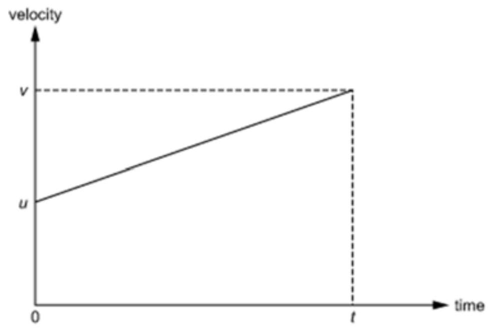
The following equations can usefully be used, as an alternative to graphical methods, to solve problems in which the acceleration is uniform (constant, both in magnitude and direction):

- $v = u + at$
- $s = \frac{1}{2}(u + v)t$
- $v^2 - u^2 = 2as$

(where u = initial velocity; v = final velocity; a = acceleration; t = time; s = displacement.)

Consider an object accelerating uniformly, from velocity u to velocity v in time t.

The velocity-time graph for this object is as follows:



- Acceleration a is given by the gradient of the line to be $(v - u)/t$.
- This is also an algebraic statement of the defining equation of acceleration, acceleration = change in velocity / time taken.
- Re-arranging, we get $at = v - u$ and so $v = u + at$.
- Displacement s is given by the area under the graph. There are various ways of finding this, but one way is to work out the area of the large rectangle and subtract the area of the triangle shown.

Accordingly:

Area of large rectangle = vt

Area of upper triangle = $\frac{1}{2} \times t \times (v - u)$

$\therefore$ Displacement = $vt - \frac{1}{2}vt + \frac{1}{2}ut = \frac{1}{2}vt + \frac{1}{2}ut$

And so $s = \frac{1}{2}(u + v)t$.

In passing, it is worth noting that, from this equation, the average velocity = total displacement / total time = $s/t = \frac{1}{2}(u + v)$.

In other words, for the specific case of uniform motion, average velocity is the average of the initial and final velocities.

The third equation follows by eliminating t between the first two:

Re-arranging the first equation, $t = (v–u)/a$.

Substituting this expression for t into the second equation, we get:

$s = \frac{1}{2}(v + u)(v–u)/a$.

Re-arranging, $2as = (v + u)(v–u) = v^2–u^2$ And so $2as = v^2–u^2$.

A stone is thrown upwards at a speed of 5.0 ms^{-1} from the top of a 30m high cliff.

At what speed does it hit the sea at the bottom of the cliff?

(Assume that the acceleration of the stone is 10 ms^{-2} downwards throughout.)

The problem with trying to solve this question using a velocity-time graph is that we have no information about the time taken. However, we can solve it very straightforwardly using the equation $2as = v^2–u^2$:

We need to choose a direction we are going to call positive, and it doesn't matter whether it is upwards or downwards; we just need to be consistent within the solution.

Let us consider the downwards direction as positive.

Initial velocity $u = –5.0\ ms^{-1}$

Net displacement = +30m

Acceleration = +10 ms^{-2}

Substituting these into the equation,

$v^2 – (–5.0)^2 = 2 \times 10 \times 30$

$\therefore v^2 – 25 = 600$

$\therefore v^2 = 600 + 25 = 625$

$\therefore v = \sqrt{625} = 25\ ms^{-1}$.

A car accelerates from 20 ms^{-1} to 30 ms^{-1} at a uniform acceleration of 5.0 ms^{-2}.

What distance does the car travel whilst it is accelerating?

Using $2as = v^2 - u^2$ and substituting in the values for u, v and a given in the question, we get: $2 \times 5.0 \times s = 30^2 - 20^2 = 900 - 400 = 500$.

∴ Distance travelled $s = 500 / (2 \times 5.0) = 500 / 10 = 50$m.

P3.2

Forces:

a. Understand that there are different types of force, including weight, normal contact, drag (including air resistance), friction, magnetic, electrostatic, thrust, upthrust, lift and tension.
b. Know and understand the factors that can affect the magnitude and direction of the forces in 3.2a.
c. Draw and interpret force diagrams.
d. Qualitatively understand resultant force, with calculations in one dimension.

Understand that there are different types of force, including weight, normal contact, drag (including air resistance), friction, magnetic, electrostatic, thrust, upthrust, lift and tension

Know and understand the factors that can affect the magnitude and direction of the forces in 3.2a

- A force is the application of a push or a pull to an object by another object.
- Forces are vector quantities, which means they have magnitude and direction.
- Forces are measured in newtons (N).
- There are various different types of force that arise in different situations. The main types of force are weight, normal contact, drag, friction, magnetic, electrostatic, upthrust, thrust, lift and tension.

The following table summarises the main types of force, what causes them, and the factors that determine their magnitudes and directions:

Type of force	Cause	Magnitude depends on	Direction
weight	mass in a gravitational field	mass gravitational field strength	downwards
normal contact	two solid objects in contact with each other		normal to surface of contact
drag	movement of an object through a fluid	speed cross-sectional area	opposite to relative motion

friction	relative sliding motion between two solid surfaces	nature of surface	opposite to relative motion
magnetic	two magnets or a current in a magnetic field	magnetic field strength current in wire	like poles repel opposite poles attract normal to current in wire
electrostatic	two charges or a charge in an electric field	electric field strength	like charges repel opposite charges attract
upthrust	solid immersed in a fluid	weight of fluid displaced	upwards
thrust	driving force from an engine	power of engine	in the direction of propulsion
lift	aerofoil (wing) moving through a fluid	speed density of fluid wing shape	normal to wing
tension	spring, string, wire etc. being stretched	extension	along string, spring, wire

Notes on the table:

- It is not uncommon for the normal contact force to be referred to as 'reaction'. However, this is to be strongly discouraged as it is misleading and can result in fundamental misunderstandings of Newton's third law. The word 'reaction' should never be used.
- Both drag (the most common form of which is air resistance) and friction are resistive forces acting on objects in motion. In the case of drag, the resistive force is caused by motion through a fluid (i.e. liquid or gas). In the case of friction, the resistive force is caused by relative motion between two solids in contact. It is not uncommon for all resistive forces to be considered together by using the label total resistive forces.

- The three forces associated with fields (weight associated with a gravitational field, magnetic force associated with a magnetic field and electrostatic force associated with an electric field) are the only types of force that can be exerted without contact. They are non-contact forces. All other forces are contact forces and are only exerted when the object is in contact with the appropriate solid / fluid / string etc.

When a football is kicked along the ground, it eventually comes to rest. A student says that 'this is because it has transferred its energy to friction'.

Explain why this is not a very good explanation and give a better explanation.

It is not a very good explanation because friction is a force and not a form of energy.

As the ball moves forward, there is a resultant friction force acting in the opposite direction. While moving, the ball does work (force × distance) against this friction force. This transfers kinetic energy of the ball to thermal energy in the surroundings.

Draw and interpret force diagrams

- The ability to analyse which forces are acting on an object and in which directions they are acting is essential to be able to solve problems involving the effects of forces on motion and equilibrium.
- Every force acting on an object is exerted by another object and is of one of the types listed in the table above.

Every force is represented by a force arrow that starts on the object on which the force acts. The arrow should start at the point on the object where the force acts (or can be considered to act) and should point away from it in the appropriate direction.

- Force arrows should be labelled with the type of force, the object that is exerting it and the magnitude of the force (either numerically or algebraically).
- Resultant force is not a type of force and is not an extra force that is acting on the object. For that reason, resultant force should never be added to force diagrams.
- Care should be taken when including arrows to represent other quantities, such as velocity or acceleration, to ensure that they cannot be mistaken for force arrows.

The construction of force diagrams is best illustrated by the consideration of a couple of examples:

Example 1

Consider an aircraft in horizontal flight. There are four main forces that act on the aircraft:

thrust exerted by the engines, acting at the engines and directed forwards

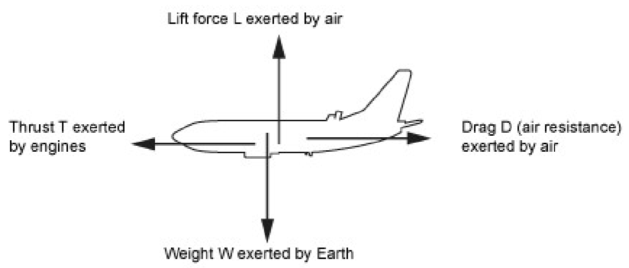
drag (air resistance) exerted by the air, acting on the mid-point of the cross-section and directed backwards

weight exerted by the Earth, acting at the centre of gravity of the aircraft and directed downwards

lift exerted by the air as the wings move through it, acting on the wings and directed upwards.

There will also be a small amount of upthrust, due to the fact that the aircraft is displacing air, but this is insignificant compared with the other forces and so may be ignored.

The force diagram can be drawn as follows:



Example 2

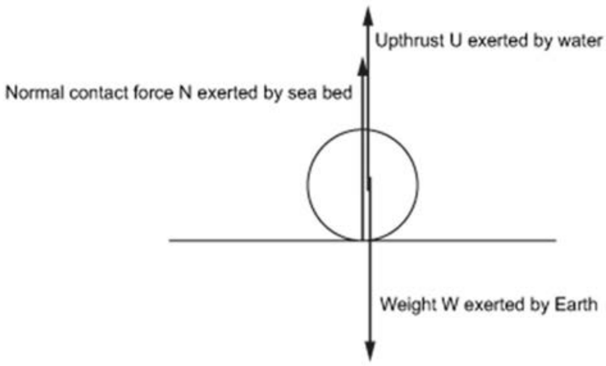
Consider a rock resting underwater on the seabed. There are three main forces that act on this rock:

weight exerted by the Earth, acting at the centre of gravity of the rock and directed downwards

normal contact force from the seabed, acting at the point of contact between the rock and the seabed and directed upwards

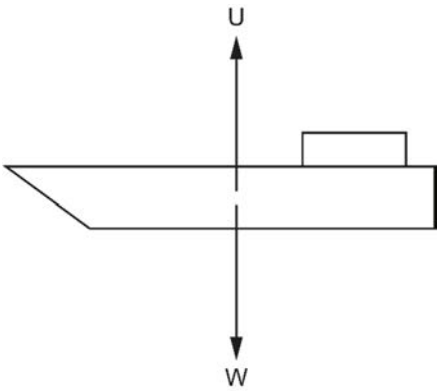
upthrust from the water displaced by the rock, acting at the centre of buoyancy of the rock and directed upwards.

There may also be some horizontal drag forces exerted on the rock by the water if the water is moving past it. In that case, there would be likely to also be horizontal friction forces acting on the rock from the contact with the seabed given that the water will be tending to cause relative movement between the rock and the seabed. But for this illustration we will focus on the vertical forces only. The force diagram can be drawn as follows:



Drawing diagrams, such as this one, where the forces all act through the same line, can be tricky. It is sometimes necessary, as here, to show the arrows slightly offset so that they can be distinguished even though they ought to be exactly in line.

A boat is moored at rest on the sea. Draw a free body diagram to show all the forces acting on the boat, and describe the nature of each force.



U is the upthrust force from the seawater (due to the fact that water has been displaced by the boat).

W is the weight of the boat, caused by the effect of the Earth's gravity.

The boat is at rest so there is no resultant force on it. The magnitudes of U and W are equal.

Qualitatively understand resultant force, with calculations in one dimension

- Once the forces acting on an object have been analysed and are known, it is possible to deduce the combined effect of all these forces. This combined effect is the vector sum of all the forces acting and is known as the resultant force acting on the object.
- It is important to understand that the resultant force is not a separate physical force acting on the object. It is not additional to weight, friction etc. It is the combination of all these forces when they are put together. That is why we do not add resultant force as an 'extra' force on force diagrams.
- The importance of the resultant force is, however, that it is the resultant force that determines the effect on the object of all the forces acting. It is the resultant force that determines, for example, whether the object accelerates or deforms.
- It follows that the reason we need to be able to calculate resultant force is that it is this value that will need to be used, for example, in the equation F = ma.
- In any particular dimension, the resultant force is calculated by subtracting the total force acting in one direction from the total force acting in the opposite direction. The resultant force will be in the direction in which the total force is larger.
- If the total force acting in one direction is equal to the total force acting in the opposite direction, then the resultant force is zero.
- An object may have forces acting in more than one dimension. In that case the resultant force in each dimension can be calculated by just using the forces that act in that dimension.

The calculation of resultant force is best illustrated by the consideration of a couple of examples:

Example 1

Consider the aircraft example above. Suppose we know that:

thrust T from the engines is 200kN

weight W from the Earth is 160kN

lift L from the air over the wings is 160kN

drag D from the air resistance is 140kN.

These forces are clearly acting in two dimensions, so calculating one single resultant force might be complicated.

However, we can consider the horizontal and vertical dimensions separately.

Horizontally, we have a 200kN trust force acting forwards and a 140kN drag force acting backwards. The difference between these forces is 200 – 140 = 60kN. The forward force is larger than the backwards force, and so the resultant horizontal force is 60kN forwards.

Vertically, we have a 160kN weight acting downwards and a 160kN lift force acting upwards. These forces are equal, and so there is no difference between them. Therefore, the resultant vertical force is zero. In this situation, the forces are said to be balanced.

It so happens that, because the vertical forces balance, the resultant horizontal force of 60kN forwards is also the total resultant force acting on the plane. But putting together the horizontal and vertical components is not so easy when there is a non-zero resultant force in both dimensions.

Example 2

Consider the rock on the seabed example in Section 3.2c. Suppose we know that:

weight W from the Earth is 4000N

upthrust U from the displaced seawater is 1500N

resultant force acting on the rock is zero.

For the resultant force to be zero, we know that the total upwards force must equal the total downwards force. There is only one downwards force, namely the 4000N weight. This means that the total upwards force must also be 4000N.

We know that the upthrust is 1500N, and so the normal contact force exerted by the seabed must be 4000N – 1500N = 2500N.

A firework of mass 0.40 kg is launched vertically upwards from the ground. The thrust from its rocket is 7.2 N. What is the resultant force on the rocket at the instant it is launched?

There are two forces acting on the rocket – thrust vertically upwards and weight vertically downwards. The resultant force is the difference of these two.

Weight is given by $W = mg = 0.40 \times 10 = 4.0$ N.

Hence the resultant force is $7.2 - 4.0 = 3.2$ N vertically upwards.

P3.3

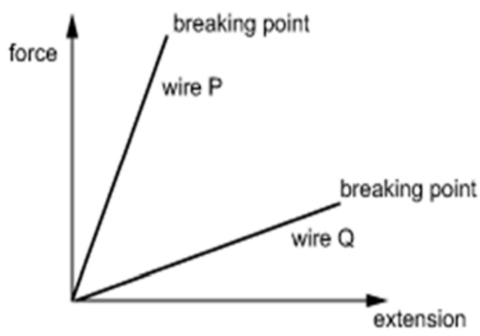
Force and extension:

a. Interpret force–extension graphs.
b. Understand elastic and inelastic extension, and elastic limits.
c. Know and be able to apply Hooke's law ($F = kx$), and understand the meaning of the limit of proportionality.
d. Understand energy stored in a stretched spring as: $E = \frac{1}{2}Fx = \frac{1}{2}kx^2$

Interpret force–extension graphs

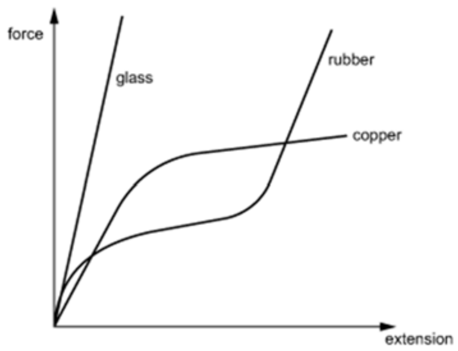
- When a spring / string / wire is pulled by equal and opposite external forces at each end as shown, it is said to be subjected to a tension force T:

- This tension force usually causes the length of the spring /string wire to increase slightly. The increase in length from the unstretched length is called the extension.
- It is usually the case that the greater the tension force, the greater the extension.
- A graph can be drawn that shows the force-extension characteristics of a particular spring / string / wire. The steeper this graph, the more force is required to produce a given extension. The shallower this graph, the greater the extension for a given force.
- A material is described as rigid if its deformation is small even with a large tension force.
- The characteristics of the spring / string / wire can change as it stretches, becoming more or less stiff. A point usually comes where the extension can increase no more and the material breaks:



- In the graph above, wire P is more rigid than wire Q
- Different types of material have different force-extension characteristics. For example:
- Copper wire stretches uniformly initially, but then suddenly stretches much more just before reaching the breaking point.
- Glass is very rigid and deforms only very slightly before breaking.

Rubber stretches non-uniformly.

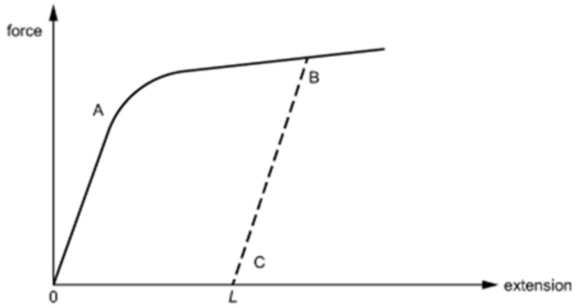


- Springs can undergo compression forces as well as tension forces. This results in a contraction rather than an extension.
- The diagram in the first bullet point of this section shows the external forces T being applied to each end of the material being stretched.
- The ends of the material exert an equal and opposite tension force T on the object that is stretching it. So, for instance, if the right-hand end is being pulled to the right by a hand with force T then the material exerts an equal force (T) on the hand to the left. This is a good example of a Newton 3 pair of forces.

Understand elastic and inelastic extension, and elastic limits

- If a spring or wire is stretched by a tension force, and then that tension force is removed, it may or may not contract back to its original length.
- The extension is elastic if the spring / wire returns to its original length when the tension force is removed.
- The extension is inelastic if the spring / wire does not return to its original length when the tension force is removed. This is sometimes known as plastic deformation. In this case the spring / wire returns to a new (greater) unstretched length, and is said to have undergone permanent stretching.
- Materials that undergo permanent stretching are sometimes known as ductile.
- The point on the force-extension graph where the extension goes from being elastic to inelastic is known as the elastic limit. (The elastic limit is often, but not always, the limit of proportionality; see below.)

Consider the force-extension characteristics of copper wire:



As long as the wire is not extended past point A, it remains within the elastic limit and the extension is elastic. So, if the wire is stretched from 0 to A by an external tension force, and that force is then removed, the extension of the wire will return to 0 and the wire will return to its original length.

Point A is the elastic limit.

If the wire is extended beyond the elastic limit, for example to point B, then the wire undergoes significant inelastic extension. When the extending force is removed, the wire does not return to its original length. There will, instead, be a small contraction, along the dotted line, back to point C. The permanent extension of the wire is L, which is the amount by which the unstretched length of the wire has increased.

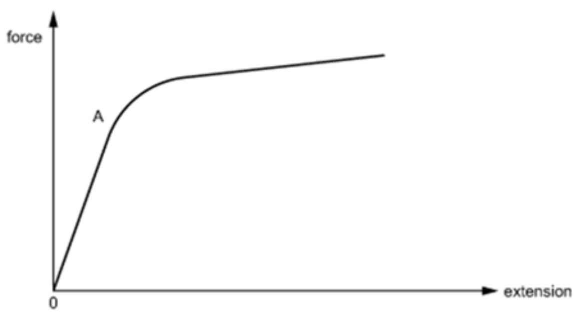
Line BC is approximately parallel to line 0A.

Know and be able to apply Hooke's law ($F = kx$), and understand the meaning of the limit of proportionality

- Springs and wires that are being extended within their elastic limits experience an extension that is proportional to the tension force. This is Hooke's law.
- The constant of proportionality in the relationship force $F \propto$ extension x is known as the spring constant k.
- It therefore follows that, for objects that obey Hooke's law, $F = kx$.
- Rearranging, this gives $k = F/x$, and this equation effectively defines spring constant as spring constant = force per unit extension.
- Units for spring constant can be Nm^{-1}, Ncm^{-1}, kNm^{-1} etc., depending on the units used for force and extension. When using the equation, it is important to ensure that the units used for the three quantities k, F and x are consistent with each other.
- Spring constant is a measure of rigidity. Materials with high spring constants require large forces to produce small extensions, and vice versa. It should be apparent that the spring constant is the gradient of the force extension graph.
- The limit of proportionality is the point on the force-extension graph beyond which Hooke's law is no longer obeyed. This is often (but not always) the same as the elastic limit. (See above.)

Example

Consider the force-extension characteristics of copper wire:



As well as being the elastic limit, point A also represents the limit of proportionality.

Suppose that, for a particular wire, the limit of proportionality occurred when a force of 4.0N caused an extension of

5.0 mm.

The spring constant of this wire is given by $k = F/x = 4.0 / 0.005 = 800\ Nm^{-1}$ (or $8.0\ Ncm^{-1}$).

Factors affecting spring constant

For a given material of wire, the greater the cross-sectional area, the greater the spring constant. And the longer the wire, the smaller the spring constant.

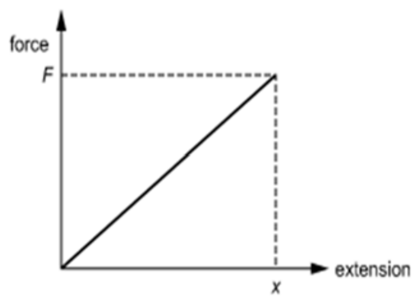
Combining springs

If two identical springs, each of spring constant k, are connected together in series, it follows that the same force will produce double the extension and so the spring constant is halved to $\frac{1}{2}$k.

If the same two springs are connected in parallel, then it follows that double the force will be needed to produce the same extension and so the spring constant is doubled to 2k.

Understand energy stored in a stretched spring as: $E = \frac{1}{2}Fx = \frac{1}{2}kx^2$

- When a spring or wire is being extended by a tension force, the work done by this force is given by the area under the force-extension graph.
- Provided the elastic limit has not been exceeded, this energy can be recovered when the external tension force is removed. Since it can be recovered, it is stored in the spring as elastic potential energy.
- It follows that elastic potential energy is given by the area under the force extension graph where the elastic limit has not been exceeded.
- Consider a spring/wire extended from its original length to extension x by external tension force F



Area under graph = $\frac{1}{2}$Fx.

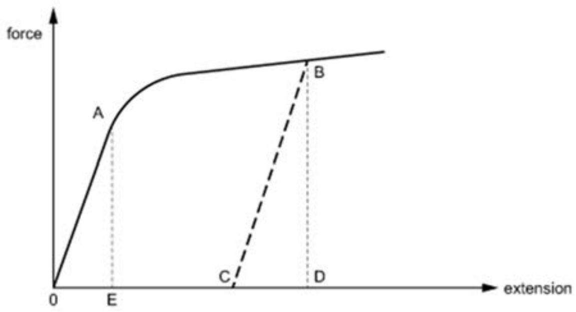
And since F = kx, it also follows that area = $\frac{1}{2}(kx)x = \frac{1}{2}kx^2$.

Thus, the energy stored in a stretched wire / spring (that obeys Hooke's law and has not exceeded the elastic limit) is given by $E = \frac{1}{2}Fx = \frac{1}{2}kx^2$.

Units of the quantities used in this formula must be consistent. For example, for a force in N (or a spring constant in Nm^{-1}), to get an energy in J, extension must be in m.

Retrievable and irretrievable energy

Consider the force-extension graph of copper wire that was discussed earlier:



If the wire is stretched only as far as point A, then all work done in stretching the wire (area under the graph as far as A) is retrievable because the elastic limit has not been exceeded. So the energy stored in the wire is the area of triangle OAE.

If the wire is stretched from 0 to B, the total work done by the tension force is given by the area under the whole curve from 0 to B. However, only an amount of energy equivalent to the area of triangle BCD is recoverable (the area under the graph as the wire returns to its new unstretched length). The potential energy stored in the wire (as recoverable energy) is therefore area BCD.

The remainder of the work done in stretching the wire from 0 to B is not recoverable. It is, instead, converted into heat as a result of the permanent extension of the wire.

A steel spring has spring constant 25 Nm^{-1} and it supports an object of mass 0.30 kg.

What is the extension of the spring?

How much elastic potential energy is stored in the spring?

$F = kx$ so $x = F/k = mg/k = (0.30 \times 10) / 25 = 0.12$ m (12 cm)

$E = \frac{1}{2}kx^2 = \frac{1}{2} \times 25 \times 0.12^2 = 0.18$ J

P3.4

Newton's laws:

a. Know and understand Newton's first law as: 'a body will remain at rest or in a state of uniform motion in a straight line unless acted on by a resultant external force'.
b. Understand mass as a property that resists change in motion (inertia).
c. Know and understand Newton's second law as: force = mass × acceleration
d. Know and understand Newton's third law as: 'if body A exerts a force on body B then body B exerts an equal and opposite force of the same type on body A'.

Know and understand Newton's first law as: 'A body will remain at rest or in a state of uniform motion in a straight line unless acted on by a resultant external force'

- Newton's first law states that 'a body will remain at rest or in a state of uniform motion in a straight line unless acted on by a resultant external force'.

There are two key elements of this law that are important to understand:

- The law is concerned with the effect of resultant force on a body. The details of how this resultant force is made up are immaterial. For example, if there are two forces of weight and normal contact force acting on a body, it is meaningless to consider what effect either of these individual forces has on the body. The effect of the forces on the body is determined solely by the resultant of the two forces.
- The law establishes that a resultant force on a body acts to change its velocity. In other words, it causes the body to accelerate. It is a common misconception that resultant forces 'cause velocity', that an object that is moving 'must have a resultant force acting on it to keep it moving'. But this is not correct; resultant forces cause acceleration, and so an object that is in motion will remain at constant velocity in the absence of a resultant external force. If a force acts on the object, that velocity will change.

The misconception in the second bullet point above often arises from confusing individual forces with resultant force in the first bullet point. By way of illustration, consider the following examples:

Example 1

An aircraft in level flight at constant speed in a straight line is travelling at constant velocity and therefore has no resultant force acting on it. But this does not mean that there are no forces acting on the aircraft. The force diagram in Example 1 in section P3.2 shows the four main forces acting on an aircraft in flight.

For the resultant force vertically to be zero, the weight downwards must have the same magnitude as the lift upwards.

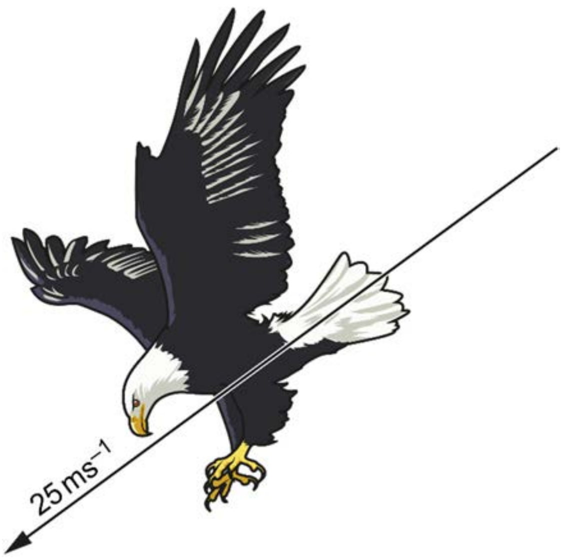
For the resultant force horizontally to be zero, the thrust forwards must have the same magnitude as the drag (air resistance) backwards.

It is clear that the engines must produce thrust to keep the aircraft at constant speed, but the thrust is not the resultant force. The thrust is needed only because there is drag acting on the aircraft and this drag needs balancing by an equal forwards force to give zero resultant force.

Example 2

By contrast, a spacecraft in deep space experiences no drag. Whilst thrust from rockets must have been used at some point in the past to accelerate the spacecraft from rest, once the rocket is moving, there is no need for thrust to 'keep the spacecraft moving'. In the absence of drag, the spacecraft will keep moving at constant velocity ad infinitum until acted on by a force. In deep space, if the rockets are fired, then the thrust produced will exert a resultant force on the spacecraft causing it to accelerate.

A bird is diving at a constant velocity of 25 ms^{-1} at an angle of 45° to the horizontal as shown.



Which statement about forces acting on the bird is correct?

1) There is a resultant downward force on the bird.

2) There is a resultant force on the bird in the direction it is moving.

3) The resultant force on the bird is zero.

4) There is a resultant force on the bird opposite to the direction in which it is moving.

The bird is moving at constant velocity (not accelerating) so there is no resultant force. This is an example of Newton's first law of motion.

There are several forces acting on the bird: its weight, lift from the wings and drag opposing its motion. However, these are vector quantities that add together to give zero resultant force.

Statement 3) is correct.

Understand mass as a property that resists change in motion (inertia)

- Mass is the property of an object that resists acceleration (changes in motion).
- Another name for this property is inertia.
- The larger the mass of an object, the greater the force needed to cause a given acceleration.
- Units of mass are grams (g), kilograms (kg) etc.

There is a direct relationship between the amount of matter present in an object and its resistance to acceleration.

Thus, the more matter that is present in the object, the greater its mass.

Mass is not to be confused with weight, which is the force of gravity acting on an object. When an object feels 'heavy' when trying to lift it, it is the weight of the object that is being detected.

The mass of an object is only sensed when trying to accelerate it (for example, by trying to push it from rest along a frictionless surface) or stop it from moving. The greater the force needed to do this, the greater the mass of the object.

Know and understand Newton's second law as: force = mass × acceleration

- Newton's second law can be stated as 'force = mass × acceleration'.
- A force of 1 Newton is defined as that force that, when acting on a body of mass 1 kg, causes an acceleration of 1 ms^{-2}.
- When using the equation F = ma, it is important to ensure that the units of the quantities used are consistent. Mass in kg and acceleration in ms^{-2} gives a force in N.
- It is important to understand that the 'force' in this equation means the resultant force acting on the object. It is solely the resultant force that determines the acceleration of the object; the details of individual forces that make up the resultant force are immaterial.
- The acceleration of the object is in the same direction as the direction of the resultant force.

Links to previous sections

Newton’s second law follows on from Newton’s first law and from an understanding of inertial mass.

Newton’s first law is effectively a special case of Newton’s second law where the resultant force is zero (zero resultant force produces zero acceleration).

Newton’s second law quantifies the concept of inertial mass by establishing that, for a constant external resultant force, the acceleration of an object is inversely proportional to its reluctance to accelerate.

F = ma is a special case

Newton’s second law is more correctly stated as ‘resultant force is proportional to the rate of change of momentum’ of an object.

In the SI system of units, the constant of proportionality is unity such that the law becomes ‘force = rate of change of momentum’.

The commonly used version of the law, ‘force = mass × acceleration’, is a special case of the more general version of the law that applies in situations where the mass is constant.

It follows that ‘F = ma’ can only be used in situations where mass is constant and where the units are as stated above. In situations where the mass is continuously changing, the more general version of the law, in terms of rate of change of momentum, has to be used.

Example

The application of Newton’s second law is best illustrated through an example.

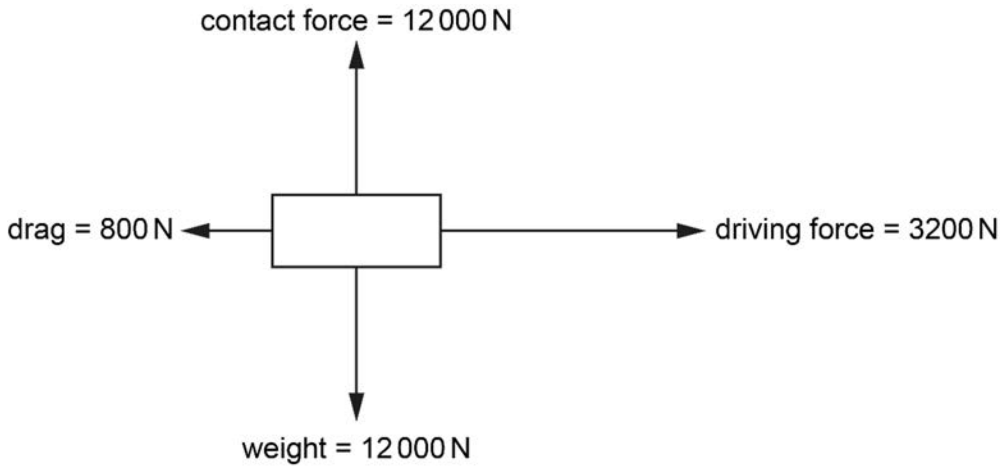
Consider the aircraft in Example 1 in section P3.2. We established in that example that the aircraft experienced zero resultant force vertically and a resultant force horizontally of 60kN forwards.

The weight of this aircraft is 160kN, which means its mass is 16 000 kg.

The acceleration of this aircraft is therefore given by a = F/m = 160 000 / 16 000 = 3.75 ms^{-2} forwards.

(Notice that the acceleration is in the same direction as the resultant force.)

The diagram shows all the forces acting on a car of mass 1200 kg.



Describe the motion of the car.

The resultant force on the car is 3200 − 800 = 2400 N to the right.

This gives it an acceleration a = F/m = 2400 / 1200 = 2.0 ms^{-2} to the right.

It could be moving to the right and increasing its speed or moving to the left and decreasing its speed.

Know and understand Newton's third law as: 'If body A exerts a force on body B then body B exerts an equal and opposite force of the same type on body A'

- Newton's third law states that 'if body A exerts a force on body B then body B exerts an equal and opposite force, of the same type, on body A'.

Newton's third law is one of the most poorly understood principles of Physics, almost entirely as a result of use of confusing language surrounding its discussion. To assist with a correct understanding of the law, the common misconceptions are discussed in detail below.

Most of the confusion is caused by incorrectly learning the law in terms of the terms 'action' and 'reaction', and use of these terms is strongly discouraged. It is virtually impossible to properly understand this law if it is learnt in terms of 'every action has an equal and opposite reaction':

- The words 'action' and 'reaction' convey an idea that one of the forces in the pair is the cause and the other one is a response to that force. This is a fundamental misconception. The two forces in a Newton 3 interaction pair of forces are exerted simultaneously, with no distinguishing of one as being the cause and the other being a consequence.
- The word 'reaction' is also used by some (but equally to be discouraged) to refer to normal contact forces. This use of the word 'reaction', to mean both a normal contact force and the Newton 3 response to an 'action', leads to the most common misunderstanding of the law, namely that normal contact force and weight are a Newton 3 pair. They are not.

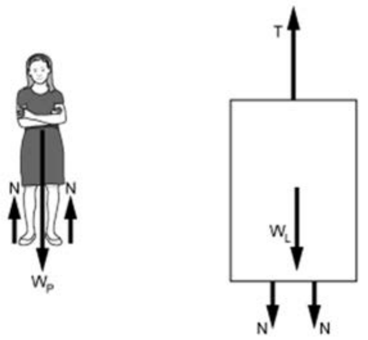
Whilst the 'equal and opposite' aspects of the law are well known, there are two aspects of the law that are less well known and appreciated. They are that:

- the two forces in the pair are of the same type (e.g. both friction, or both normal contact, or both gravitational etc.)
- the two forces in the pair act on different objects (one on body A and the other on body B).

Neither of these fundamental aspects of Newton's third law are expressed in the common incorrect versions of the law that are often learnt. Two forces that are of different types or that both act on the same object cannot possibly be a Newton 3 pair.

Example

Consider the example of a person standing in a lift. The forces acting on the person and on the lift are, separately, as follows:



The person experiences weight downwards and normal contact force upwards. (To help to illustrate the point, the normal contact forces are shown separately at each of the two points of contact, but they are often shown collectively as just a single normal contact force.) The upwards and downwards forces may or may not be equal, depending on whether the person is accelerating. If the person is not accelerating, then the resultant force on him is zero and the upwards and downwards forces will be equal. But this is nothing to do with Newton's third law – it is actually a consequence of Newton's first law.

The lift experiences its own weight downwards, the normal contact forces from the person downwards, and the tension in the cable upwards. Again, the upwards and downwards forces may or may not balance but if they do, that is a consequence of Newton's first law. Again, nothing to do with Newton's third law.

Where Newton's third law comes into this example is that it tells us, regardless of whether the system is accelerating or not, there are pairs of forces that will always be equal and opposite. They are the normal contact force exerted by the lift on the person, and the normal contact force exerted by the person on the lift. Note the following features of this example of a Newton 3 pair:

The two forces in the pair are of the same type (both normal contact).

The two forces on the pair are equal and opposite.

The two forces in the pair act on different objects (one on the person, exerted by the lift, and the other on the lift, exerted by the person).

This example illustrates the characteristics of every Newton 3 pair of forces and, hopefully, illustrates the importance of understanding the 'exerted by body A on body B' and 'exerted by body B on body A' aspects of the statement of the law.

Students who learn that 'every action has an equal and opposite reaction' usually fall into the trap of thinking that the weight of the person and the normal contact force acting on the person form a Newton 3 pair. It is hoped that this example illustrates why the terms 'action' and 'reaction' are unhelpful in properly understanding Newton's third law and should never be used.

Putting component parts of a system together

If we now consider the person and lift together, as a single object, it becomes apparent that the normal contact forces between them are internal forces within the system that cancel out (they are equal and opposite, so they must), leaving just the two weights and the tension in the cable as the external forces acting on the compound system:

An apple rests on the ground. Identify the forces acting on the apple and explain how Newton's third law applies to this situation.

The two forces acting on the apple are its weight acting vertically downwards, and a contact force from the ground acting vertically upwards.

Weight is a gravitational force exerted on the apple by the Earth, so the apple must exert an equal gravitational force in the opposite direction on the Earth.

The ground exerts an upward contact force on the apple, so the apple must exert a downward contact force of the same magnitude on the ground.

P3.5

Mass and weight:

a. Know and understand the difference between mass and weight.
b. Know and be able to apply gravitational field strength, g, approximated as 10 N kg^{-1} on Earth.
c. Know and be able to apply the relationship between mass and weight: $w = mg$
d. Understand free-fall acceleration.
e. Know the factors affecting air resistance.
f. Understand terminal velocity and the forces involved.

Know and understand the difference between mass and weight

- Mass is a measure of the resistance of an object to acceleration (inertia). It is a reflection of the amount of material present in a sample – the greater this is, the harder it is to change the motion of (i.e. accelerate) the object.
- Mass is a scalar quantity, with units of grams (g), kilograms (kg) etc. It has no direction associated with it.
- Weight is a force. It is the name given to the gravitational force acting on an object that is in a gravitational field.
- Weight, being a force, is a vector quantity, with units of newtons (N), kilonewtons (kN) etc. The direction of this force is always in the direction of the gravitational field causing it – near the surface of a planet such as Earth, downwards, towards the centre of the planet.

Mass and weight are commonly confused concepts – particularly as the word 'weight' is used in everyday language to mean 'mass'. Of course, there is a relationship between the two (that we will explore in the next section) – the greater the mass of an object, the greater will be the gravitational force (weight) acting on it in a given gravitational field.

This relationship actually forms the basis for measuring mass, which may add to the common confusion. Mass (the property of inertia) is quite difficult to measure directly, so most devices designed to measure mass are actually measuring weight but are calibrated to give the equivalent mass of that weight in the gravitational field for which it is calibrated.

For example, the types of 'weighing scales' commonly found in households, and balances found in laboratories, are usually designed to measure a force applied to it. This force is indirectly, through the balance of forces and equilibrium conditions, equivalent to the weight of the object. But the scale that the instrument reads is calibrated to give a mass, in g or kg, that is equivalent to that weight on Earth.

If the same measuring instrument was used in a situation where the object is not in equilibrium (for example, in an accelerating lift) or not in the gravitational field for which it was calibrated (for example, on the Moon), then it would give an incorrect reading for the mass of the object. This despite the fact that the mass has not actually changed. There is still the same amount of matter present, and it has the same resistance to acceleration. But it is not the mass being measured – it is the force that the object exerts on the instrument that is being measured.

So, the fact that instruments designed to give a mass in kg are actually giving the mass-equivalence of a force is possibly another reason for the common confusion between mass and weight.

Know and be able to apply gravitational field strength, g, approximated as 10 N kg^{-1} on Earth

- In any given gravitational field, the gravitational force acting on an object is proportional to its mass.
- The gravitational field strength is defined as the magnitude of the gravitational force acting per unit mass of an object in the field.
- Gravitational field strength is usually denoted by the symbol g.
- Unit of gravitational field strength = unit of force / unit of mass = N kg^{-1}.
- The gravitational field near the surface of the Earth is approximately 10 N kg^{-1}.

Other gravitational fields

The gravitational field created by a planet near its surface depends on the mass and radius of the planet. In general, the more massive the planet, the greater the gravitational field it creates. For example:

- The gravitational field near the surface of the Moon is approximately 1.6 N kg^{-1}.
- The gravitational field near the surface of Jupiter is approximately 26 N kg^{-1}.
- The gravitational field near the surface of the Sun is approximately 280 N kg^{-1}.

This means that if an object was moved from the Earth to another planet, its mass would remain the same (same amount of material so same resistance to acceleration) but its weight would change according to the gravitational field of the other planet.

Inertial and gravitational mass

The concept of mass as the property of resistance to acceleration (inertial mass) and the concept of mass as the property on which a gravitational force acts in a gravitational field (gravitational mass) are essentially entirely different concepts. However, both properties can be measured and scientists have found no difference between the two. For practical purposes, mass as a property that resists acceleration and mass as a property on which a force acts in a gravitational field appear to be entirely equivalent.

Know and be able to apply the relationship between mass and weight: w = mg

- It follows from the definition of gravitational field (as the gravitational force acting per unit mass of an object) that the gravitational force acting on an object is given by the product of gravitational field and mass.
- Weight is the gravitational force acting on the object.
- Therefore, the weight W of an object is related to its mass m and the gravitational field g by the equation W = mg.
- The units used for mass and weight must be consistent with the units used for gravitational field. For example, if g is in N kg^{-1}, then m must be in kg and W in N.
- Weight will always act in the same direction as the direction of the gravitational field.

It is a common misconception that a mass of 1kg is equal to a weight of 10N. However, mass and weight are completely different concepts.

Suppose, for instance, that we have a car travelling at a speed of 10 ms^{-1}. It is true that this car will travel a distance of 10m in a time of 1s. But we would not dream of stating that 1s is equal to 10m – a distance cannot equal a time, because distance and time are different concepts. And, of course, for a car travelling at a different speed, the distance travelled in a time of 1s would be different. So, all we can say is that a car travelling at 10 ms^{-1} will travel 10m in 1s. That is not the same thing as saying 1s equals 10m.

And so, it is with mass and weight. It is true to say that a mass of 1kg, in the Earth's gravitational field, will experience a weight of 10N. But that does not mean that 1 kg is equal to 10N. And if this mass of 1 kg was taken to the Moon, its weight would be 1.6 N, not 10 N.

Mass and weight are different concepts. There is a relationship between them (via gravitational field), but they cannot be equated as being different measurements of the same thing.

The Earth's gravitational field strength is often taken as 10 Nkg^{-1} but actually it varies from place to place. In Oslo it is about 9.83 Nkg^{-1} whereas in Kuala Lumpur it is only 9.77 Nkg^{-1}. A person has a weight of 812 N in Oslo.

What is the mass of the person?

How much would the person weigh in Kuala Lumpur?

W = mg so m = W/g = 812/9.83 = 82.6 kg.

The mass is the same in Oslo and Kuala Lumpur, so the weight in Kuala Lumpur will be W = mg = 82.6 × 9.77 = 807 N.

Understand free-fall acceleration

- An object that is falling due to gravity has no other forces acting on it apart from the force due to the gravitational field (its weight).
- It follows from the absence of any other forces that the resultant force acting on the object is equal to its weight.
- The resultant force is therefore given by F = W = mg.
- Newton's second law tells us that the acceleration of an object of mass m is equal to F/m, where F is the resultant force on the object.
- The acceleration of an object falling freely due to gravity is thus given by a = F/m = (mg)/m = g.
- Thus, an object falling freely due to gravity falls with an acceleration that (in ms^{-2}) is numerically equal to the gravitational field strength (in N kg^{-1}).
- This acceleration is called the acceleration of free-fall.

Weightlessness

Technically, an object is ‘weightless’ when it has no weight. For an object with mass, this can only be true if it is not in a gravitational field – i.e. in deep space.

More commonly, objects are described as ‘weightless’ in situations where they do actually have weight but where this weight is not sensed. This is then ‘apparent weightlessness’ rather than true weightlessness.

The thing that causes us to sense our weight is when there is another force acting on us to stop us from accelerating. For example, when we sit on a chair, there is a normal contact force from the chair that balances our weight and stops us falling to the floor. It is this normal contact force that is the thing that we ‘feel’ and that gives rise to our sensation of our weight.

If the forces that stop us from accelerating are removed, we stop sensing them and therefore stop sensing our weight.

But, of course, in the absence of other forces, our weight causes us to accelerate at the acceleration of free-fall.

In other words, in any situation where we have weight and that weight is actually causing us to accelerate at the acceleration of free-fall, there are no other forces acting and we will feel ‘weightless’. We are not actually weightless (because we do have weight), but we are experiencing ‘apparent weightlessness’.

This sensation of apparent weightlessness is experienced by astronauts in orbit around the Earth (because they are constantly falling towards the Earth with an acceleration equal to the acceleration of free-fall). And it can be simulated for astronauts in training by accelerating an aircraft downwards towards the Earth’s surface at the acceleration of free-fall.

A person standing in a lift that is accelerating downwards at the acceleration of free-fall would feel ‘weightless’ because, to achieve that acceleration, the support force from the floor of the lift would need to become zero. The person therefore no longer experiences a contact force and therefore experiences the sensation of ‘apparent weightlessness’. Relative to the lift (which is also accelerating at the acceleration of free-fall), the person appears to be floating in space.

Which of the following statements is/are correct?

1) If a hammer and a feather were dropped at the same time from the same height above the Moon's surface, they would hit the surface at the same time.

2) Astronauts inside the International Space Station are weightless because the space station orbits outside the Earth's atmosphere.

3) If a plane dives vertically downwards with an acceleration of 10 ms^{-2} then passengers inside the plane will feel 'weightless'.

The Moon has no atmosphere, so the only forces acting on the hammer and the feather are their weights. This means that they are in free fall. Objects in free fall in the same gravitational field fall with the same acceleration, so they hit the surface at the same time. Statement 1) is correct.

Astronauts inside the International Space Station and the Space Station itself are both free falling in the Earth's gravitational field, so they both fall at the same rate and the astronauts feel 'weightless'. However, it is their weight that makes them fall, so they are only apparently weightless because they are falling with the same acceleration.

Statement 2) is incorrect.

Passengers inside the plane also fall with an acceleration of 10 ms^{-2} so they are able to 'float' inside the cabin without any force from their surroundings. The fact that they are falling is due to their weight, so they are only apparently 'weightless'. Statement 3) is correct.

Know the factors affecting air resistance

- Air resistance is a type of drag force, experienced when the fluid being moved through is air.
- It acts on an object in the opposite direction to its motion through the air.
- The magnitude of the air resistance force increases with increasing speed of motion.
- The magnitude of the air resistance force also depends on the nature and effective cross-sectional area of the aspect of the object that presents itself to the air flow. In general, the larger the effective cross-sectional area of the object normal to the direction of motion, the larger the air resistance force.
- The final factor affecting the air resistance force is whether the air flow is streamlined or turbulent. Turbulent air flow gives rise to a larger air resistance force than streamlined air flow.
- Whether the air flow is streamlined or turbulent depends in part on how aerodynamic the surface is over which the air is flowing. The more aerodynamic the surface, the more likely the air flow is to be streamlined.

The relationship between speed of motion and turbulence is a complex one.

The greater the speed of movement of the object through the air, the more likely it is that the air flow will become turbulent. And the type of air flow has an effect on the relationship between speed and the magnitude of the air resistance force.

In general, for streamlined air flow, air resistance tends to be proportional to speed.

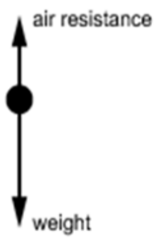
In general, for turbulent air flow, air resistance tends to be proportional to speed^2.

In both cases, the constant of proportionality will reflect the effective cross-sectional area of the object.

It therefore follows that significant increases in speed result in a very significant increase in air resistance – firstly, because air resistance increases with speed anyway, and secondly because when the air flow becomes turbulent, the effect of the increase in speed on air resistance is squared.

Understand terminal velocity and the forces involved

- An object falling due to gravity in air experiences two main forces: weight downwards and air resistance upwards.



- We know that the weight is constant and is given by W = mg
- We also know he air resistance increases with speed [and is zero when speed is zero]
- It therefore follows that an object falling from rest experiences an initial acceleration equal to the acceleration of free-fall. However, as the object accelerates (and the speed increases), the air resistance increases, resulting in a decrease in the resultant force downwards.
- As the object gets faster, its acceleration therefore decreases.

- Eventually the object is moving fast enough for the air resistance to equal the weight. At this point the resultant force (and hence acceleration) becomes zero and the object can no longer move any faster.
- The speed at which the air resistance equals the weight is called the terminal velocity of the object.

If an object is moving initially faster than its terminal velocity, the air resistance force is greater than the weight. This gives rise to a resultant force in the opposite direction to motion that causes the speed of the object to decrease until it becomes equal to the terminal velocity.

If an object moving at terminal velocity changes size or shape, such that a different effective cross-sectional area is presented to the air flow, the air resistance force changes, leading to a new terminal velocity. An example of this is a parachutist opening their parachute. An increased effective cross-sectional area requires a lower speed to generate the same air resistance force, and so the speed at which the air resistance and weight are the same becomes lower.

Air is only one example of a fluid through which an object moves, generating drag. Water is another example – a boat moving through water will experience a drag force in the opposite direction to its motion that increases with speed. For any given driving force from the engine, there will be an effective terminal speed at which the boat can move through the water.

A free fall parachutist is falling at a constant velocity of 50 ms^{-1} when the parachute opens. Soon after this the parachutist is falling at a constant velocity of 8.0 ms^{-1}.

Which of the following statements is/are correct?

1) The total air resistance when the parachute is open and the parachutist is falling at 8.0 ms^{-1} is greater than when it is closed and the parachutist is falling at 50 ms^{-1}.

2) There is a brief period between these two terminal velocities when the resultant force on the parachutist is upwards.

3) When in free fall at terminal velocity the resultant force on the parachutist is downwards.

4) When the parachute is open and the parachutist is at the slower terminal velocity the resultant force on the parachutist is upwards.

At terminal velocity (either 50 ms^{-1} with the parachute closed or 8.0 ms^{-1} with parachute open) the parachutist is not accelerating, so the resultant force is zero. This makes statements 1), 3) and 4) incorrect.

However, when the parachute first opens, the parachutist is falling faster than terminal velocity with a parachute so the air resistance force on the parachute is greater than the weight of the parachutist so there is a resultant upward force that slows the fall to 8.0 ms^{-1} (when forces are again balanced). Therefore statement 2) is the only correct statement.

P3.6

Momentum:

a. Know and be able to apply: momentum = mass × velocity, $p = mv$
b. Know and be able to use the law of conservation of momentum in calculations in one dimension.
c. Know and be able to apply: force = rate of change of momentum.

Know and be able to apply: momentum = mass × velocity, p = mv

- An object with mass that is moving has momentum. The momentum of the object is defined as the product of the mass of the object and its velocity.
- This leads to the equation p = mv, where p is the momentum of an object of mass m moving with velocity v.
- Since velocity is a vector quantity, momentum is also a vector quantity. The momentum of the object is in the same direction as its velocity.
- The units of momentum are the units of mass × the units of velocity. For masses in kg and velocities in ms^{-1}, the units of momentum are often left as kg m s^{-1}.
- However, this unit can also be (and sometimes is) written as Ns.
- When combining the momenta of different components of a system together, to find the total momentum of the system, it is important to take direction into account. The combined momentum is the vector sum of the separate components.

Example 1

Consider a car of mass 1200 kg moving at velocity 20 ms^{-1} due north.

The momentum of this car is mass × velocity = 1200 × 20 = 24 000 kg m s^{-1} due north.

This can also be written as 24 000 Ns due north.

Example 2

Consider a lorry moving at velocity 10 ms^{-1} in the positive direction with momentum +200 000 Ns.

Mass of lorry = momentum / velocity = 200 000 / 10 = 20 000 kg.

Example 3

Consider a system consisting of two balls.

Ball A, of mass 0.10kg, is moving with velocity 5.0 m s^{-1} to the right.

Ball B, of mass 0.20kg, is moving with velocity 4.0 m s^{-1} to the left.

Momentum of ball A = mass × velocity = 0.10 × 5.0 = 0.50 kg m s^{-1} to the right.

Momentum of ball B = mass × velocity = 0.20 × 4.0 = 0.80 kg m s^{-1} to the left.

We can combine these two momenta together to find the total momentum of the system, but in doing so we need to remember that momentum is a vector quantity. This means that we need to take direction into account.

Total momentum of system =

0.50 kg m s^{-1} to the right + 0.80 kg m s^{-1} to the left = 0.30 kg m s^{-1} to the left.

A rubber ball of mass 0.35 kg is thrown against a wall and bounces back in the opposite direction. The velocity of the ball immediately before the collision with the wall is 5.0 ms^{-1} horizontally to the right. The velocity of the ball immediately after the collision is 4.0 ms^{-1} horizontally to the left. What is the change of momentum of the ball as a result of the collision?

Momentum is a vector so we must take direction into account.

The momentum before the collision is mu = 0.30 × 5.0 = 1.5 $kgms^{-1}$ (to the right).

The momentum after the collision is mv = 0.30 × 4.0 = 1.2 $kgms^{-1}$ (to the left).

The change in momentum is final momentum minus initial momentum = 1.2 $kgms^{-1}$ – (–1.5 $kgms^{-1}$) = 2.7 $kgms^{-1}$ to the left.

Know and be able to use the law of conservation of momentum in calculations in one dimension

- In any closed system involving the interaction between two (or more) objects, the total momentum of the system before the interaction is equal to the total momentum of the system after the interaction.
- This is known as the law of conservation of momentum and applies universally, without exception.
- A closed system is one on which no external forces act.
- In applying the law of conservation of momentum, it is important to remember that momentum is a vector quantity and that momentum is conserved in every direction.

Example

Consider the system consisting of two balls from Example 3 above.

Ball A, of mass 0.10kg, is moving with velocity 5.0 m s^{-1} to the right.

Ball B, of mass 0.20kg, is moving with velocity 4.0 m s^{-1} to the left.

Suppose the two balls collide, and after the collision, ball B is moving with velocity 1.0 m s^{-1} to the left.

Initially:

Momentum of ball A = mass × velocity = 0.10 × 5.0 = 0.50 kg m s^{-1} to the right.

Momentum of ball B = mass × velocity = 0.20 × 4.0 = 0.80 kg m s^{-1} to the left.

Therefore, total momentum of system = 0.50 kg m s^{-1} to the right + 0.80 kg m s^{-1} to the left = 0.30 kg m s^{-1} to the left.

After the collision:

The law of conservation of momentum tells us that the final momentum of the system is the same as the initial momentum of the system, so the total final momentum = 0.30 kg m s^{-1} to the left.

Final momentum of ball B = mass × velocity = 0.20 × 1.0 = 0.20 m s^{-1} to the left.

From this we can deduce that the final momentum of ball A = 0.10 m s^{-1} to the left.

Therefore, final velocity of ball A = momentum / mass = 0.10 left / 0.10 = 1.0 m s^{-1} to the left.

We have deduced that, in this example, the final velocities of both balls are the same, which means that they effectively coalesce and move together with the same velocity.

An adult and a child are stationary next to each other on an ice rink. The adult has a mass of 70 kg and the child has a mass of 35 kg. The adult pushes the child and the child slides away at 0.80 ms^{-1}.

What is the speed and direction of the adult immediately after releasing the child? (Assume that friction is negligible.)

Momentum is conserved and momentum is a vector, so the momentum gained by the child must be equal in magnitude but opposite in direction to the momentum gained by the adult.

Child's momentum = 35 × 0.80 = 28 $kgms^{-1}$

Adult's momentum = 70v = 28 $kgms^{-1}$ (where v is the speed of the adult) v = 28/70 = 0.40 ms^{-1}

The direction of the adult's motion is opposite to the direction in which the child moves.

Know and be able to apply: force = rate of change of momentum

- The action of an external resultant force on an object will change its momentum.
- The change in momentum = external resultant force × time.
- This is why the unit of momentum is sometimes written as Ns.
- Hence the external resultant force is equal to the rate of change of momentum of the object.

Proof

Consider an object of mass m moving with an initial velocity u.

Suppose an external resultant force F acts on the object for time t.

The acceleration of the object is given by Newton's second law: acceleration = force F / mass m.

Therefore, change in velocity = acceleration × time = Ft/m.

Therefore, final velocity = u + Ft/m.

Therefore, final momentum = mass × velocity = m(u + Ft/m) = mu + Ft.

Therefore, change in momentum = final momentum – initial momentum = (mu + Ft)–mu = Ft.

So, change in momentum = Ft, and force F = (change in momentum) /t = rate of change of momentum.

Conservation of momentum and Newton's third law

Consider two objects, X and Y, colliding with each other, in the absence of any external forces.

There are no external forces, but each object exerts a force on the other during the collision.

We know, from Newton's third law, that the force exerted by X on Y is equal and opposite to the force exerted by Y on X.

Let the magnitude of this force be F.

The forces that each object exerts on the other are clearly exerted for the same time, t.

Hence, change in momentum of X (Ft) is equal and opposite to the change in momentum of Y (Ft).

Hence, the total momentum of the system remains constant.

We can see from this analysis that the law of conservation of momentum and Newton's third law both follow directly from each other. They are essentially equivalent to each other.

It is sensible to bend your knees when you land after jumping. Which of the following statements about this is/are correct?

1) Bending your knees on landing reduces the acceleration of your body as it comes to rest.

2) Bending your knees on landing reduces your momentum just before landing.

3) Bending your knees on landing reduces the loss of gravitational potential energy in the jump.

4) Bending your knees on landing reduces the force on your body when you land.

5) Bending your knees on landing reduces the rate of change of momentum as you land.

The momentum gained and the change of gravitational potential energy as you fall are both determined by how far you have fallen and not by how you land. So, statements 2) and 3) are both incorrect.

Bending your knees as you land increases the time of stopping. This reduces the acceleration of your body because acceleration is change of velocity divided by time and you have the same change of velocity over a longer time. So, statement 1) is correct.

You also have the same change of momentum over a longer time, so the rate of change of momentum is reduced and statement 5) is correct. Newton's second law states that force is equal to rate of change of momentum so if the rate of change of momentum is reduced, then so is the force on your body and statement 4) is correct.

P3.7

Energy:

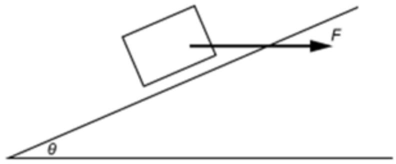
a. Know and be able to apply: work = force × distance moved (in direction of force)
b. Understand work done as a transfer of energy.
c. Know and be able to apply: gravitational potential energy = mgh, where h is the difference in height of the object.
d. Know and be able to apply: kinetic energy = $\frac{1}{2}mv^2$
e. Know and be able to apply: $\text{power} = \frac{\text{energy transfer}}{\text{time}}$
f. Know and be able to use in calculations the law of conservation of energy.
g. Understand the concepts of useful energy and wasted energy.
h. Know and be able to apply: $\text{percentage efficiency} = \frac{\text{useful output}}{\text{total input}} \times 100$

Know and be able to apply: work = force × distance moved (in direction of force)

- When a force acting on an object causes the object to move through a distance, energy is transferred to or from the object by the force. This transfer of energy is called work.
- This is only true if the distance moved (or a component of the distance moved) by the object is in the same line as the force. If the distance (or distance component) is in the same direction as the force, then energy is transferred to the object. If the distance (or distance component) is in the opposite direction to the force, then energy is transferred from the object.
- The numerical quantity of work is defined as work = force × distance moved (in the direction of the force). This leads to the equation W = Fs, where W is the work done by the force F on the object when it moves distance component s along the line of action of the force.
- Although force is a vector quantity, so is the distance component (as it is effectively a displacement). When the two are multiplied together, the result is the scalar quantity of work. Whilst work can have a positive or negative sign (depending on whether energy is removed from or supplied to the object), it has no directional component.
- The units of work are the units of force × the units of distance. For forces in N and distances in m, the units of work are Nm. However, this unit has a special name – the joule (J), and this equation effectively defines the standard unit of energy as that amount of work done when a force of 1N moves an object through a distance of 1m.
- When working with data given in other units, it is important to make sure that compatible units are used before using the formula W = Fs. In general, this means converting energies to J, forces to N, and distances to m.

When working with distances that are at an angle θ to the force, the component of distance parallel to the force is found by introducing a cos θ factor.

For example, consider a horizontal external force F pushing an object a distance d up and along a slope that is inclined at angle θ to the horizontal:



Distance component moved parallel to force = $d \cos\theta$.

∴ Work done by force F = force × distance moved in direction of force = $Fd \cos\theta$.

Understand work done as a transfer of energy

- Work is the name given to the energy imparted to (or removed from) an object by the action of an external force.
- When the external force does work on the object, the energy of the object increases by an amount of energy equal to the work done.
- When the object does work on its surroundings, the energy of the object decreases by an amount of energy equal to the work done.
- Newton's third law tells us that if body A exerts a force on body B, then body B exerts an equal and opposite force on body A. During the action of the force, both objects move the same distance in the same direction as each other. It therefore follows that, if body A does work W on body B, then the energy of body A increases by W and the energy of body B decreases by W. (In this case, the direction of movement is in the direction of the force exerted by A on B.) We can see, therefore, that a direct consequence of Newton's third law is conservation of energy in a situation where work is being done.

Example

Consider two balls, A and B, moving towards each other and colliding. During the collision, they each exert a force of 60N on the other and whilst in contact, they move a distance of 5.0m in the direction of initial motion of A.

Force exerted by A on B is in the same direction as distance moved.

Therefore, work done by A on B = force × distance = 60 × 5 = 300J.

Force exerted by B on A is in the opposite direction to distance travelled.

Therefore, work done by B on A = force × distance = 60 × (-5) = -300J.

We can see that there is no net gain or loss of energy of the system. We simply have 300J of energy being transferred from A to B via the mechanism of work (energy transfer through forces).

Energy of A has increased by 300J, and energy of B has decreased by 300J.

Note

The form that energy imparted to a system (or energy removed from a system) takes depends on what the effect of the acting force is.

For example, if an object is being moved upwards at constant speed, there is no change in kinetic energy and the energy imparted takes the form of gravitational potential energy.

If an object is being accelerated along a horizontal surface, at constant height, then there is no change in potential energy and the energy imparted takes the form of kinetic.

In more complex situations, energy imparted to a system through work can take the form of several different types of energy.

Know and be able to apply: gravitational potential energy = mgh, where h is the difference in height of the object

- An object of mass m raised through vertical height h gains gravitational potential energy given by potential energy = mgh.
- Care must be taken, when using this formula, to ensure that units are compatible. If mass is in kg, g is in Nkg^{-1} and h is in m, then potential energy will be in Joules (J).

Proof

Consider an object of mass m being raised at constant speed by an external vertical force F.

The speed of the object is not changing, which means that all of the work done by the external force becomes gravitational potential energy.

Since there is no acceleration, the resultant force on the object must be zero.

Hence the external force F is equal (and opposite) to the weight of the object mg.

If the object is raised through vertical height h (which is in the same direction as the external force), then work done by external force = force × distance moved = mg × h = mgh.

The work done becomes potential energy, and so potential energy = mgh.

Arbitrary zero

When solving problems involving potential energy, there is no definitive point at which potential energy is taken to be zero. Where we set zero potential energy is arbitrary and can vary from problem to problem.

What matters is that the difference in energy between one height and another is always the same (and given by mgh).

For example, a book of mass 0.20kg resting on a table 1.2m high has 2.4J of potential energy more when on the table than when on the floor below.

If the arbitrary zero of potential energy is taken at floor level, then the book has 0J of potential energy on the floor, and

2.4J of potential energy on the table.

If the arbitrary zero of potential energy is taken at table level, then the book has 0J of potential energy on the table and -2.4J of potential energy when on the floor.

Where the zero is set does not matter – the differences in energy between the various heights will always be the same.

A trunk of mass 500 kg is pulled 4.0 m up a sloping ramp and loaded onto a truck. The force needed to pull the trunk up the ramp is 2000 N and the trunk rises 1.2 m.

a) How much work was done against friction forces as the trunk was loaded onto the truck? b) What was the average friction force?

The work done on the trunk by the force pulling it is given by force × distance = 2000 × 4.0 = 8000 J

The increase in gravitational potential energy of the trunk is mgh = 500 × 10 × 1.2 = 6000 J

The work done against friction is therefore 8000 − 6000 = 2000 J

The work done against friction is equal to the friction force (F) multiplied by the distance the trunk moves (4.0 m). Therefore 4.0 × F = 2000 and F = 500 N

Know and be able to apply: kinetic energy = $\frac{1}{2}$ mv^2

- An object of mass m moving at speed v has kinetic energy given by kinetic energy = $\frac{1}{2}mv^2$.
- Care must be taken, when using this formula, to ensure that units are compatible. If mass is in kg and speed is in ms^{-1}, then kinetic energy will be in joules (J).

Proof

Consider an object of mass m being accelerated from rest to speed v along a smooth horizontal surface by an external vertical force F.

The height of the object is not changing, and there are no resistive forces, which means that all of the work done by the external force becomes kinetic energy.

Since F is the only external force acting, acceleration of object = force / mass = F/m.

From the equation of uniform motion $v^2 = u^2 + 2as$, and since the object is accelerating from rest, we can deduce that the distance travelled by the mass whilst accelerating is given by distance = $v^2/2a$.

But a = F/m.

Therefore, distance travelled = $mv^2/2F$.

Therefore, work done by external force in accelerating the mass = force × distance = F × ($mv^2/2F$) = $\frac{1}{2}mv^2$.

This work all becomes kinetic energy.

Hence, the kinetic object of a mass m travelling at speed v is given by kinetic energy = $\frac{1}{2}mv^2$.

Know and be able to apply:

$$\text{power} = \frac{\text{energy transfer}}{\text{time}}$$

- When energy is transferred from one system to another, the rate at which the energy is transferred is called power.
- Power is defined as the rate of transfer of energy or the rate of doing work. This leads to the equations P = W/t and P = E/t, where P = power, W = work done, E = energy transferred and t = time.
- The units of power are the units of energy (or work) / units of time. For energies in J and times in s, the unit of power is Js^{-1}. The joule per second is given a special name, the watt (W). So energy in joules and time in seconds gives power in watts.
- When using the equation, it is important to ensure that the units of the quantities used are consistent. This generally means that all times have to be converted to seconds (unless unusual units like J min^{-1} are used for power). Energy can be left in kJ, in which case power will be in kW.

The kilowatt-hour

An unusual unit is sometimes used in the context of electrical energy. This unit, kWh, is actually a unit of energy. It is the energy transferred in a time of 1 hour when the rate of transfer is 1 kW.

If we convert the time and power into standard units, we have a time of 3600 seconds and a power of 1000W. This means that the energy involved = power × time = 1000 × 3600 = 3 600 000 J.

Thus an energy of 1 kWh is equivalent to 3 600 000 J.

A car of mass 1200 kg accelerates from rest to 20 ms^{-1} in 8.0 s. What is the useful output power of the car's engine?

The car's kinetic energy increases from zero to $\frac{1}{2} mv^2 = \frac{1}{2} \times 1200 \times 20^2 = 240\,000$ J in 8.0 s.

Power is given by $P = E/t = 240\,000 / 8.0 = 30\,000$ W.

The total power generated by the engine will be much higher than this because it is not 100% efficient.

Know and be able to use in calculations the law of conservation of energy

- The law of conservation of energy states that energy is never created or destroyed; it is merely transferred from one form into another.
- The different forms of energy can be divided into active forms and stored forms of energy. Stored energy (energy that is stored in some form for future use) is known as potential energy.
- Active forms of energy are:
    - electrical energy (transferred by a current in a circuit)
    - heat (thermal energy)
    - light
    - kinetic energy (energy of an object that is moving) - sound.

- Stored forms of energy are:
    - chemical potential energy (stored in a battery or cell)
    - gravitational potential energy (stored due to height) - strain potential (energy stored in a stretched spring).

- When energy is supplied to a system by the doing of work, the same amount of energy is removed from the system doing the work.
- The law of conservation of energy can be applied numerically to a variety of different situations using the relevant formulae.
- Analysing mechanical situations using the concepts of energy often makes for a simpler solution to the problem than using forces and the equations of uniform motion. Indeed, in cases where the motion is not uniform, where the equations of uniform motion are not applicable, energy methods can still be used.

An object of mass 1.2kg is attached to the lower end of an initially unstretched spring. The upper end of the spring is fixed. The object is then released so that it oscillates. The lowest point of the oscillation is a distance of 0.20m below the highest point (where the object is released).

(The spring obeys Hooke's law throughout. Gravitational field strength = 10Nkg^{-1}.) What is the spring constant of the spring?

At the top of the oscillations, the energy of the oscillations is all gravitational potential. (No strain energy because the spring is unstretched, and no kinetic energy because the object is stationary.)

At the bottom of the oscillations, there is again no kinetic energy (because the object is momentarily stationary) and so all of the gravitational potential energy lost has become strain energy in the stretched spring.

Loss of gravitational potential energy = mgh = 1.2 × 10 × 0.20 = 2.4J.

Gain in strain energy = $\frac{1}{2}kx^2 = \frac{1}{2}k \times 0.20^2 = 0.020k$.

∴ By conservation of energy, 0.020k = 2.4J. ∴ k = spring constant = 2.4 / 0.020 = 120Nm^{-1}.

Understand the concepts of useful energy and wasted energy

- In any device that is designed to convert energy from one form into another, some energy transfer is always into forms of energy that are not the purpose of the device.
- The ideas of useful energy and wasted energy are linked to the purpose of the device.
- Useful energy is energy that is converted into the form for which the device is intended.
- Wasted energy is energy that is converted into forms of energy other than that for which the device is intended.
- These technical terms are not intended to convey a value judgement on whether a particular form of energy is absolutely useful or not. Nor is it intended to imply that there is no use for wasted energy.
- Any given form of energy may be useful in the context of one application and wasted in the context of another, as illustrated by some of the examples below.

Example 1 – filament lamp

The purpose of a filament lamp is to convert electrical energy into light. A filament lamp also converts a lot of the electrical energy supplied into heat. However, producing heat is not the purpose of a filament lamp, and so the light produced is useful whilst the heat produced is wasted energy.

This does not mean that the heat produced necessarily goes to waste – it may be that the warming effect of having the light on may be found to be useful by a consumer. But in carrying out technical analysis of the performance of the lamp, the heat produced is considered to be technically wasted energy.

Example 2 – electric heater

The purpose of an electric heater is to convert electrical energy into heat. The element also glows orange, which means that some energy is being converted into light. However, producing light is not the purpose of a heater, and so the heat produced is useful whilst the light produced is wasted.

We can see from these two examples that there is no subjective value judgement being placed on whether or not producing heat or light is worthwhile. The distinctions of ‘useful’ and ‘wasted’ are purely technical distinctions based on the purpose of the energy transfer device.

Know and be able to apply:

$$\text{percentage efficiency} = \frac{\text{useful output}}{\text{total input}} \times 100$$

- For any device that is designed to convert energy from one form into another, the total input energy must always be equal to the total output energy. This is a direct consequence of the law of conservation of energy.
- The total output energy is always made up from the useful output energy and the wasted output energy. Total output energy = useful output energy + wasted output energy.
- The proportion of the input energy that is converted into the form of energy that is intended (i.e. the useful output energy) is known as the efficiency of the device.
- Efficiency is therefore equal to the ratio (useful output energy / total input energy).
- Efficiency can either be left as a fraction, or, more often, can be multiplied by 100 to express it as a percentage.
- A device with an efficiency of 100% would be one that converts all of its input energy into the form of output energy intended. However, in practice, 100% efficiency is never achievable.
- Whilst expressed above as the ratio of energies, efficiency can equally well be treated as the ratio of powers:

% efficiency = (useful output energy / total input energy) × 100

% efficiency = (useful output power / total input power) × 100

- Efficiency is dimensionless – it is the ratio between two powers or two energies. The units used for the powers or the energies must be the same as each other. But, apart from the % (if expressed as a percentage), efficiency has no units.

A heater that operates from a 240V mains supply draws an operating current of 5.0A. The heater produces an output of thermal energy at a rate of 960W.

What is the efficiency of the heater?

Total input power = V I = 240 × 5.0 = 1200W.

Useful output power = 960W.

These are already in compatible units, so no unit conversions to deal with.

Efficiency = useful output power / total input power = 960 / 1200 = 0.80.

This could be left as a fraction, 0.80, or multiplied by 100 to give efficiency = 80%.

P4. Thermal physics

P4.1

Conduction:

a. Know and understand thermal conductors and insulators, with examples.
b. Know and be able to apply factors affecting rate of conduction.

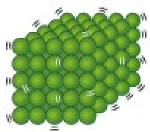
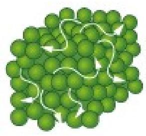
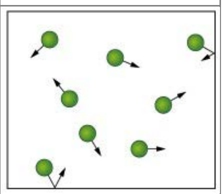
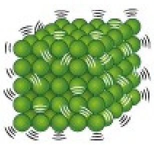
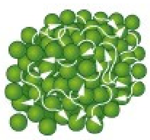
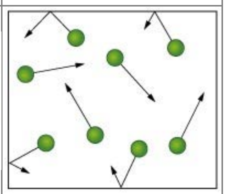
Know and understand thermal conductors and insulators, with examples

Thermal energy

Solids, liquids and gases are all made of microscopic particles – atoms or molecules (or ions). These particles are in motion: in solids the particles vibrate, while in fluids (liquids and gases) the particles move from place to place.

Thermal energy is the energy a substance has because of the motions of its microscopic particles. Thermal energy is not the same as temperature (a measure of 'hotness') – but when the thermal energy of a substance increases, its temperature usually increases too. The higher the temperature of a substance, the faster its particles move, on average.

The table illustrates this for the three states of matter.

	Solid	**Liquid**	**Gas**
Colder			
Hotter			

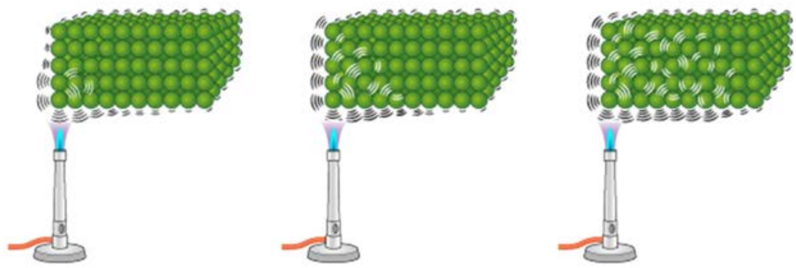
Conduction

When thermal energy is being transferred from one place to another, it is often called heat. Heat transfers from regions at higher temperature to regions at lower temperature.

Conduction, or thermal conduction, is the transfer of heat from one place to another through the passing on of kinetic energy between the microscopic particles of a substance. This happens mainly within solids and liquids, where the particles are in close proximity. Conduction can also transfer heat from a substance in one state to a substance in another state (including the gaseous state), if the two are in contact.

The diagram illustrates thermal conduction in a solid. The particles in the hotter region vibrate more energetically.

Over time, some of this energy is passed along to neighbouring particles, so that they also vibrate more energetically. This process continues through the solid, so that the temperature rises even in the region farthest from the heat source.



In liquids, energy is transferred via collisions between the moving particles.

Thermal conductors and insulators

Some materials are better conductors than others. ‘Good conductor’ means a fast conductor, through which heat transfers by conduction relatively quickly. Examples are metals (in both the solid and liquid states), with copper, silver and gold being particularly good conductors.

Poor conductors are called insulators. ‘Insulator’ means a slow conductor, through which heat transfers by conduction relatively slowly. Examples are plastics, wood, gases (for example, air), and materials that contain a lot of trapped air, such as foam and fibreglass.

Heat cannot transfer through a vacuum by conduction, since a vacuum does not contain particles.

Conduction in different states of matter

Gases are typically poorer conductors than liquids and solids. This is because the particles in a gas are far apart relative to their size. Collisions are not frequent enough to transfer kinetic energy between particles as quickly as in liquids and solids.

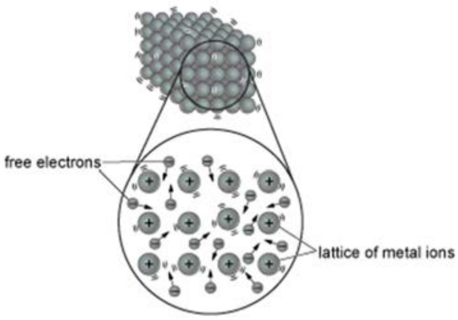
Liquids are typically poorer conductors than solids, because in a liquid the particles aren't held tightly together, and it takes longer for kinetic energy to be transferred between particles.

Conduction in metals

In metals in the solid and liquid states, there are free electrons, which can move through the lattice of metal ions.

When part of a metal becomes hotter, not only the ions but also the free electrons gain kinetic energy. Like the atoms in a non-metal, ions can only pass on kinetic energy relatively slowly, from one ion to its immediate neighbours and so on. However, free electrons can transfer energy much faster, by moving through the lattice and colliding with ions and with each other. This is why metals are particularly good thermal conductors.

Ions and free electrons in a metal



Uses of thermal conductors and insulators

For many applications, particular materials are chosen because of their thermal conduction properties. For example, good thermal conductors are used to make the parts of saucepans and frying pans that hold food, and the part of an iron that makes contact with clothing. Thermal insulators are used to make blankets, home insulation (to slow heat loss from the home in cold weather), and the handles of saucepans and irons.

A double-glazed window pane – a window pane made of two layers of glass with a layer of trapped air in between – is a better insulator than a single pane of glass. This is mainly because of the insulating effect of the layer of air.

Write these substances in order of thermal conductivity, from the best insulator to the best conductor.

liquid water

water vapour

solid aluminium

ice (frozen water)

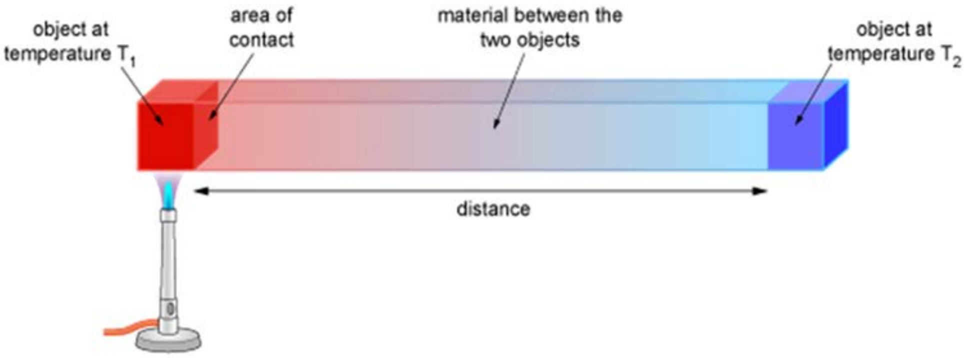
Explain why liquid silver is a better thermal conductor than solid rubber.

Considering the two solids, aluminium is a metal and therefore a good thermal conductor, while ice is a non-metal and therefore a poor thermal conductor. However, ice is a better conductor than liquid water which, in turn, is a better conductor than water vapour. This is because, in general, a substance is a better conductor in the solid state than in the liquid state, and a better conductor in the liquid state than in the gaseous state. The order is water vapour, liquid water, ice, solid aluminium.

Silver is a metal and hence has free electrons which can move throughout the silver in both its solid and liquid states. The free electrons transfer energy rapidly through the material. Rubber has no free electrons and so it conducts heat much more slowly.

Know and be able to apply factors affecting rate of conduction

The diagram shows two objects at temperatures T_1 and T_2, connected by a length of material with a uniform cross section.



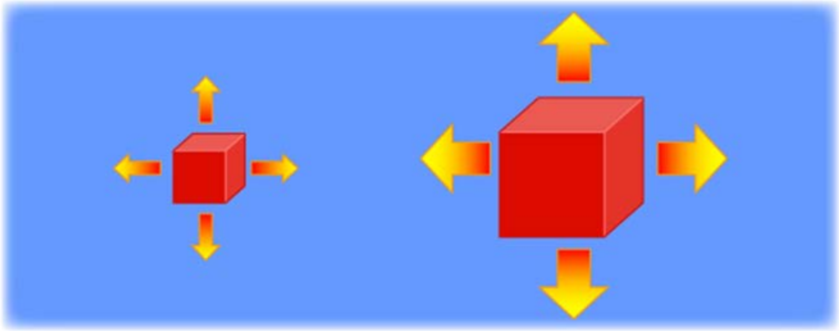
Heat conducts from regions of higher temperature to regions of lower temperature. In the diagram the temperature of the object on the left, T_1, is higher than the temperature of the object on the right, T_2. So heat transfers from left to right, through the material that is between the two objects.

The rate of heat transfer by conduction through the material depends on several factors, which are summarised in the table.

Factor affecting rate of heat transfer by conduction	Rate of heat transfer by conduction is higher for
temperature difference between the two objects ($T_1 - T_2$)	higher temperature difference
nature of the substance between the two objects	better thermal conductor
distance between the two objects	shorter distance
area of object surfaces in contact with connecting material	larger area

In the diagram above, only one face of each object is in contact with the connecting material. Unless the other faces of the two objects are surrounded by vacuum, they must be in contact with some substance (for example, air). There will also be conduction between each object and this substance (unless there is no temperature difference).

For an object that is surrounded by another substance which is at a different temperature, the area of contact is the total surface area of the object. Consider the two objects in the diagram below. They are both at the same temperature, which is higher than the temperature of the substance that surrounds them. But the object on the right has a larger surface area. This allows collisions to occur more frequently between its surface particles and particles of the surrounding substance – transferring energy at a higher rate. So the object on the right loses heat by conduction to the surroundings at a higher rate than the object on the left.



The rate of conduction to or from an object can be slowed by surrounding the object with a layer of insulating material.

For each pair of changes to a double-glazed house window described below, state and explain whether the changes could be made without changing the rate of thermal conduction through the window.

Increasing the thickness of the glass panes and decreasing the area of the window.

Replacing the air between the panes with a gas that is a better insulator and increasing the area of the window.

Decreasing the thickness of the glass panes and replacing the air between the panes with a vacuum.

No. Increasing the thickness of the glass decreases the rate of conduction, and so does decreasing the area of the window. This combination of changes must decrease the rate of conduction.

Yes. Replacing the air between the panes with a better insulator decreases the rate of conduction, while increasing the area of the window increases the rate of conduction. It is possible for this combination of changes to leave the rate of conduction unchanged.

No. Conduction is not possible through a vacuum, so replacing the air with a vacuum must reduce the rate of conduction through the window to zero.

P4.2

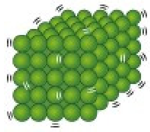
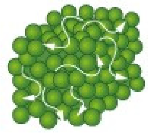
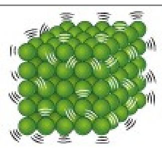
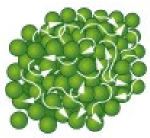
Convection:

a. Understand and be able to apply the effect of temperature on density of fluid.
b. Understand and be able to apply fluid flow caused by differences in density.

Understand and be able to apply the effect of temperature on density of fluid

When the temperature of a fluid increases, the average speed of its microscopic particles increases. The particles collide with each other more frequently and with greater force. If the fluid is not confined to a rigid container, its particles move further apart, on average.

So a fluid expands when its temperature increases – not because the particles themselves expand (which they do not), but because their average separation increases.

	Solid	**Liquid**
Colder		
Hotter		

When the particle separation increases, the density of the fluid decreases, because there are now fewer particles (and hence less mass) per unit volume.

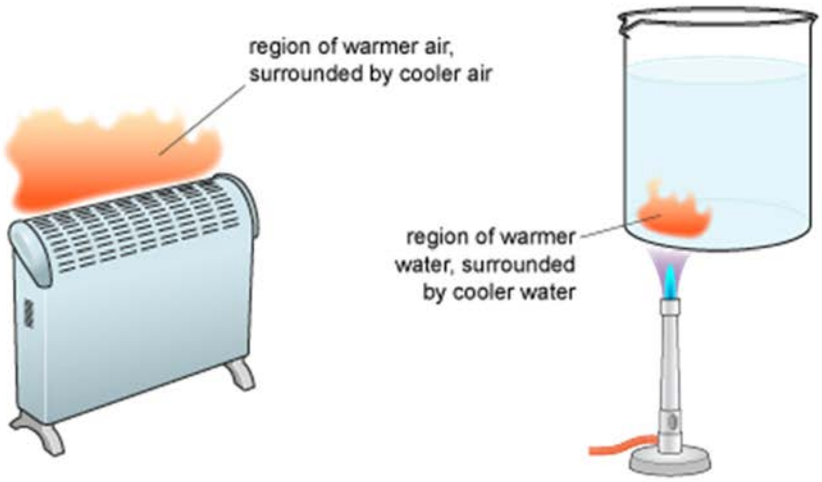
A hot air balloon floats because the air inside it is heated by a flame so that it is hotter – and therefore less dense – than the surrounding air. Even allowing for the additional mass of the balloon itself and the basket it carries, this makes the balloon's overall density lower than the density of the surrounding air. So the balloon floats upwards. It can be made to fall by switching off the flame and letting the air inside it cool.

Flame heating the air inside a balloon

Understand and be able to apply fluid flow caused by differences in density

Convection is often summarised as 'hotter fluid rises and colder fluid moves in to take its place'. This process is described and explained in more detail below.

Convection occurs when a region within a fluid is heated. This can happen to the air nearest to an electric heater in a room, for example; or to the water, in a beaker, that is nearest to a Bunsen burner. In both cases, heat transfers from the heat source to the fluid mainly by conduction.

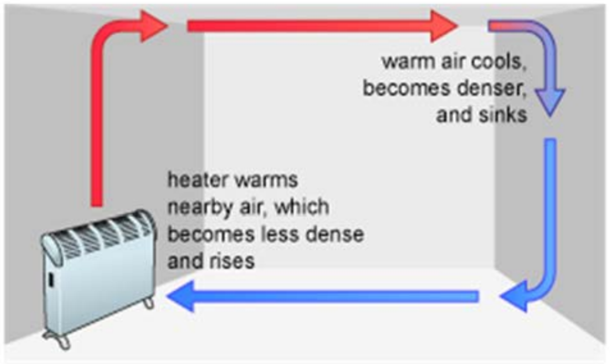


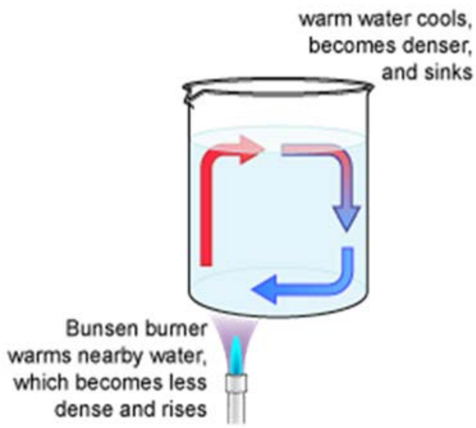
The warmer fluid has lower density than the surrounding cooler fluid. So the warmer fluid moves upwards (just as a hot air balloon floats upwards because it is less dense than the surrounding air).

The upwards movement of warmer fluid pushes cooler fluid out of the way, and the result is that cooler fluid moves around to take the place where the warmer fluid was.

As the warmer fluid rises, it gradually cools (by conduction of heat to the cooler fluid around it), becomes less dense, and tends to sink.

If the heat source (in the examples above, the radiator or the Bunsen burner) continues to supply heat, a cycle of moving fluid, called a convection current, can develop within the fluid. Fluid nearest the heat source is continuously warmed, rises, and cools, while cooler fluid continuously moves to take its place next to the heat source.





If a heat source warms fluid at the top of a container, a convection current is not set up, because the warmer, less dense, fluid stays at the top.

Convection transfers heat through a fluid much more quickly than conduction. Both can happen at the same time within a fluid.

Comparing conduction and convection

The table shows similarities and differences between conduction and convection.

Conduction	Convection
Requires the presence of particles.	Requires the presence of particles.
Can occur in fluids.	Can occur in fluids.
Can occur in solids.	Cannot occur in solids.
Heat is transferred by microscopic motions of individual particles – kinetic energy is passed from one particle to the next.	Heat is transferred by macroscopic (bulk) motion of large numbers of particles – the particles themselves move around, carrying their kinetic energy with them.

Using convection

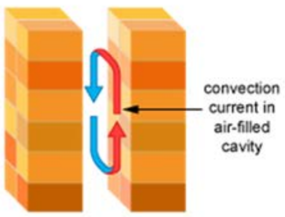
Household appliances designed to affect or control the temperature in a room are usually positioned to make the best use of convection. For example, a room heater is usually placed near the floor, to allow the warmed air to rise and circulate around the room. An air conditioning unit, designed to cool the air in a room, is usually placed near the ceiling, allowing the cooled air to sink and circulate.

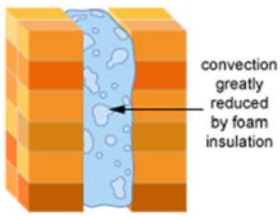
Reducing heat transfer by reducing convection

Many houses have two layers of outer wall, with a cavity (empty space) between them. These cavities were originally designed to be filled with air, to reduce heat loss from the house in cold weather by conduction (since air is a good thermal insulator). However, when heat from the house transfers through the inner wall to the air in the cavity, convection makes this air circulate. This speeds up the transfer of heat to the outer wall and then out to the surroundings.

If the cavity is filled with insulating material – such as foam insulation, which contains many very small pockets of trapped air – then convection is greatly reduced (since it can only occur within each tiny air pocket). This reduces the rate of heat loss from the house.

Cavity wall insulation





Some refrigerators have a small freezer compartment inside, and inside this are ‘cooling coils’. This is where cooling of air takes place within the appliance.

What is the best position for the freezer compartment – at the top, middle or bottom of the refrigerator? Explain your answer.

Air cooled by the freezer is denser than the air below, and so the cooler air sinks. Warmer air (which may have been warmed by room temperature food placed in the fridge, or by conduction of heat through the fridge walls) rises to take its place and is cooled in turn by the cooling coils. This causes a convection current which circulates cooled air around the inside of the appliance. The best position is at the top.

During a sunny day, the surface of the land reaches a higher temperature than the surface of the sea. This causes a breeze to blow.

State and explain the direction of this breeze.

Explain why there is a breeze moving in the opposite direction at a higher altitude.

The air above the land and sea is warmed by conduction, but the air above the land becomes warmer because the land's temperature is higher. This causes a convection current in which air above the land rises and is replaced by cooler air moving across from over the sea. So a breeze blows from the sea towards the land.

This is the uppermost part of the convection current, which flows horizontally from above the land to above the sea.

P4.3

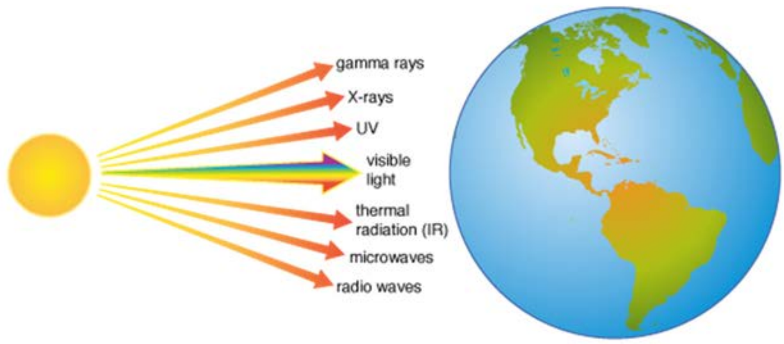
Thermal radiation:

a. Understand thermal radiation as electromagnetic waves in the infrared region.
b. Know and be able to apply absorption and emission of radiation.
c. Know and be able to apply factors affecting rate of absorption and emission of thermal radiation.

Understand thermal radiation as electromagnetic waves in the infrared region

Thermal radiation, also called infrared (IR) radiation, is a type of wave that is one of the parts of the electromagnetic spectrum. Like all waves of the electromagnetic spectrum, it travels at the speed of light and does not need a medium in which to travel.

So unlike conduction and convection, thermal radiation can transfer energy through a vacuum. This is how heat is transferred from the Sun to the Earth.



Radiation therefore involves transfer of heat directly from one object to another, without heating up (or passing heat through) a medium connecting the two objects. An example of this is a radiant heater, which heats objects nearby without heating up the air in between. The thermal energy is transferred by infrared radiation.

Infrared radiation is not visible to the human eye, but it can be sensed by the warmth it produces on hitting the skin.

Which of the following statements apply to thermal radiation?

1) It can travel through a vacuum.

2) It is a type of ionising radiation.

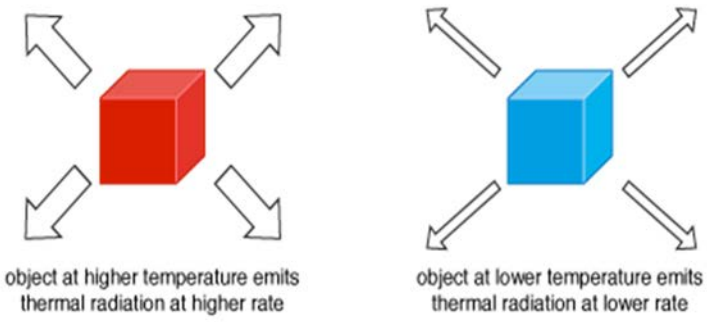
3) It travels slower than light.

Thermal radiation (or infrared radiation) is a type of electromagnetic radiation. All types of electromagnetic radiation can travel without a medium (in a vacuum) and all travel at the speed of light. Thermal radiation is not emitted during the decay of unstable nuclei, so it is not a type of ionising radiation. Only statement 1) applies.

Know and be able to apply absorption and emission of radiation

Emission of thermal radiation

Any object or substance with a temperature above absolute zero (the lowest temperature that is possible) emits thermal radiation. The higher the temperature of an object, the higher the rate at which it emits thermal radiation (where 'rate' means energy per second, or power).

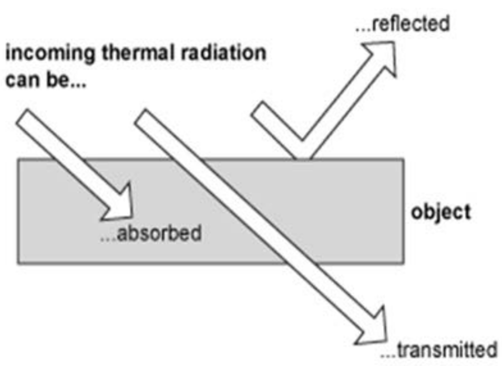


When an object emits thermal radiation, thermal energy of the object is transferred to energy of the radiation (as long as the object is not changing state). So if an object were to emit thermal radiation without any other energy transfers occurring, its temperature would decrease.

Absorption of thermal radiation

When thermal radiation hits an object or substance, some of the radiation may be absorbed. Any that is not absorbed is either reflected (bounces off) or transmitted (passes through).

The three possible outcomes when thermal radiation hits an object



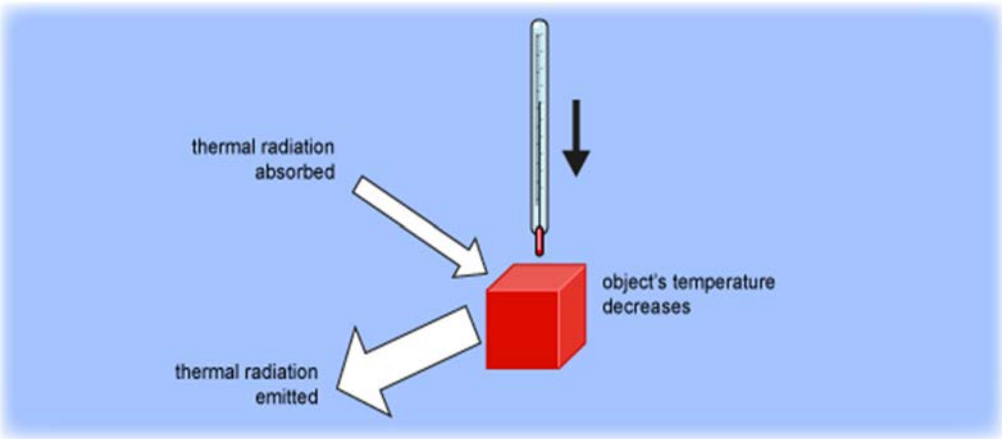
When an object absorbs thermal radiation, energy of the radiation is transferred to thermal energy of the object (as long as the object is not changing state). So if an object were to absorb thermal radiation without any other energy transfers occurring, its temperature would increase.

Combined effect of emission and absorption

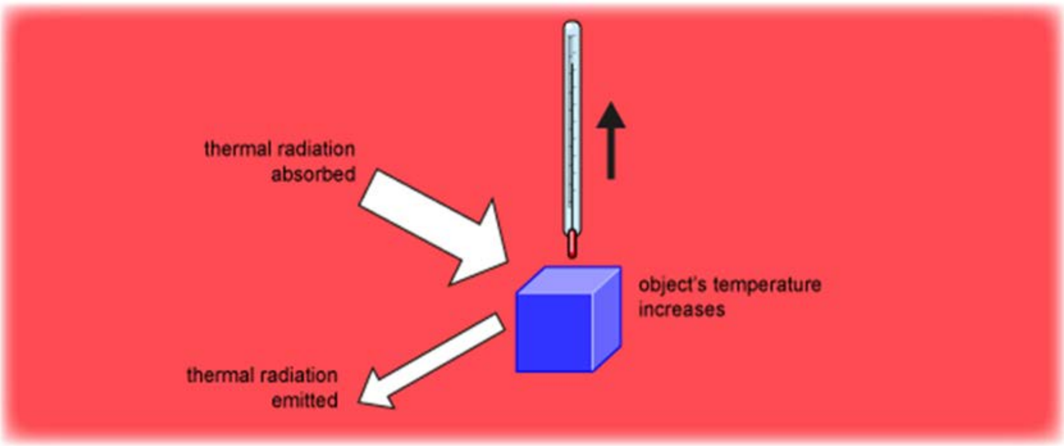
All objects both emit and absorb thermal radiation all the time.

If an object is at a higher temperature than its surroundings, then it emits thermal radiation at a higher rate than it absorbs it. There is a net loss of thermal energy from the object and its temperature decreases (in the absence of other energy transfers).

Emission and absorption



If an object is at a lower temperature than its surroundings, then it absorbs thermal radiation at a higher rate than it emits it. There is a net gain in thermal energy of the object and its temperature increases (in the absence of any other energy transfers).



Air absorbs thermal radiation only weakly – hence thermal radiation from the Sun is able to pass through the atmosphere and reach the Earth's surface.

Thermal radiation is incident on a window at a rate of 1000 W. The window reflects this radiation at a rate of 100 W, and at the same time it transmits radiation at a rate of 750 W. The window also emits thermal radiation at a rate of 200 W.

Calculate the net rate at which the window's thermal energy changes as a result of the processes described, and state whether it is a net gain or loss of energy by the window.

Of the 1000 W of incident thermal radiation, 100 W is reflected and 750 W passes through the window, so the rest, 1000 − 100 − 750 = 150 W, is absorbed by the window.

The window also loses thermal radiation by emission at a rate of 200 W, so the net change in thermal energy is 150 − 200 = -50 W, or a loss of energy at a rate of 50 W.

Know and be able to apply factors affecting rate of absorption and emission of thermal radiation

The higher the temperature of an object, the higher the rate at which it emits thermal radiation.

The rates of emission and absorption of thermal radiation by an object both depend on several other factors, summarised in the table.

Factor affecting rates of emission and absorption	Rates of both absorption and emission are <u>lower</u> for	Rates of both absorption and emission are <u>higher</u> for
texture of object's surface	shiny surfaces	matt (dull) surfaces
surface area of object	smaller surface area	larger surface area

A good absorber of thermal radiation is a good emitter, and a poor absorber is a poor emitter (and a good reflector).

Practical applications

A survival blanket may be wrapped around someone who is at risk of becoming dangerously cold. The blanket is shiny so that it is a poor absorber and emitter, and a good reflector, of thermal radiation. It is better than, for example, a dull matt blanket at reflecting thermal radiation from the person back towards them, and it is poorer at emitting thermal radiation to the surroundings.

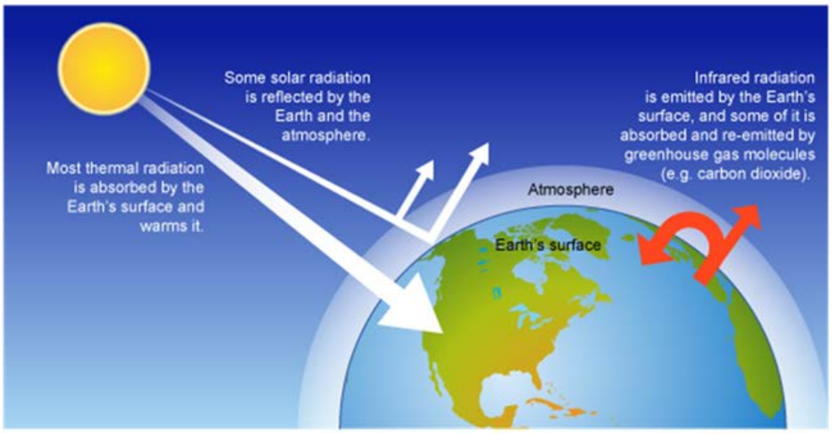
A person wrapped in a survival blanket

A solar water heater uses thermal radiation from the Sun to heat water. The water runs through pipes, and these are painted matt black because a dull, matt surface is a relatively good absorber of thermal radiation. (It is also a good emitter, but since thermal radiation arrives from the Sun at a high rate, the rate of absorption is higher than the rate of emission from the warm pipes.)

Solar water heater

Effect of thermal radiation on the Earth

The temperature of the Earth is determined mainly by the rates at which the Earth's surface and its atmosphere absorb and emit thermal radiation over time. Global warming, also called the greenhouse effect, is happening because heat is being transferred to the Earth and its atmosphere faster than it is being transferred away. The diagram gives a simplified overview of what is happening.



Photos: Science Photo Library

P4.4

Heat capacity:

a. Understand the effect of energy transferred to or from an object on its temperature.
b. Know and be able to apply:

$$\text{specific heat capacity} = \frac{\text{thermal energy}}{\text{mass} \times \text{temperature change}}$$

where temperature is measured in °C and specific heat capacity, c, is measured in J kg^{-1} °C^{-1}.

Understand the effect of energy transferred to or from an object on its temperature

Heat may be transferred to or from an object by conduction, convection or radiation. More than one of these processes can occur at the same time. The overall effect on the object's temperature depends on the net heat transferred: the difference between the energy transferred *to* the object and the energy transferred *from* the object.

When more heat is transferred to an object than from it (in other words, when there is a net transfer of heat into an object), the object's temperature increases (unless it is changing state). The microscopic particles of the object move more energetically.

When more heat is transferred from than to an object, the object's temperature decreases (unless it is changing state). Its particles move less energetically.

If there is a transfer of thermal energy to or from an object (and it is not changing state), the change in the temperature of the object depends on:

- the mass of the object the type of material it is made of;
- the transfer of thermal energy per unit change in temperature of an object is called the heat capacity of the object.

The unit of heat capacity is J °C^{-1}.

Know and be able to apply:

$$specific\ heat\ capacity = \frac{thermal\ energy}{mass \times temperature\ change}$$

where temperature is measured in °C and specific heat capacity, c , is measured in J kg^{-1} °C^{-1}

The specific heat capacity of a substance is the heat capacity (thermal energy transferred per unit temperature change) per unit mass of the substance – since 'specific' means 'per unit mass'.

The specific heat capacity of a substance can be calculated by measuring the temperature change of a sample of known mass when a measured thermal energy is transferred to or from the sample. The specific heat capacity is:

$$specific\ heat\ capacity = \frac{thermal\ energy}{mass \times temperature\ change}$$

where energy is measured in J, mass is measured in kg and temperature is measured in °C.

The unit of specific heat capacity is J kg^{-1} °C^{-1}.

Heat capacity is a property of a particular object or sample.

Specific heat capacity is a property of a type of substance.

The table shows some examples of specific heat capacities of different substances, written to two significant figures.

Substance	**Specific heat capacity / J kg−1 °C−1**
lead (solid)	130
copper (solid)	390
aluminium (solid)	900
water (solid)	2100
ethanol (liquid)	2400
water (liquid)	4200
hydrogen (gas)	14000

Calculate the specific heat capacity of magnesium if the temperature of a 3.0 kg sample of aluminium increases by 15°C when 45 000 J of thermal energy is transferred to the sample from the surroundings.

$$c = \frac{thermal\ energy}{mass \times temperature\ change}$$

$$= \frac{45000}{3 \times 15}$$

= 1000 J kg^{-1} $°C^{-1}$

Calculate the energy transferred to the surroundings when the temperature of an aluminium block of mass 2.0 kg decreases from 25.0°C to 20.0°C.

(The specific heat capacity of aluminium is 900 J kg^{-1} $°C^{-1}$.)

The temperature change is 25.0°C − 20.0°C = 5.0°C.

thermal energy = c × mass × temperature change

= 900 × 2.0 × 5.0

= 9000 J

P5. Matter

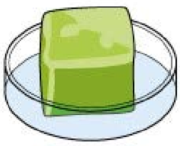
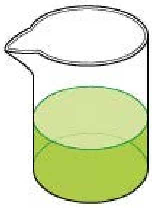
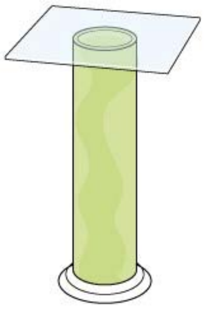
P5.1

States of matter:

a. Know the characteristic properties of solids, liquids and gases.
b. Know and be able to apply particle models of solids, liquids and gases.
c. Know and be able to explain properties of solids, liquids and gases in terms of particle motion and the forces and distances between the particles.

Know the characteristic properties of solids, liquids and gases

Each of the three states of matter – solid, liquid and gas – has characteristic properties as shown in the table.

	Solid 	**Liquid** 	**Gas** 
Fluidity	rigid	fluid (takes shape of bottom of container)	fluid (takes shape of whole container)
Compressibility	insignificant	insignificant	high (volume can decrease – or increase – significantly)
Typical density	high (around 10^3–10^4 kg m^{-3})	high (around 10^3 kg m^{-3})	low (around 1 kg m^{-3} at room temperature and atmospheric pressure)

A substance such as a sponge may seem to contradict the idea that solids are hardly compressible, since it can be greatly compressed. A sponge is made of a solid material with many pockets of air inside it. When the sponge is squeezed, air is pushed out of it, and this is why very significant compression occurs. At the same time, the solid 'skeleton' of the sponge hardly changes its volume.

Liquids and gases can be grouped together under the term 'fluids', so named because both states can flow.

Which state of matter best fits each description?

A substance in this state is difficult to compress and can flow.

A substance in this state occupies much more volume than it does in either of the other two states.

A particular substance in this state has a density of 19 000 kg m^{-3}.

Liquid. Both solids and liquids are hard to compress, but solids cannot flow – they are rigid.

Gas. Gases have much lower densities than liquids and solids, because the same amount of material takes up a much larger volume.

Solid. Solids typically have densities in the range 10^3 – 10^4 kg m^{-3}.

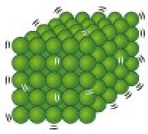
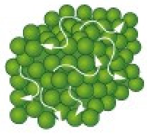
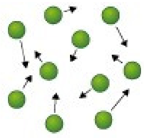
A density of 1.9×10^4 kg m^{-3} is even higher than this, so this is most likely to be a solid.

Know and be able to apply particle models of solids, liquids and gases

The particle model states that all matter is made of microscopic particles which are in constant motion.

The particle model can be used to explain the existence of the three states of matter: solid, liquid and gas. In the model, the particles are arranged differently in each state, and move in different ways. The spaces between particles are empty.

The table describes the forces between particles, the arrangement of particles, and the motion of particles in the three states, according to the particle model.

	Solid 	Liquid 	Gas 
Forces between particles	neighbouring particles are strongly attracted (or 'bonded') to each other	neighbouring particles are attracted to each other, but the bonds are weaker than in solids	no forces between particles (except during collisions with each other)
Arrangement of particles	very close together, in a fixed arrangement usually arranged in a regular pattern, or lattice	very close together (though usually not quite as close as in a solid) no regular pattern	far apart (on average) compared with the particle size no regular pattern
Motion of particles	each particle vibrates about a fixed position within the lattice particles do not change places with each other	particles move randomly from place to place within the liquid	particles move randomly and at high speed (relative to particles in a liquid) particles travel in straight lines between collisions (with objects or with each other)

In the particle model, the term 'particle' refers to the smallest unit of a substance, which may be made of atoms, molecules, or ions. The particle model does not distinguish between the types of particle; it predicts the same behaviours for the solid, liquid and gas states regardless of the type of substance.

According to the particle model, which state has each of the following microscopic characteristics?

a) Particles very close together (relative to their size) and moving randomly from place to place.

b) Vibrating particles.

c) No forces between particles most of the time.

Liquid. In both liquids and solids, the particles are very close together, but in a solid they do not move from place to place.

Solid. Particles in a solid carry out repeating motions, or vibrations, that are centred on fixed positions within the lattice.

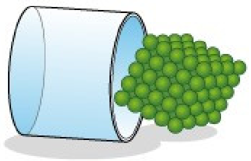
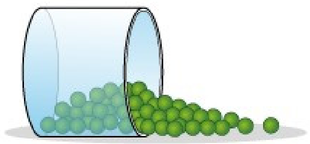
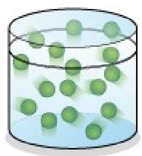
Gas. Particles in a gas do not exert forces on each other except during collisions.

Know and be able to explain properties of solids, liquids and gases in terms of particle motion and the forces and distances between the particles.

The particle model can explain macroscopic properties of the three states of matter in terms of the forces, arrangements and motions of the microscopic particles from which they are made.

/…. continued on next page

The table summarises explanations that can be deduced from the particle model.

State	Macroscopic property	Explanation in terms of particles
Solid	Has a fixed shape; cannot flow. 	The particles in a solid are held in a fixed arrangement (because of their strong attractions to each other).
	Compressibility is insignificant; has a fixed volume.	The particles in a solid are very close together.
Liquid	Variable shape; takes shape of bottom of container. 	The particles in a liquid are not held in a fixed arrangement, but there are attractions between them that are strong enough to hold the liquid together.
	Compressibility is insignificant; has a fixed volume.	The particles in a liquid are very close together.
	Can flow.	The particles in a liquid are not held in a fixed arrangement, so they are able to move past each other.
Gas	Variable shape; fills container. 	The particles in a gas are not attracted to each other and are in random motion, so they spread throughout the available space.
	Can be compressed; does not have a fixed volume.	The particles in a gas are far apart, so they can be pushed closer together.
	Low density compared with solids and liquids.	In gases there is a lot of space between the particles, hence a lower number of particles per unit volume than in liquids and gases. So gases have a lower mass per unit volume.

When describing macroscopic properties and microscopic particle behaviour, care should be taken not to confuse the two. For example, if a material is compressible (a macroscopic property), that does not mean its particles are compressible – rather, they can be moved closer together. Similarly, the density (a macroscopic property) of water vapour is lower than that of liquid water, but that does not mean that water particles are denser in the liquid state – they are simply closer together, so that there are more of them per unit volume.

The particle model can also be used to explain changes of state, pressure and temperature in gases, and the relationship between the pressure and volume of a gas.

Use the particle model to explain why:

Convection cannot occur in a solid.

A drop of food colouring in a glass of water gradually spreads out.

A gas exerts a force on any object it is in contact with.

Convection involves the movement of warmer, less dense material upwards while cooler, denser material moves to take its place. The particles in a solid are in a fixed arrangement and cannot change places with each other, so convection in a solid is not possible.

The particles of a liquid are in constant random motion. Collisions between liquid particles and particles of food colouring move the food colouring around within the liquid, causing it to spread out.

Particles in gases are in random motion, so they collide with objects. Each collision exerts a small force.

P5.2

Ideal gases:

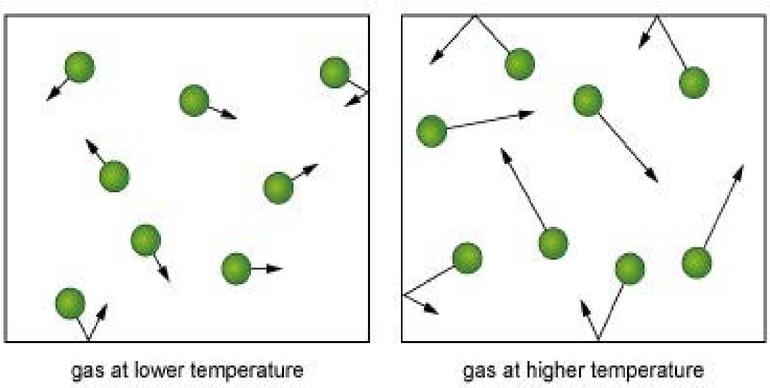
a. Be able to explain pressure and temperature in terms of the behaviour of particles.
b. Understand and be able to apply the effect of pressure (P) on gas volume (V) at constant temperature, i.e. PV = constant.

Be able to explain pressure and temperature in terms of the behaviour of particles

According to the particle model, a gas consists of identical particles which are in random motion and do not exert forces on each other (except during collisions). These assumptions, together with a small number of others (which include: the gas particles take up negligible volume; and they obey Newton's laws) make up a simple model of a gas called an ideal gas. Many gases, when under conditions close to room temperature and atmospheric pressure, behave very much like an ideal gas.

The temperature of an ideal gas is a measure of the 'hotness' of the gas. At a macroscopic level, temperature can be measured with an instrument such as a thermometer. The higher the temperature of a gas, the higher the average speed of its microscopic particles.

How gas changes with increased temperature



A gas exerts a force on any object that it is in contact with, because of the random motions of the gas particles. Particles in a gas collide with objects, and each collision exerts a tiny force. These add up to an average force per unit area, or pressure, exerted by the gas.

Temperature is a macroscopic property, so it would not be correct to describe individual particles of a gas as having a particular temperature, or being hot or cold. The gas as a whole has a temperature, while the behaviour of its particles can be described in terms of their average speed. When the temperature of a gas increases, the average speed of its particles increases.

Pressure is also a macroscopic property of a gas, so it would not be correct to describe individual particles as having a particular pressure. The gas as a whole has a pressure, which is due to the microscopic motions of its particles.

A sample of gas is held in a sealed container. The temperature of the gas is increased.

Describe and explain how this affects the pressure which the gas exerts on the walls of the container.

The pressure increases, because:

the particles are moving faster on average, so collisions with the container walls are more frequent

the particles are moving faster on average, so collisions with the container are more forceful on average.

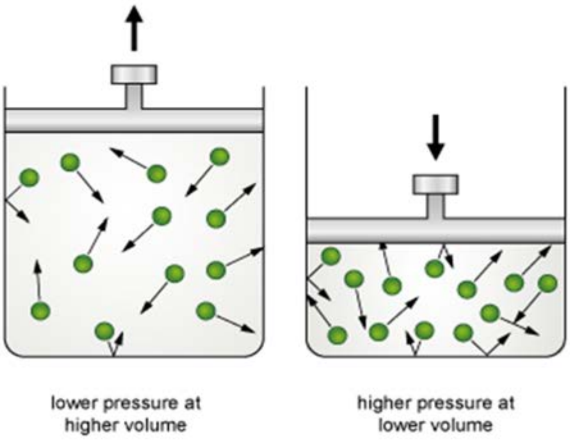
These two effects both increase the average force per unit area, or pressure, of the gas on the container walls.

Understand and be able to apply the effect of pressure (P) on gas volume (V) at constant temperature, i.e. PV = constant

Consider a sealed container of gas, with the gas kept at constant temperature. The volume of the gas is decreased (by pushing in a plunger, for example). The average speed of the gas particles is still the same as before, since the temperature is constant – but the gas particles now have less distance to travel between collisions with the container walls.

So the frequency of collisions between gas particles and the container increases, while the average force of each collision stays the same. As a result, the pressure exerted by the gas increases.

Volume and pressure



So if the volume of a gas is decreased (at constant temperature), the gas pressure increases. The converse is also true: if the volume of a gas is increased (at constant temperature), the gas pressure decreases.

The relationship between pressure, P, and volume, V, for a fixed amount of ideal gas at a constant temperature can be written:

PV = constant

(The value of the constant depends on the number of particles in the gas sample and the temperature of the sample.)

In practice, the temperature of a gas increases when its volume is decreased. However, if the gas returns to its original temperature – for example, by returning to thermal equilibrium with its surroundings – then the pressure and volume before and after the volume decrease will obey the relationship PV = constant.

If a gas has initial pressure P_1 and initial volume V_1, and its volume is changed to V_2 so that its pressure changes to a new value, P_2, then:

$P_1 V_1$ = constant, and

$P_2 V_2$ = constant, so $P_1 V_1 = P_2 V_2$

provided the container stays sealed so that no gas can get in or out, and the temperature is the same when P_2 and V_2 are measured as when P_1 and V_1 are measured.

Another way to express the relationship between pressure and volume for a fixed volume of gas is:

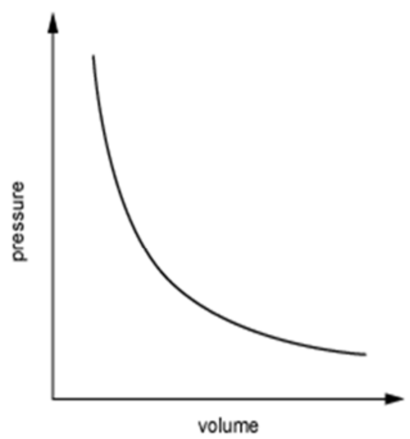
$$P = \frac{constant}{V}$$

or

$$P \propto \frac{1}{V}$$

or, pressure is inversely proportional to volume.

So the graph of P against V for a particular mass of ideal gas at a particular temperature is an inverse proportion graph, as shown.



The pressure of a sample of gas in a sealed container is 1.0×10^5 Pa and the volume is 60 cm^3. The container volume is reduced, without any gas escaping, to 40 cm^3. What is the pressure of the gas?

Before the change in volume,

$PV = 1.0 \times 10^5 \times 60$

$= 6.0 \times 10^6$ Pa cm^3

After the change in volume,

$PV = 6.0 \times 10^6$ Pa cm^3

$P = 6.0 \times 10^6$ Pa cm^3 / 40 cm^3

$= 1.5 \times 10^5$ Pa

A slightly quicker method is to avoid calculating the constant and simply write:

$P_1V_1 = P_2V_2$

$P_1 = \frac{P_2V_2}{V_1}$

$= 6.0 \times 10^6$ Pa cm^3 / 40 cm^3

$= 1.5 \times 10^5$ Pa

P5.3

State changes:

a. Understand the terms *melting point* and *boiling point*.
b. Know and understand the terms *latent heat of fusion* and *latent heat of vaporisation*.
c. Know and be able to apply specific latent heat calculations.

Understand the terms melting point and boiling point

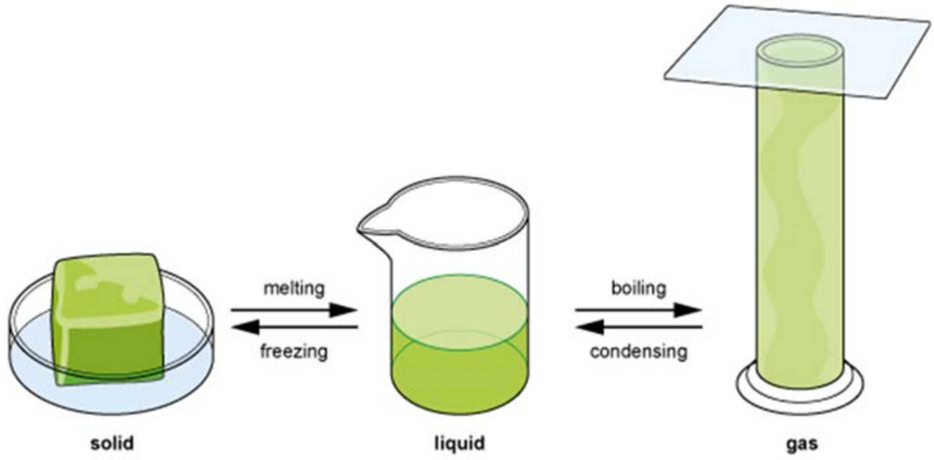
Nearly every pure substance has a melting point and a boiling point. (There are a few exceptions, such as carbon dioxide, which sublime – change directly between the solid and gas states.)

The melting point of a substance is the temperature at which it changes between the solid and liquid states. A solid turns to liquid (by melting) at temperatures above the melting point. A liquid turns to solid at temperatures below the melting point.

The boiling point of a substance is the temperature at which it changes between the liquid and gas states throughout the bulk of the sample. A liquid turns to gas (by boiling) at temperatures above the boiling point. A gas turns to liquid at temperatures below the boiling point.

The table summarises the state of a substance at different temperatures.

Temperature of substance	State of substance
below its melting point	solid
at its melting point	solid/liquid
between its melting point and its boiling point	liquid
at its boiling point	liquid/gas
above its boiling point	gas



The table shows some typical examples of melting and boiling points. These are for illustration purposes only, and values do **not** need to be learnt.

Substance	Melting point / °C	Boiling point / °C
water	0	100
ethanol (alcohol)	−117	78
copper	1085	2560
aluminium	660	2520
mercury	−39	357
nitrogen	−210	−196
hydrogen	−259	−253

An impure substance melts over a range of temperatures and boils over a range of temperatures, so it cannot be said to have a specific melting point or boiling point.

Write the state of each of the substances below at a temperature of 24°C .

Bromine, which has melting point −7°C and boiling point 59°C.

Fluorine, which has melting point −220°C and boiling point −188°C.

Iodine, which has melting point 114°C and boiling point 184°C.

Liquid. 24°C is above the melting point of bromine but below the boiling point.

Gas. 24°C is above the boiling point of fluorine.

Solid. 24°C is below the melting point of iodine.

Know and understand the terms latent heat of fusion and latent heat of vaporisation

When thermal energy is transferred to a body, its temperature rises, and when thermal energy is transferred from a body, its temperature falls – unless it is changing state.

While a sample of a pure substance is changing state either from solid to liquid or from liquid to gas, it absorbs thermal energy without increasing its temperature. The absorbed energy is needed to increase the separations between the particles. Only when the entire sample has changed state will further absorption of thermal energy cause its temperature to increase again.

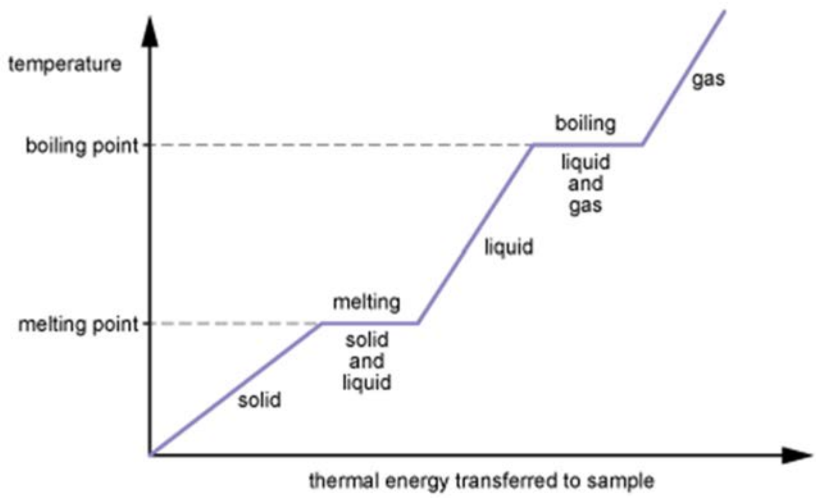
While a sample of a pure substance is changing state either from gas to liquid or from liquid to solid, it releases thermal energy without decreasing its temperature. As the attractions between the particles increase and their separations decrease, energy is transferred from the sample to thermal energy of its surroundings. Only when the entire sample has changed state will further release of thermal energy cause its temperature to drop.

If a sample of a pure substance is melting or freezing, both the solid and liquid parts of the sample remain at the melting point of the substance during the state change.

If a sample of a pure substance is boiling or condensing, both the liquid and gaseous parts of the sample remain at the boiling point of the substance during the state change.

The thermal energy transferred to or from a sample during a state change is called latent heat, and its quantity depends on the type of substance, the mass of the sample, and which type of state change occurs. Latent heat of fusion applies to melting (or freezing), while latent heat of vaporisation applies to boiling (or condensing).

What happens to a pure substance, initially in the solid state, which is heated steadily until it becomes a gas



The graph shows that the temperature rises when the substance is not changing state, and remains constant when the substance is changing state.

A graph for a cooling sample, initially in the gaseous state, looks similar, except that it is reversed so that the temperature starts high and decreases.

This graph shows temperature against thermal energy transferred, but it is also possible to plot temperature against time. If thermal energy is transferred to the sample at a constant rate, then the temperature–time graph is simply a scaled version of the temperature–energy graph.

A sample of an impure substance does not remain at a constant temperature when changing state. The latent heat is absorbed or released over a range of temperatures.

During each of the processes below, name the type of latent heat involved, and state whether it is absorbed or released by the substance during the process.

Ice melting

Gaseous propane changing to liquid propane.

Latent heat of fusion is absorbed. This is a change of state from solid to liquid. While the solid is melting, it absorbs thermal energy without changing its temperature. This is the latent heat of fusion.

Latent heat of vaporisation is released. This is a change of state from gas to liquid. While the gas is condensing, it releases thermal energy without changing its temperature. This is the latent heat of vaporisation.

Know and be able to apply specific latent heat calculations

The specific latent heat of a substance is the latent heat – or thermal energy transferred during a complete change of state – per unit mass of the substance, since 'specific' means 'per unit mass'.

So specific latent heat, L, is related to the amount of thermal energy transferred, E, and the mass m of the sample, according to the formula: $E = mL$

The unit of specific latent heat is J kg^{-1}.

If the state change is between solid and liquid, then L is the specific latent heat of fusion, and the energy E is the thermal energy transferred to the sample to change it from entirely solid at the melting point to entirely liquid at the melting point.

If the state change is between liquid and gas, then L is the specific latent heat of vaporisation, and the energy E is the thermal energy transferred to the sample to change it from entirely liquid at the boiling point to entirely gas at the boiling point.

Example latent heat calculations are shown below.

Calculate the energy transferred when 3.0 kg of ice at 0°C changes to liquid water at 0°C.

(The specific latent heat of fusion of water is 330 kJ kg^{-1}.)

$E = mL$

$= 3.0 \text{ kg} \times 330\,000 \text{ J kg}^{-1}$

$= 990\,000$ J, or 990 kJ

Calculate the latent heat of vaporisation of water if 6900 kJ of thermal energy is transferred to the surroundings when

3.0 kg of water vapour at 100°C changes to liquid water at 100°C.

$L = \frac{E}{m} = \frac{6900000}{3.0}$

$= 2\,300\,000 \text{ J kg}^{-1}$, or 2300 kJ kg^{-1}

If a sample is heated so that its temperature rises and then a state change occurs (or vice versa), the energy transferred to the sample has two components:

- the energy required to raise the temperature,
- and the energy required to change the state.

These can be calculated separately using the specific heat capacity and the relevant specific latent heat of the substance. The total energy transferred is the sum of these two components.

P5.4

Density:

a. Know and be able to apply: $\text{density} = \frac{\text{mass}}{\text{volume}}$, $\rho = \frac{m}{V}$
b. Understand the experimental determination of densities.
c. Be able to compare the densities of solids, liquids and gases.

Know and be able to apply: $\boldsymbol{density = \frac{mass}{volume}; \ \rho = \frac{m}{V}}$

The density of a substance is the mass per unit volume of that substance. The mass, volume and density of a sample of a substance are related by the formula:

$$density = \frac{mass}{volume}$$

Density is typically expressed in units of kg m^{-3} or g cm^{-3} (though g mL^{-1} can be used for a liquid).

While different samples of a particular substance may have different masses, all will have the same density. So density is one of the properties that can be used to distinguish one material from another.

To convert a density from g cm^{-3} to kg m^{-3}, use the fact that:

1 kg = 10^3 g

1 m^3= $(100cm)^3$ = 10^6 cm^3

For example, aluminium has a density of 2.7 g cm^{-3}, which is equivalent to 2700 kg m^{-3}:

$$\frac{2.7g}{1cm^3} = \frac{2.7 \times 10^{-3}kg}{1 \times 10^{-6}m^3} = 2.7 \times 10^3 kgm^{-3}$$

An object that is made of more than one substance has an average density, given by its total mass divided by its total volume.

The density of solid aluminium is 2.700 g cm^{-3}.

a) Write the density of aluminium in kg m^{-3}.
b) Calculate the volume of a 0.500 kg sample of aluminium.

a)

1 g cm^{-3} = 10^{-3} kg cm^{-3}

=$10^{-3} \times 10^{6}$ kg m^{-3}

=1000 kg m^{-3}

Therefore

2.700 g cm^{-3} = 2.700 × 1000 = 2700 kg m^{-3}

b)

$$V = \frac{m}{\rho}$$

= 500 g/2.7 gcm-^{3}

= 185.18 … cm^{3}

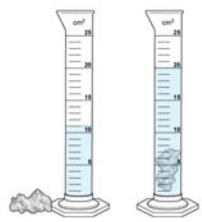
= 185 cm^{3} (to 3 s.f.)

Understand the experimental determination of densities

The density of a substance can be calculated by measuring the mass and volume of a sample of that substance, and then using the relationship:

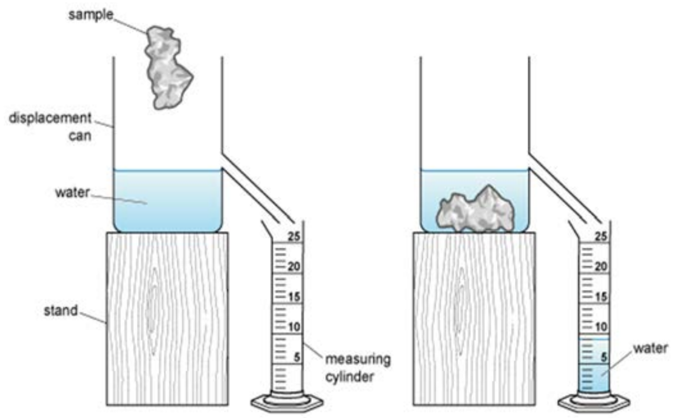
$$density = \frac{mass}{volume}$$

The experimental method varies depending on the nature of the sample, as shown in the table.

Nature of sample	Method for measuring mass	Method for measuring volume
Solid cuboid (or other shape whose dimensions can be measured)	Measure the mass using a balance. 	Measure the sample's dimensions using a ruler or other suitable instrument. Use these measurements to calculate the volume of the sample.
Solid with irregular shape (or with dimensions that cannot be measured using available instruments)	Measure the mass using a balance. 	Partly fill a suitably sized measuring cylinder with water. Measure the volume of the water. Place the sample carefully into the cylinder. Measure the new volume. Find the difference between the two measurements. 
Liquid	Use a balance to find the mass of a measuring cylinder. Pour the sample into the cylinder. Measure the mass of cylinder+sample. Find the difference between the two mass measurements.	Measure the liquid volume using the scale on the measuring cylinder.

To use the method for measuring the volume of an irregularly shaped solid sample, there must be enough water initially in the cylinder to completely cover the sample, and the water must not overflow (or rise above the top of the measuring scale) when the sample is introduced.

An alternative method for measuring the volume of an irregularly shaped solid sample is to use a displacement can, as shown in the diagram below. The can is filled with water up to the level of its spout. The object is placed in the water, causing some water to overflow into a measuring cylinder. The volume of water in the measuring cylinder equals the volume of the solid sample. *Measuring using a displacement can*



The methods described for measuring the volume of an irregularly shaped solid sample can be used as long as the sample does not react with or dissolve in the water and the sample sinks in water (which it will if it is denser than water).

If the sample floats on water, an adapted method can be used. Tie a heavy block of known volume to the sample. Measure the volume of this combination using the displacement of water. Subtract the volume of the heavy block to find the volume of the sample.

An irregularly shaped solid object has a mass of 35.2 g.

To measure its volume, a student tries to use a displacement can, but finds that the object floats in water. The student ties a block of mass 447 g and density 8.94 g cm^{-3} to the object so that it sinks. The volume of water displaced by the block and the object is 105.0 cm^3.

Calculate the density of the object.

The volume of the block is $V = \frac{m}{\rho} = \frac{447}{8.94} = 50\ cm^3$

The volume of the object is therefore 105.0 − 50.0 = 55.0 cm^3.

The density of the object is $\rho = \frac{m}{V} = \frac{35.2}{55.0} = 0.640\ g\ cm^{-3}$

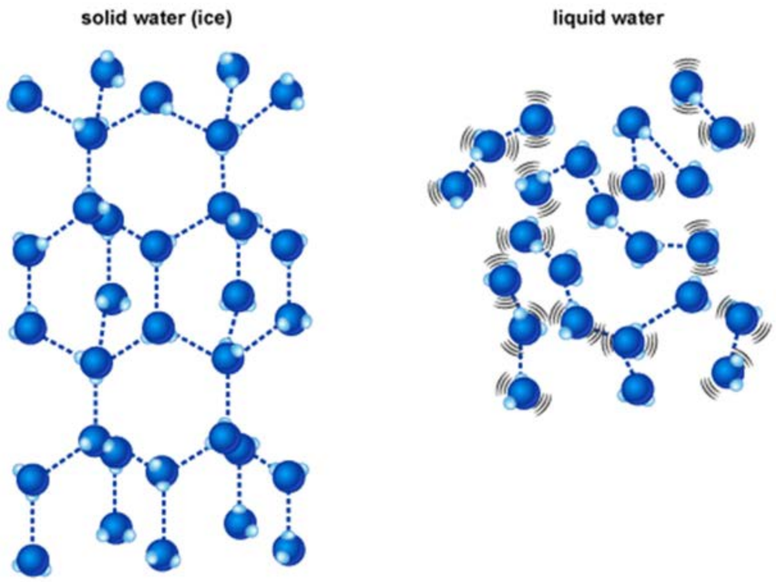
Be able to compare the densities of solids, liquids and gases

Densities of solids and liquids

The different densities of different substances can be explained by the particle model: the density of a substance depends on the mass of its particles and how closely packed the particles are. These both vary considerably between substances.

Most substances are slightly denser in their solid state than in their liquid state. This is because the particles are slightly closer together in the solid state, so solids have slightly more mass per unit volume. The difference is usually small (a substance is typically around 1.1 times denser in its solid state than in its liquid state).

There are exceptions, for example water: ice is less dense than liquid water. This is because the regularly arranged molecules in ice are slightly further apart than the randomly arranged molecules in liquid water.



Densities of gases

All substances are much less dense in the gaseous state than in the liquid and solid states. This is because the particles in a gas have much greater separation, so there are far fewer particles – adding up to far less mass – per unit volume. The density of a liquid typically increases by a factor of around 1000 when it changes into a gas.

Gas density depends on the temperature and pressure (which both affect the particle separation), so the density of a gas is usually quoted at particular values of temperature and pressure.

Both solids and liquids have a wide range of densities, with the densities of some solids being lower than the densities of some liquids. For example, most types of wood are less dense than liquid water (and hence float on water), and there are many substances in the solid state that are less dense than liquid mercury.

(It is even possible for the solid form of one substance to be less dense than the gaseous form of another. For example, graphene aerogel is a solid that is less dense than air.)

The densities of many solids and liquids are in the range 1000–10 000 kg m^{-3} (but some substances have densities that are considerably higher or lower than this: for example, the density of the packing material expanded polystyrene is about 20 kg m^{-3}, while the density of solid gold is about 19 000 kg m^{-3}).

Densities of gases are typically around 1 kg m^{-3} (at room temperature and atmospheric pressure).

Use the particle model to explain why solids and liquids usually have similar densities but gases have much lower densities.

According to the particle model, the particles in solids and liquids are very close together relative to their size. The particles in gases are far apart relative to their size. So in a gas, there are far fewer particles per unit volume, and therefore much lower mass per unit volume, corresponding to a much lower density.

P5.5

Pressure:

a. Know and be able to apply: $\text{pressure} = \frac{\text{force}}{\text{area}}$
b. Know and be able to apply: hydrostatic pressure = $h\rho g$, where h is the height, or depth, of the liquid.

Know and be able to apply: $\boldsymbol{pressure = \frac{force}{area}}$

When a force is exerted on a surface, the pressure on the surface is defined by the relationship

$$pressure = \frac{force}{area}$$

where area refers to the area of contact over which the force acts. So the pressure is the force per unit area acting on the surface.

If the force is in newtons and the area is in square metres, then the pressure is in pascals (symbol Pa), where 1 N $m^{-2}$is defined as 1 Pa.

If an object rests on a horizontal surface, then the force exerted on the surface equals the weight of the object. For example, if a box of mass 12 kg and base area 0.50 m^2 rests on a horizontal surface, the pressure exerted by the box on the surface is

$$\text{pressure} = \frac{\text{weight}}{\text{area}} = \frac{\text{mass} \times \text{gravitational field strength}}{\text{area}} = \frac{12 \times 10}{0.50} = 240 \text{ Pa}$$

As the equation implies, pressure can be increased either by increasing the applied force, or by decreasing the area over which the force is applied. Pressure can be decreased either by decreasing the applied force, or by increasing the area over which the force is applied.

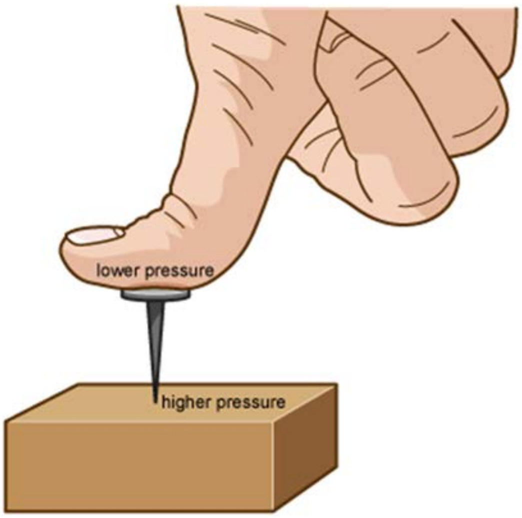
Items that are designed to exert a high pressure include sewing needles and knives. These have a small area of contact at the sharp end, which maximises the pressure resulting from the force applied by the user at the other end.

This enables the pin to pass through cloth and the knife to cut through food.

Items that are designed to exert a low pressure include snowshoes and skis. These have a large area of contact with the surface of the snow, to minimise the pressure resulting from the weight of the person wearing them. This prevents the wearer from sinking into the snow.

A drawing pin (or thumbtack) has a sharp point at one end and a large area at the other end. It can be pushed into a notice board by pressing the large end with the thumb. The force at the sharp end equals the force exerted by the thumb, but the pressure at the sharp end is much greater. Thus, the drawing pin can be pushed in without causing pain or damage to the thumb.

Pressure applied to a drawing pin



The pressure on a surface underneath an object is the sum of two contributions: the pressure due to the weight of the object, and the pressure due to any gas above it (such as the atmosphere). The phrase 'the pressure due to the solid' can be used to refer to the contribution from the solid only.

A table has four legs. Each leg has a square cross-section, of width 5.0 cm, in contact with the ground. The weight of the table is 250 N.

Calculate the pressure exerted by the table legs on the ground.

The total area of contact between the table and the ground is

$4 \times 0.050^2 = 0.010 \text{ m}^2$

(The area needs to be in m^2 for calculating the pressure in Pa.)

The pressure exerted by the feet on the ground is

$pressure = \frac{force}{area} = \frac{250}{0.010} = 25000\ Pa$

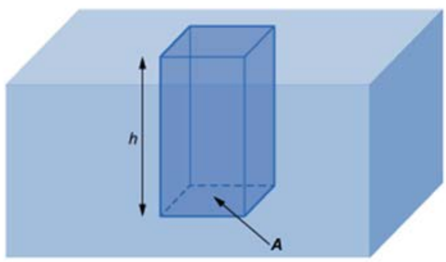
Know and be able to apply: hydrostatic pressure = hρg, where h is the height, or depth, of the liquid

Fluids are substances that can flow: liquids and gases. The pressure in a fluid increases with depth, and acts in all directions. The pressure at a point within a fluid is due to the weight of fluid above that point – so pressure increases with depth h, with fluid density ρ, and with gravitational field strength g, according to the relationship: hydrostatic pressure = hρg

If h, ρ and g are expressed in units of metres, kilograms per cubic metre, and newtons per kilogram respectively, then the calculated pressure is in pascals.

As the formula shows, pressure depends only on depth, fluid density and gravitational field strength; it does not depend on the shape of the container. For example, if the containers in the diagram below contain the same type of liquid, the pressures at the two points marked P in the diagram are equal.

The equation for calculating hydrostatic pressure can be derived by considering a cuboid of liquid within a liquid, with the upper surface of the cuboid at the surface of the liquid, as shown. The cross-sectional area of the cuboid is A and its height is h.



The pressure at the lower surface of the cuboid is due to the weight of the liquid above it.

The volume of liquid is Ah, and the mass of the liquid is

m = Vρ = Ahρ

The weight of the liquid is

mg = Ahρg

The pressure at the lower surface of the cuboid is

$P = \frac{Ah\rho g}{A} = h\rho g$

The pressure at a given depth within a liquid is the sum of two contributions: the pressure due to the liquid, and the pressure due to any gas above it (such as the atmosphere). The phrase 'the pressure due to the liquid' can be used to refer to the contribution from the liquid only.

At a particular point on the sea floor, the pressure is 3.16×10^6 Pa.

The density of the seawater is 1020 kg m^{-3} and atmospheric pressure is 1.0×10^5 Pa.

Gravitational field strength g = 10 N kg^{-1}.

Find the pressure on the sea floor due to the water.

Calculate the depth of the sea at this point.

The pressure due to the water is the total pressure minus the pressure due to the atmosphere:

$3.16 \times 10^6 - 1.0 \times 10^5 = 3.06 \times 10^6$ Pa.

Rearranging the hydrostatic pressure formula and substituting gives:

$$h = \frac{P}{\rho g} = \frac{3.06 \times 10^6}{1020 \times 10} = 300 \text{ m}$$

P6. Waves

P6.1

Wave properties:

a. Understand the transfer of energy without net movement of matter.
b. Know and understand transverse and longitudinal waves.
c. Know and understand the terms: peak, trough, compression and rarefaction.
d. Recall examples of waves, including electromagnetic waves and sound.
e. Know and be able to use the terms: amplitude, wavelength, frequency and period.
f. Know and be able to apply: $\text{frequency} = \frac{1}{\text{period}}$, $f = \frac{1}{T}$
g. Know and be able to apply: $\text{wave speed} = \frac{\text{distance}}{\text{time}}$
h. Know and be able to apply: wave speed = frequency × wavelength, $v = f\lambda$

Understand the transfer of energy without net movement of matter

Waves consist of a pattern of vibrations (oscillations).

When the wave is absorbed it transfers energy to the object that absorbs it, but there is no net transfer of matter in this direction.

Examples of energy transfer by waves

- Sound: When I speak to you, my vibrating vocal cords make the air nearby vibrate in a particular way. This pattern is passed on by the wave until the air in your ear vibrates in the same pattern and you hear my voice. No air from my mouth has passed into your ear!
- Radio waves (EM waves): When electrons vibrate in a radio transmitter, they send out electromagnetic waves that make different electrons vibrate at the same frequency in the receiving antenna, transferring energy to them. No electrons are transferred from the transmitter to the receiver.
- Water waves: Ripples on the surface of a pond consist of a pattern of peaks and troughs moving outwards. Whilst these do make the water surface move up and down, and do transfer energy outwards, they do not transfer any water away from the source.

A teacher claps his hands together to get the attention of his class.

1 Which one or more of the following statements is correct?

2 Energy was transferred from his hands to the air.

3 Energy was transferred through the air as a pattern of vibrations.

4 Vibrating air was transferred from his hands to the ears of pupils in his class.

5 Energy was transferred from the vibrating air to the ears of pupils in his class.

Clapping his hands transfers energy to the air 1) to create a a pattern of vibration (sound) in the air. This is a sound wave. The pattern travels outwards into the room 2) as a sound wave. When the vibrations reach the ears of his pupils, they transfer energy to them 4) and the pupils can hear the clap. However, no air from his hands has reached the ears of his pupils (so 3) is incorrect).

Which of the following statements about waves is incorrect?

1 Ocean waves transfer matter because they move boats up and down as they pass.

2 Seismic waves from earthquakes transfer energy because they can destroy buildings.

3 Sound waves from a loudspeaker are given energy from the speaker as it vibrates.

4 Dark surfaces gain energy when they absorb the light that falls onto them.

Whilst ocean waves do move boats up and down, there is no net movement of the water, it just oscillates about a fixed position, so 1) is incorrect. 2), 3) and 4) are all correct and illustrate how waves transfer energy from their source to the objects that absorb them.

Know and understand transverse and longitudinal waves

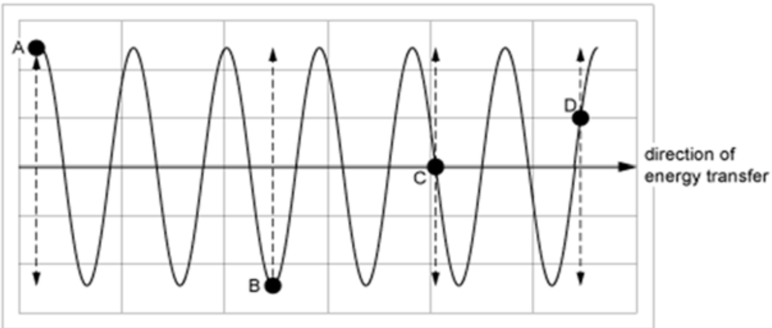
A wave consists of oscillating particles (or fields).

These vibrations can be along the same line as the wave is travelling (transverse waves) or at right angles to it (longitudinal waves).

- Transverse waves: the vibration direction is perpendicular to the wave direction.
- Longitudinal waves: the vibration direction is parallel to the wave direction.

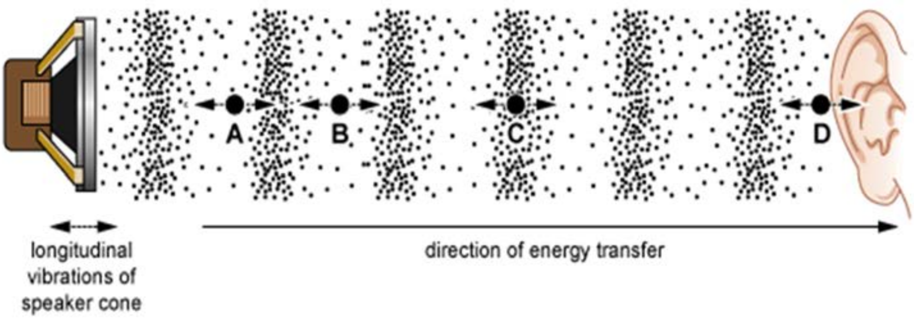
Oscillations in a transverse wave

The diagram below shows the positions of four particles, A, B, C, and D as a transverse wave moves past. All particles in the wave oscillate at 90° to the direction of energy transfer. Their vibrations are shown as vertical dashed arrows.



Oscillations in a longitudinal wave

The diagram below shows the positions of four particles, A, B, C, and D as a longitudinal wave (in this case a sound wave) moves past. All particles in the wave oscillate parallel to the direction of energy transfer. Their vibrations are shown as horizontal arrows.

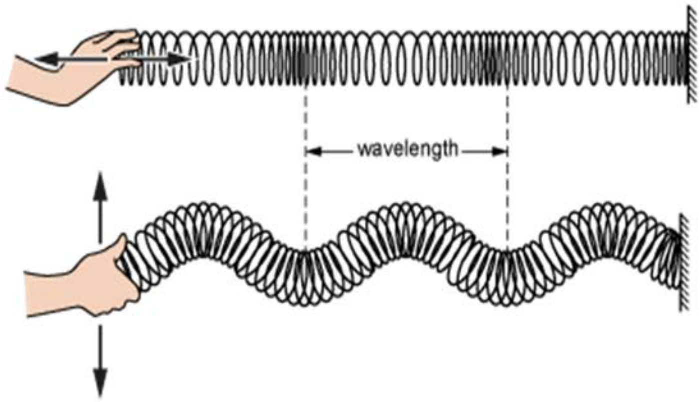


Transverse and longitudinal waves

Transverse	Longitudinal
All electromagnetic waves	Sound
Waves on a string	Ultrasound
Seismic S-waves	Compression waves on a slinky/spring
	Seismic P-waves

Water waves are a mixture of transverse and longitudinal waves so particles of water move vertically and horizontally (in an elliptical motion) as the wave passes. However, the particles have no net movement in the direction of energy transfer.

The diagrams below show how a teacher uses a slinky spring to demonstrate different types of wave motion.



Look at the table below. Which row is correct about the top diagram?

Look at the table below. Which row is correct about the bottom diagram?

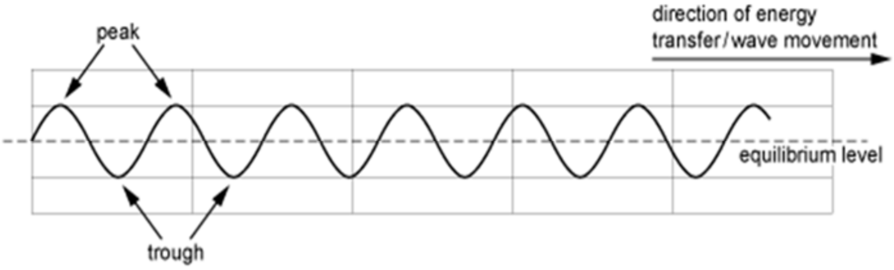
	Vibration direction relative to direction of energy transfer	Type of wave produced by the teacher
A	parallel	transverse
B	perpendicular	longitudinal
C	parallel	longitudinal
D	perpendicular	transverse

In the top diagram, the teacher's hand moves horizontally, compressing and extending the spring. This makes the coils vibrate horizontally, parallel to the direction of energy transfer. A wave in which the vibration directions are parallel to the direction of energy transfer is a longitudinal wave, so C is the correct row.

In the bottom diagram, the teacher's hand moves vertically. This makes the coils vibrate vertically, perpendicular to the direction of energy transfer. A wave in which the vibration directions are parallel to the direction of energy transfer is a transverse wave, so D is the correct row.

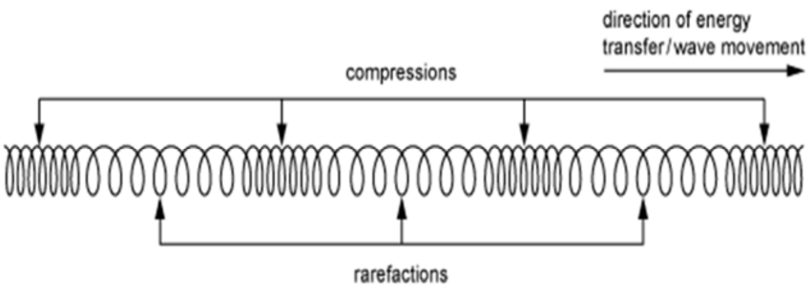
Know and understand the terms: peak, trough, compression and rarefaction

In a transverse wave the pattern of vibration creates a series of equally spaced peaks and troughs that travel away from the source of the waves.



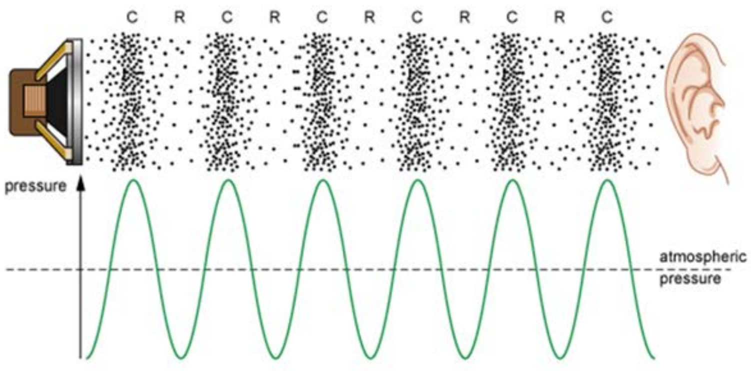
In a longitudinal wave the pattern of vibration creates a series of equally spaced compressions and rarefactions that travel away from the source of the waves.

- A compression occurs when particles in the medium are pushed together as the wave passes
- A rarefaction occurs when the particles in the medium are pulled further apart as the wave passes.



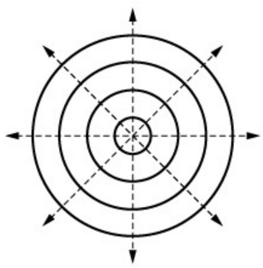
Compressions and rarefactions in a sound wave

Compressions (C) and rarefactions (R) in the air cause the pressure to vary above and below atmospheric pressure. A sound wave can be described in terms of these pressure variations or in terms of the particle displacements in the wave. Regions where the particles are pushed closer together are at a higher pressure than atmospheric and regions where they are farther apart are at a lower pressure than atmospheric.



Peaks and troughs in a water wave

The image below shows how ripples spreading out from a point on the surface of water creates a pattern of circular peaks and troughs that expand outwards from the source. Waves are often represented by wavefronts that map out the pattern of peaks (or for a longitudinal wave, the pattern of compressions). A wavefront diagram is shown on the right.



Two swimmers are floating in seawater. A wave passes them and they both move up and down, completing one cycle of motion in 5.0 s. When one of the swimmers is at the peak of the wave the other is at the next trough. The swimmers are 12 m apart in the direction in which the wave is travelling.

How far apart are adjacent peaks in the wave?

What is the frequency of the wave?

What is the speed of the wave?

The distance between a peak and the adjacent trough is half of the distance between two adjacent peaks so the distance between adjacent peaks is $2 \times 12 = 24$ m (equal to the wavelength of the waves).

Each swimmer will complete one cycle of movement as one complete wave passes. This takes 5.0 s so the number of cycles per second, i.e. the frequency, is $1/5 = 0.20$ Hz.

It will take 2.5 s for the second swimmer to move from the bottom of his motion to the top (half of a cycle). This is the time it takes for the peak to move from the first swimmer to the second, a distance of 12 m. The speed is distance / time = 12 / 2.5 = 4.8 m/s.

Alternatively, $v = f\lambda = 0.20 \times 24 = 4.8$ m/s.

Recall examples of waves, including electromagnetic waves and sound

Waves consist of patterns of vibration (oscillations).

There are two main types of wave: mechanical waves and electromagnetic waves.

- Mechanical waves (e.g. sound waves) consist of vibrating particles, so they can only move through a material medium. They cannot travel through a vacuum because there are no particles to vibrate.
- Electromagnetic waves do not need a material medium. They are vibrations of the electric and magnetic fields so they can travel through a vacuum. When electromagnetic waves travel through a vacuum they always travel at the speed of light.

Mechanical waves	**Electromagnetic waves**
Sound	Radio
Ultrasound	Microwave
Seismic waves	Infra-red
Water waves	Visible light
Waves on strings	Ultra-violet
Waves on a slinky	X-ray
	Gamma-ray

Types of mechanical waves

Sound: longitudinal vibrations of the air (or another material medium) with frequencies audible to humans.

Ultrasound: longitudinal vibrations of the air (or another material medium) at frequencies above the highest frequency audible to humans ($f > 20$ kHz).

Seismic waves: longitudinal and transverse vibrations of matter inside the Earth travelling outwards from earthquakes.

Water waves: vibrations of water particles on or below the surface of water. Water waves are partly transverse and partly longitudinal.

Waves on a string: if one end of a string is fixed and the other end is made to vibrate perpendicular to the string, transverse waves will travel along the string.

Waves on a spring or slinky: this is often used to demonstrate types of wave and wave properties. If a long spring or slinky is laid on the floor and one end is waved sideways then transverse waves can be seen travelling along the slinky. If the end is moved in and out (along the line of the slinky) then longitudinal waves can be seen travelling along the slinky. You might also see the waves reflected from the far end.

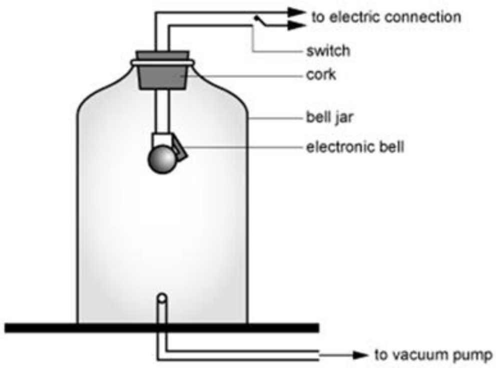
The nature of electromagnetic waves

Charged particles such as electrons set up an electric field in the space around them. When they are made to vibrate a magnetic field is also produced. The pattern of electric field and magnetic field vibrations travels outwards as an electromagnetic wave.

Visible light makes up a small region of the much wider electromagnetic spectrum.

Here is a diagram of a famous experiment.

An electric bell is suspended inside a glass jar and the air is gradually pumped out of the jar. Which of the following statements is correct about the situation when all the air has been pumped out of the jar?



1 An observer will be able to see the bell because light waves are longitudinal.

2 An observer will be able to hear the bell because sound waves are longitudinal.

3 An observer will be able to see the bell because light is a mechanical wave.

4 An observer will not be able to hear the bell because sound is a mechanical wave.

5 An observer will not be able to hear the bell because sound is a transverse wave.

When the air is removed from the jar there is a vacuum inside. Light is an electromagnetic wave, so 3) is incorrect, so it can travel through a vacuum and an observer can still see the bell. All electromagnetic waves are transverse, so 1) is incorrect. Sound is a mechanical wave, so it can only travel through a material medium and cannot travel through the vacuum inside the jar, so an observer cannot hear it, so 4) is correct. Sound waves are longitudinal waves, but this is not the reason that they cannot travel through a vacuum, so 2) and 5) are both incorrect.

The film *Alien* was advertised with the tagline: ‘In space no one can hear you scream’.

Which one or more of the following statements could explain this?

1 Space is a vacuum.

2 Sound is a mechanical wave.

3 Electromagnetic waves can travel through a vacuum.

4 Sound is a longitudinal wave.

5 Light is an electromagnetic wave.

A scream is a sound. Sound is a mechanical wave so it can only travel through a material medium, so 2) is correct. Space is a vacuum containing no particles, so sound cannot travel through space, so 1) is correct). 3) and 5) are both about electromagnetic waves so whilst both are true, they are not relevant to a scream. 4) is also true but not the reason sound cannot travel in a vacuum.

Know and be able to use the terms: amplitude, wavelength, frequency and period Wavelength

- The distance between adjacent peaks (or troughs) in a transverse wave.
- The distance between adjacent compressions (or rarefactions) in a longitudinal wave.

Amplitude

- The maximum displacement of a particle in the wave from its equilibrium position.

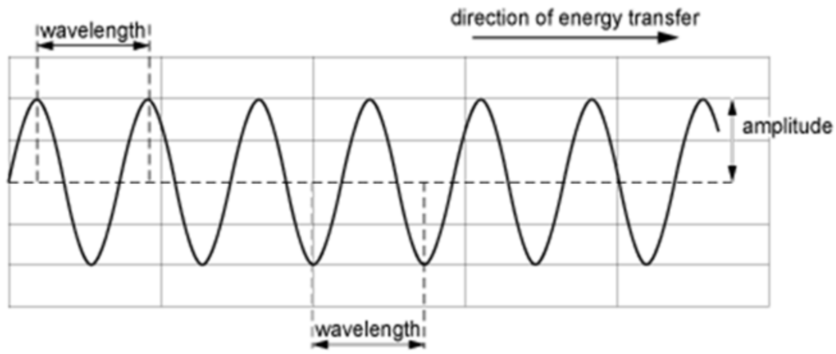
Period T

- The time taken to complete one cycle of vibration (oscillation). Measured in seconds.

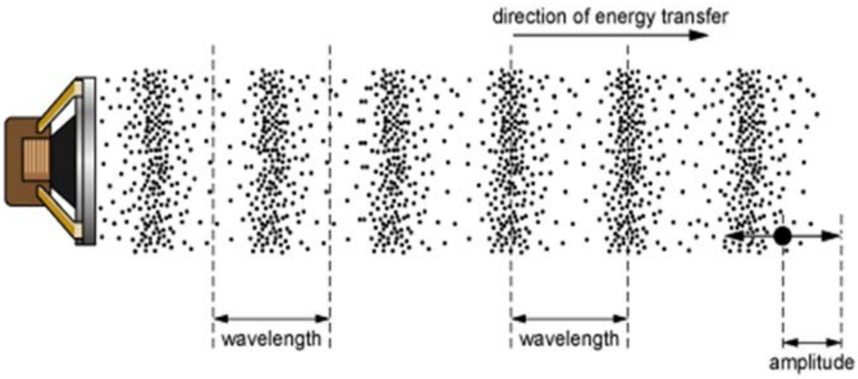
Frequency f

- The number of vibrations (oscillations) per unit time at a point in the wave. Measured in hertz (Hz). 1 Hz = 1 oscillation per second.

Transverse waves

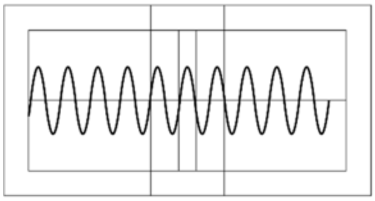


Longitudinal waves

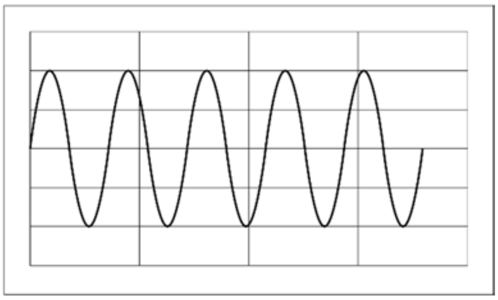


The diagrams below show two different transverse waves drawn to the same scale. The waves both have the same speed.

(1)



(2)



Which row in the table below states the correct relationships between the two waves?

	Longer wavelength	Higher frequency	Larger amplitude
A	Wave 1)	Wave 1)	Wave 1)
B	Wave 1)	Wave 1)	Wave 2)
C	Wave 1)	Wave 2)	Wave 1)
D	Wave 1)	Wave 2)	Wave 2)
E	Wave 2)	Wave 1)	Wave 1)
F	Wave 2)	Wave 1)	Wave 2)
G	Wave 2)	Wave 2)	Wave 1)
H	Wave 2)	Wave 2)	Wave 2)

Wavelength is the distance between adjacent peaks. Since the two diagrams are drawn to the same scale, it is clear that the distance between adjacent peaks is greater for Wave 2).

Frequency is the number of oscillations per unit time. Since the two waves have the same speed, the peaks in Wave 1) are closer together and more will pass a point per unit time than in Wave 2) so Wave 1) has the higher frequency.

Amplitude is the maximum displacement from the equilibrium position. Since the two waves are drawn to the same scale and the vertical displacement of Wave 2) is greater, Wave 2) has the greater amplitude.

The correct row is therefore F: Wave 2), Wave 1), Wave 2).

A vibrator is used to generate ripples on the surface of water in a shallow container. The ripples travel outwards at 1.2 cm s^{-1} and a cork floating in the water bobs up and down 20 times in a minute as the waves pass. The distance between the top and bottom of its motion is 8.0 mm.

Calculate:

the period of the wave

the frequency of the wave

the wavelength of the wave

the amplitude of the wave.

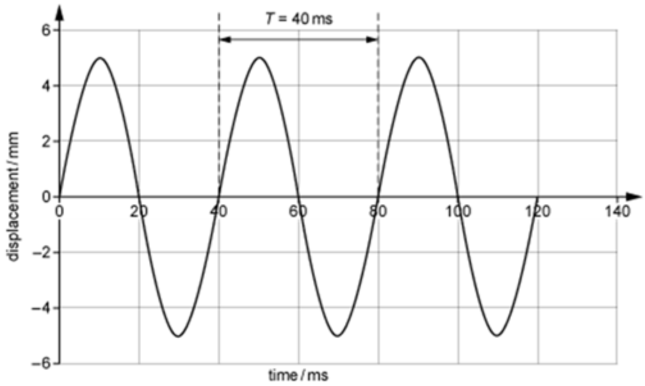
The cork completes 20 cycles of vibration in 60 s so the time for one complete cycle is 60/20 = 3.0 s.

If there are 20 complete cycles in 60 s there will be 20/60 complete cycles in 1 s, so the frequency is 0.33 Hz.

The wavelength is equal to the distance moved by the wave in one cycle i.e. 1.2 × 3.0 = 3.6 cm.

Amplitude is maximum displacement from equilibrium. This is half the total vertical distance from the trough to the peak, i.e. 8.0/2 = 4.0 mm.

Know and be able to apply $frequency = \frac{1}{period}$ $\quad f = \frac{1}{T}$



Calculating frequency from a graph of particle displacement against time:

From the graph: T = 40 ms = 0.040 s

Frequency: $\quad f = \frac{1}{T} = \frac{1}{0.040} =$ 25 Hz

Prefixes are often used for short time intervals or high frequencies:

Prefixes

milli - 1 ms = 0.001 s = 10^{-3} s

micro - 1 µs = 0.000001 s = 10^{-6} s

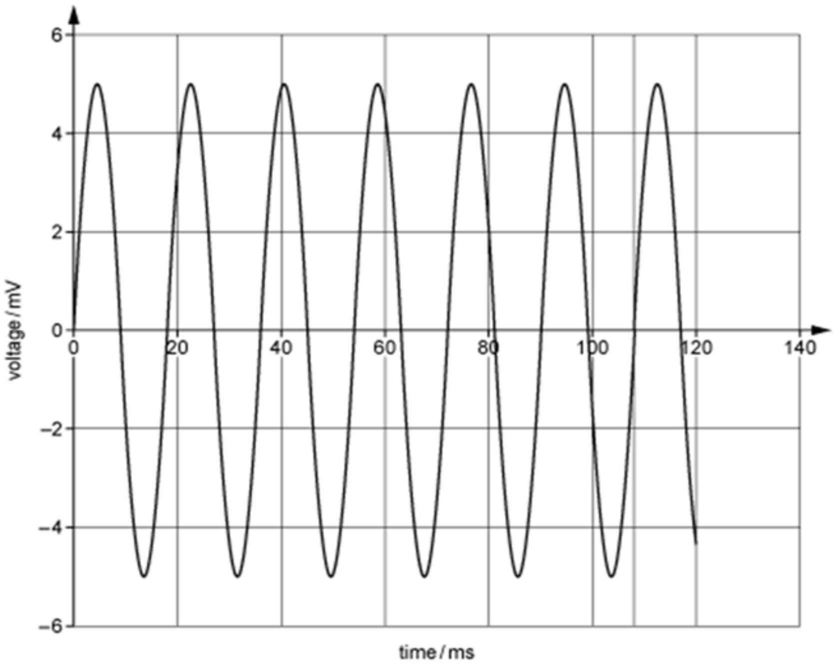
nano - 1 ns = 0.000000001 s = 10^{-9} s

kilo - 1 kHz = 1000 Hz = 10^{3} Hz

mega - 1 MHz = 1000 000 Hz = 10^{6} Hz

giga - 1 GHz = 1000 000 000 Hz = 10^{9} Hz

When a microphone detects sound, it converts the sound vibrations into a varying voltage signal. The signal produced by a particular microphone is shown below.



Use the graph to determine the frequency of the sound.

6 complete cycles are shown: 6 T = 108 ms

so T = 18 ms

$f = \frac{1}{T}$ = 56 Hz

Humans can hear sounds in the range 20 Hz to 20 kHz. What is the shortest time period of sound vibrations that can be heard?

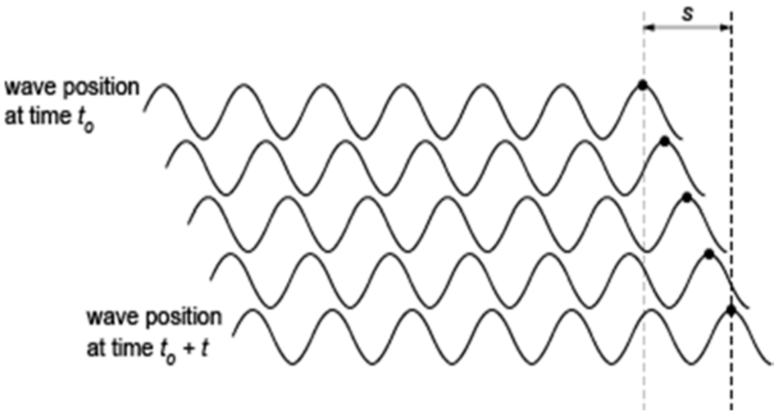
The shortest time period will correspond to the highest frequency, i.e. 20 kHz or 20 000 Hz.

Rearranging the equation for frequency, we get: $T = \frac{1}{f}$ so T = $\frac{1}{20000}$ = 0.000050 s.

Know and be able to apply: $\boldsymbol{wave\ speed = \frac{distance}{time}}$

Wave speed is the speed at which an identifiable point in the wave pattern moves, e.g. a peak or a trough in a transverse wave or a compression or a rarefaction in a longitudinal wave. If the point moves a distance s in a time t then the wave speed is $v = \frac{s}{t}$.

The diagram below shows the position of a peak in a transverse wave at five equally spaced instants.



One peak has been marked with a dot.

This peak has moved a distance s in a time t

The wave speed is equal to distance / time: $v = \frac{s}{t}$

A similar approach could be used to follow the movement of a compression in a longitudinal wave. One way to determine wave speed is to measure the time taken for the wave to travel a measured distance.

Ranging methods to measure distance

If we know the speed of a wave, we can use it to measure distances. For example:

Laser range finding

A short pulse of laser light is reflected from an object and the time for the pulse to travel to the object and back is measured. This method is used for surveying and has been used to measure and monitor the distance to the Moon.

Ultrasound scanning

A short pulse of ultrasound is reflected from boundaries inside a patient and the time taken for the pulse to return is measured and used to determine the distance to the boundary or to construct an image, e.g. of an unborn baby.

Distance is calculated using:

distance to object = $\frac{1}{2}$ × wave speed × time for pulse to complete return trip

(The factor of 2 is because the pulse travels to the distant object and returns.)

The distance from the Earth to the Moon is 380 000 km and increasing at a rate of 3.8 cm per year. Radio waves are used to communicate with astronauts on the surface of the Moon.

Calculate the minimum communication delay caused by the Moon's distance.

By how much will this delay increase in one year?

Speed of radio waves in space = 3.00×10^{8} ms^{-1}

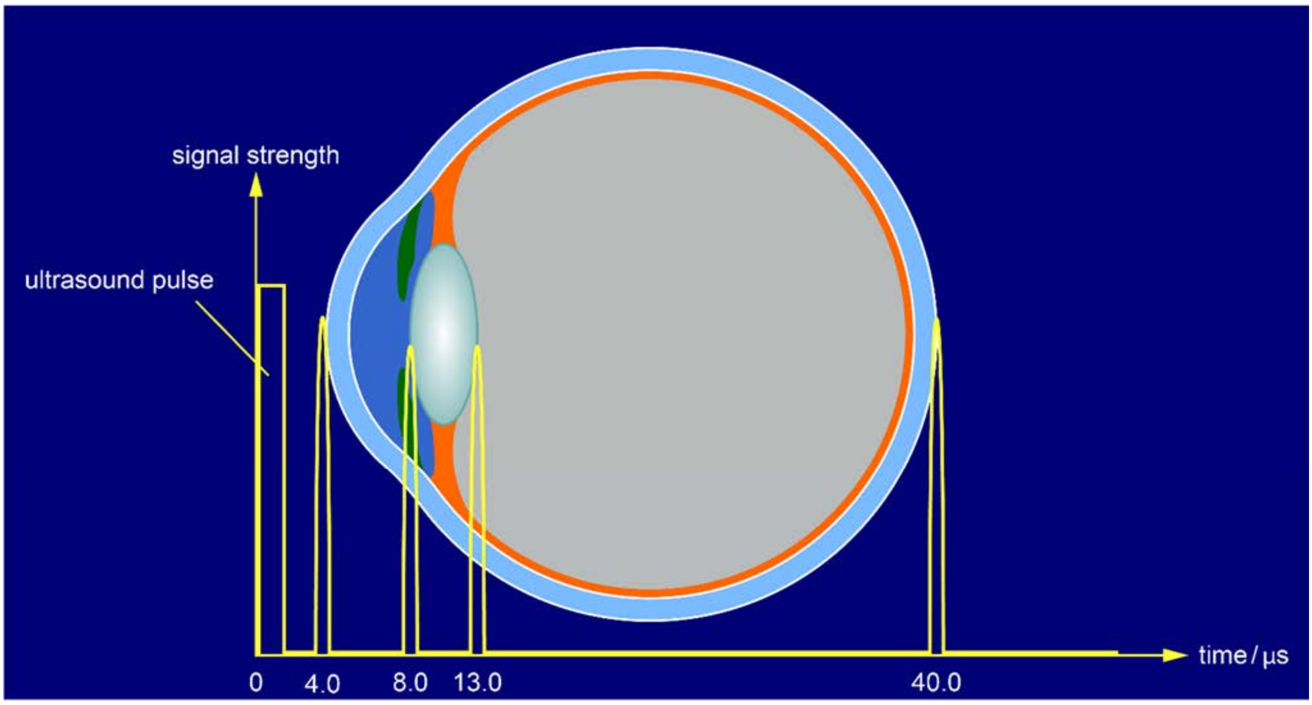
distance to object = $\frac{1}{2}$ × wave speed × (time for pulse to complete return trip)

$$\text{time} = \frac{2 \times \text{distance}}{\text{wave speed}} = \frac{2 \times 3.80 \times 10^{8}\text{m}}{3.00 \times 10^{8}\text{ms}^{-1}} = 2.53\text{ s}$$

extra distance to object = $\frac{1}{2}$ × wave speed × (extra time for pulse to complete return trip)

$$\text{time} = \frac{2 \times \text{extra distance}}{\text{wave speed}} = \frac{2 \times 3.80 \times 10^{-2}\text{m}}{3.00 \times 10^{8}\text{ms}^{-1}} = 2.53 \times 10^{-10} = 0.253\text{ ns}$$

Ultrasound scans (A-scans) are used to measure the size and structure of the eye, particularly when diagnosing and planning to remove cataracts. The procedure transmits an ultrasound pulse from the surface of the cornea and detects reflections from each tissue boundary. Typical results are shown here:



The graph shows the times at which different reflections return.

Use this to work out:

The length of the eye.

The thickness of the lens.

The speed of ultrasound inside the eye is (on average) 1550 ms^{-1}.

Reading from the graph, the times for reflections to return are:

From front surface of eye: $t_0 = 4.0$ μs

From back surface of eye: $t_1 = 40.0$ μs

From back surface of lens: $t_2 = 13.0$ μs

From front surface of lens: $t_3 = 8.0$ μs

Length of eye $= \frac{1}{2}v\,(t_1 - t_0) = \frac{1}{2}\,1550 \times (36 \times 10^{-6}) = 0.028$ m, 2.8 cm

Thickness of lens $= \frac{1}{2}v\,(t_2 - t_3) = \frac{1}{2}\,1550 \times (5.0 \times 10^{-6}) = 0.0039$ m, 3.9 mm

Know and be able to apply: wave speed = frequency × wavelength, v = fλ wave speed (m/s) = frequency (Hz) × wavelength (m)

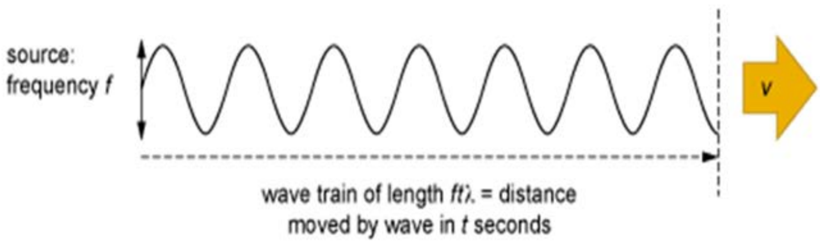
$v = f\lambda$

Derivation

A source emits waves at frequency f and with wavelength λ. In a time t the source emits ft complete waves, each of length λ. The length of this 'wave train' is ft × λ metres. The wave at the front of the wave train was emitted at t = 0 and has travelled a distance of ftλ metres.

Its speed is:

$$v = \frac{distance}{time} = \frac{ft\lambda}{t} = f\lambda$$



The range of human hearing is 20 Hz to 20 kHz. What is the corresponding range of wavelengths? The speed of sound in air is 330 ms^{-1}.

At 20 Hz:

$$\lambda = \frac{v}{f} = \frac{330}{20} = 16.5\ m$$

At 20 kHz (= 20 000 Hz):

$$\lambda = \frac{v}{f} = \frac{330}{20000} = 16.5\ mm$$

When a source of electromagnetic waves moves away from us, the waves we receive are Doppler shifted to longer wavelengths and lower frequencies, but the wave speed is unchanged and equals the speed of light, 3.00×10^8 ms^{-1}. Light emitted from a distant galaxy with a wavelength of 430 nm is detected on Earth at a frequency of 6.67×10^{14} Hz.

Calculate the size of the shift in wavelength as a result of the motion of the galaxy.

The wavelength received on Earth is:

$$\lambda = \frac{\text{speed of wave}}{\text{frequency received on Earth}} = \frac{3.00 \times 10^8}{6.67 \times 10^{14}} = 4.50 \times 10^{-7} = 450 \text{ nm}$$

Wavelength shift = 450 − 430 = 20 nm (increase).

Photo: Pixabay

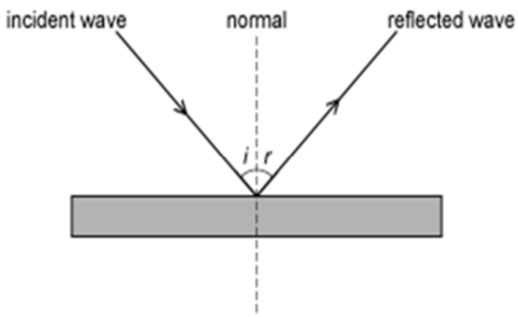
P6.2

Wave behaviour:

a. Know and understand reflection at a surface.
b. Know and understand refraction at a boundary.
c. Know and understand the effect of reflection and refraction on the speed, frequency, wavelength and direction of waves.
d. Know and understand the analogy of reflection and refraction of light with that of water waves.
e. Know and understand the Doppler effect.

Know and understand reflection at a surface

When a wave strikes a surface, all or part of the wave energy can reflect off the surface. The diagram below shows how waves reflect from a smooth surface.



i is the incident angle: the angle between the normal and the direction of the incident wave r is the reflected angle: the angle between the normal and the direction of the reflected wave

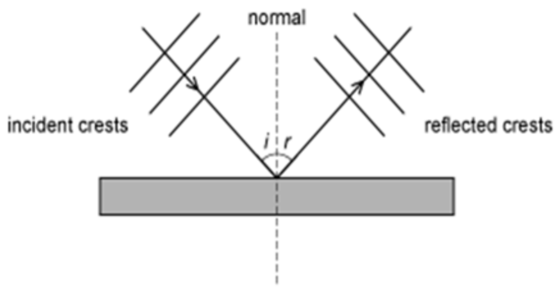
The law of reflection

incident angle = reflected angle

(Both rays and the normal must also lie in the same plane).

Reflection of crests and troughs

Sometimes wave diagrams show the positions of the crests (or troughs) of a wave before and after reflection. This is also what you will see if you watch the reflection of water waves from a barrier (e.g. in a ripple tank). Crests and troughs are perpendicular to the rays and the diagram below shows reflection of ripples from a straight barrier:



Reflection from smooth and rough surfaces

The law of reflection applies at every point where waves hit a surface. If the surface is smooth all the normals are parallel to one another so all the waves are reflected in an orderly way (e.g. reflection from a mirror) and images can be formed. If the surface is rough the normals at each point are in different directions so each ray is reflected in a random direction (e.g. reflection from a white sheet of paper).

Astronomers have measured the distance to the Moon by timing a laser pulse to travel to the Moon and reflect back to Earth. The average time for the pulse to do this was 2.57 s. What is the distance in km to the Moon?

(The speed of light waves is 3.00×10^8 ms^{-1})

The light waves in the pulse reflect from the surface of the Moon and return to the Earth where they are detected. The total distance travelled is twice the distance from the Earth to the Moon.

Total distance = speed × time = $3.00 \times 10^8 \times 2.57 = 7.71 \times 10^8$ m

Distance to Moon = $\frac{1}{2} \times 7.71 \times 10^8$ m = 3.86×10^8 m = 386 000 km

A man standing 170 m from a vertical cliff bangs two pieces of wood together and hears the reflected sound (an echo) a short time later. At what frequency should he bang the pieces of wood together so that the echoes come exactly halfway between the hits?

0.5 Hz

1.0 Hz

2.0 Hz

4.0 Hz

For the echoes to come halfway between hits the time between hits must be double the time taken for the echoes to travel to the cliff and back.

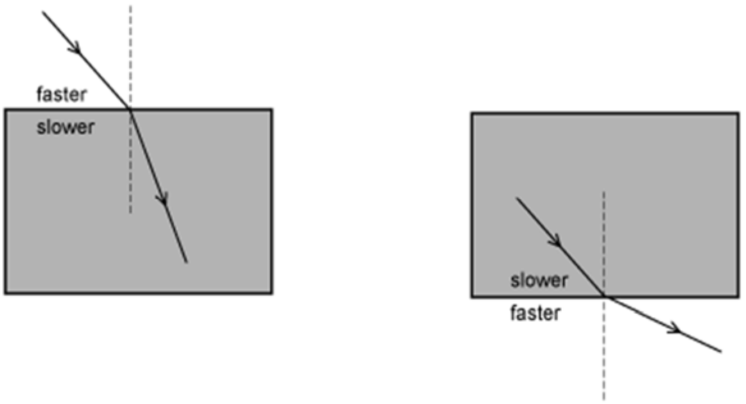
This is a distance of 340 m and takes a time of 340/340 = 1.0 s. The time between hits must therefore be 2.0 s and the frequency must be f = 1/T = 1/2.0 = 0.50 Hz.

Know and understand refraction at a boundary

When waves cross a boundary between two different media in which the waves travel at different speeds they refract.

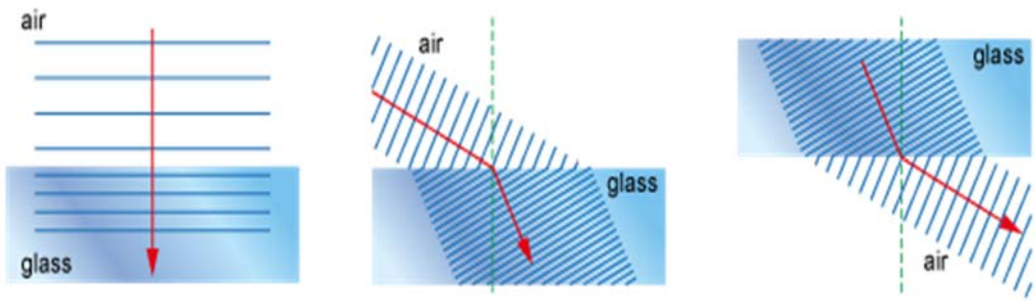
If they are not travelling parallel to the normal, then the wave direction changes as they cross the boundary.

- If a light ray slows down, it refracts towards the normal [e.g. air to glass]
- If a light ray speeds up, it refracts away from the normal [e.g. glass to air]



If the waves are travelling along the normal they continue in the same direction (but still change speed).

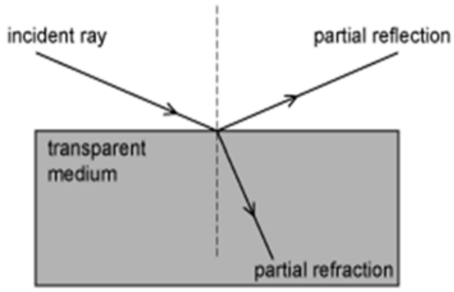
Diagrams of refraction often show the effect on wave crests (or troughs). These are perpendicular to the direction of travel of the waves. The diagrams below show glass crossing an air/glass or glass/air boundary. Light travels slower in glass than in air.



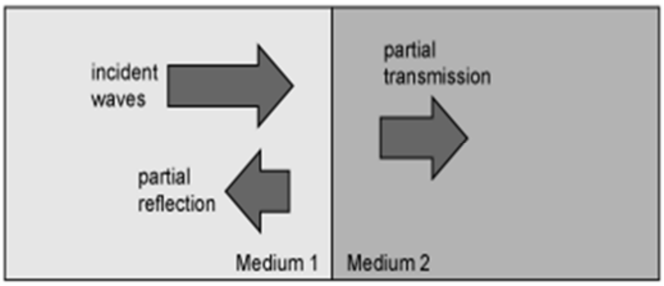
The easiest way to draw a diagram like this is to draw the rays first and then add the crests (wavefronts) perpendicular to them. Note that the crests are continuous at the boundary.

Partial reflection

It is unusual for 100% of the incident wave energy to be reflected, some is likely to be absorbed by the surface material and, if the material is transparent, some will refract into the material.



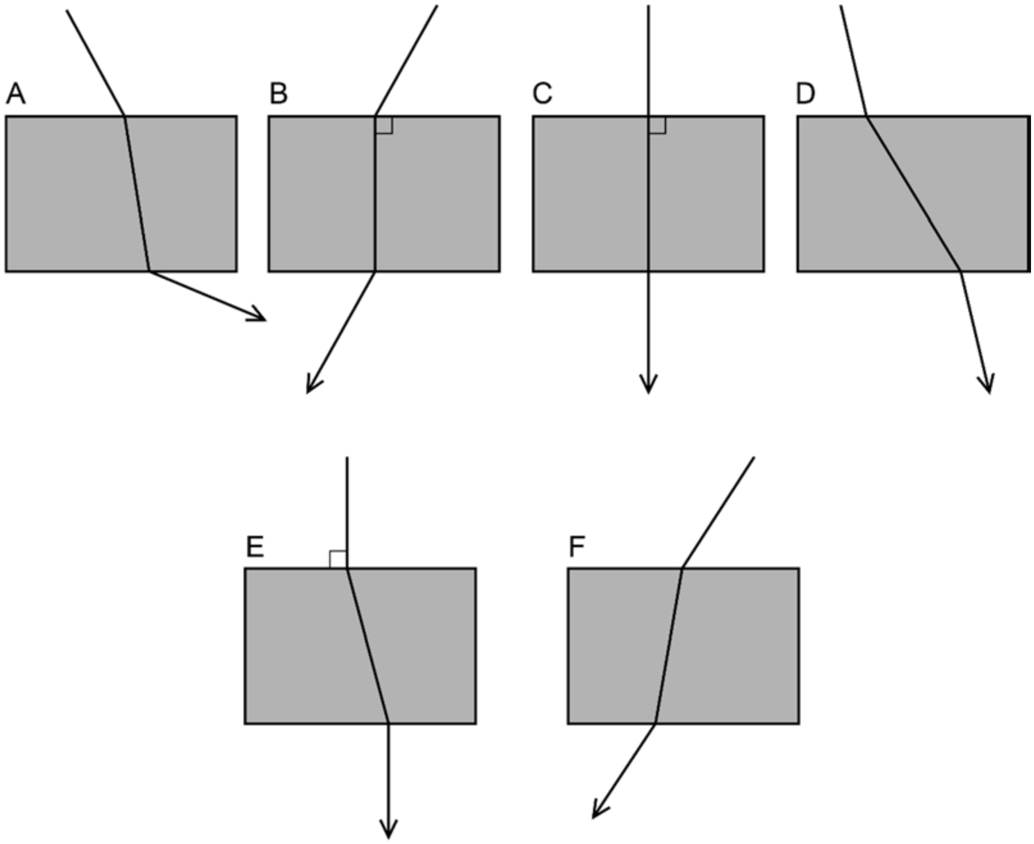
At a boundary between two media some of the wave energy will be absorbed, some will be transmitted and some will be reflected:



Energy is conserved at the boundary:

Incident energy = reflected energy + transmitted energy + absorbed energy

Here are six diagrams showing rays of light passing from air through a transparent glass block. Which one or more of these diagrams is or are possible as a result of refraction?



Light slows down when it moves from air to glass and refracts toward the normal. It speeds up when it moves from glass to air and refracts away from the normal. The sides of the block are parallel so the effect at the bottom of the block will be the inverse of the effect at the top.

A (not possible): refraction is in the correct sense at both boundaries, but the emerging ray has refracted more than the entering ray. These should be parallel.

B (not possible): the entering ray refracts toward the normal but could only continue along the normal if its original direction was along the normal.

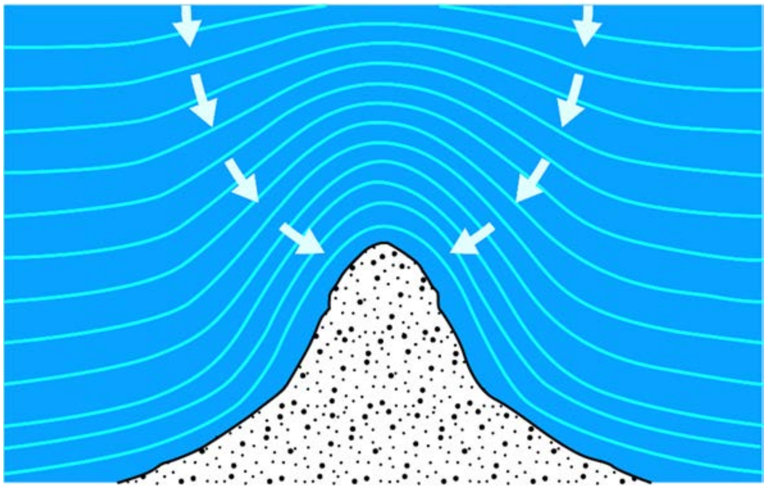
C (possible): the incident ray is parallel to the normal so although it slows down it does not change direction and continues along the normal.

D (not possible): the ray has refracted in the wrong direction at both boundaries.

E (not possible): if the incident ray was along the normal it would continue in the same direction through the block, as in C.

F (possible): the ray has refracted in the correct direction and by the same amount at both boundaries. The emerging ray is parallel to the incident ray.

The diagram below shows the direction of movement of wave crests as they approach a coastline. The depth of water gets lower closer to the coast.



Which of the following statements could help to explain the changes in direction and wavelength of the waves?

1 The waves travel more slowly in shallower water.

2 The waves refract toward the coastline.

3 The waves travel faster as they approach the shore.

It is clear from the diagram that the waves have a shorter wavelength closer to the shore. The wave frequency cannot change so the speed must be less close to the coastline – statement 1) is correct. A change of wave speed causes refraction and the direction changes toward the coastline – statement 2) is correct. Statement 3) would contradict the two points above and cause the waves to refract away from the coastline, so statement 3) does not help to explain the effect.

Know and understand the effect of reflection and refraction on the speed, frequency, wavelength and direction of waves

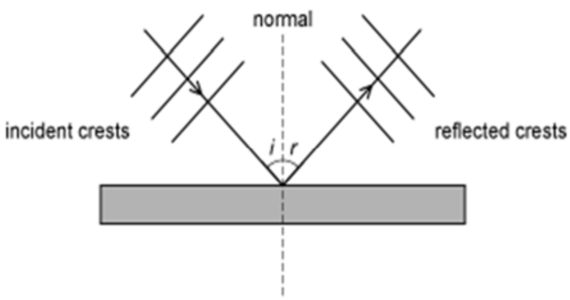
Reflection changes the direction of a wave. However, the reflected wave remains in the same medium so its speed, frequency and wavelength are all unchanged.

Refraction is caused by a change of wave speed when the wave crosses the boundary between two different media. This also affects the wavelength but has no effect on frequency. The direction of the wave also changes unless the incident wave is travelling along the normal to the boundary (i.e. the incident angle is zero, i = 0).

Effect on crossing boundary	Frequency	Wavelength	Direction of refraction
Increase of speed on crossing boundary	No change	Increases	Away from normal
Decrease of speed on crossing boundary	No change	Decreases	Toward normal

Explaining the effects of reflection

Look at the diagram below, which shows waves reflecting from a surface.



Frequency

Every wave crest that reaches the surface reflects from it and continues in the same medium so the number of waves per second before and after reflection is the same – the frequency is unchanged by refraction.

Wavelength

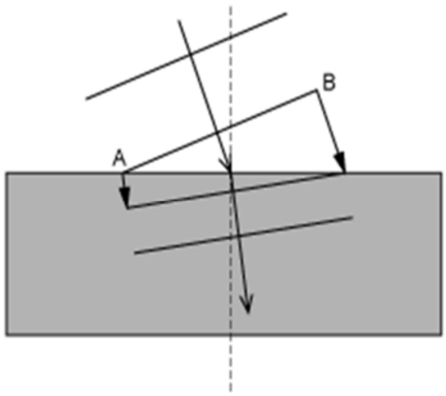
As the wave remains in the same medium, its speed is constant so the distance moved during one cycle of oscillation, the wavelength, is also constant. Another way to see this is to use the wave equation $v = f\lambda$, since v and f are unchanged, the wavelength is the same before and after reflection.

Direction

The wave direction is changed so that the angle between the incoming ray and the normal is equal to the angle between the reflected ray and the normal.

Explaining the effects of refraction

Look at the diagram below, which shows waves reaching a boundary, slowing down and refracting.



Frequency

Every wave crest that reaches the boundary enters the second medium so the number of waves per second in each medium is the same – the frequency is unchanged by refraction.

Wavelength

As the wave enters the second medium, its speed changes so the distance moved during one cycle of oscillation, the wavelength, also changes in the same way as the speed. In this case the wave slows down so this distance is reduced and the wavelength is shorter. Another way to see this is to use the wave equation $v = f\lambda$, since v decreases and f is unchanged, so wavelength must also decrease.

Direction

If the wave is travelling at a non-zero angle to the normal, different parts of the wave crest enter the second medium at different times. In this case, the left-hand end A) enters first and slows down first, moving a shorter distance than the right-hand end B) of the same wave crest. This causes the direction of the wave to change. When the wave slows down it refracts toward the normal.

An astronomer is using a telescope to observe a distant star. Light from the star has to pass through the Earth's atmosphere to reach the telescope and as it does so, it slows down. Which of the statements below, about the apparent position of the star is/are correct?

1 If the star is vertically above the telescope its apparent position in the sky will be the same as its actual position in the sky.

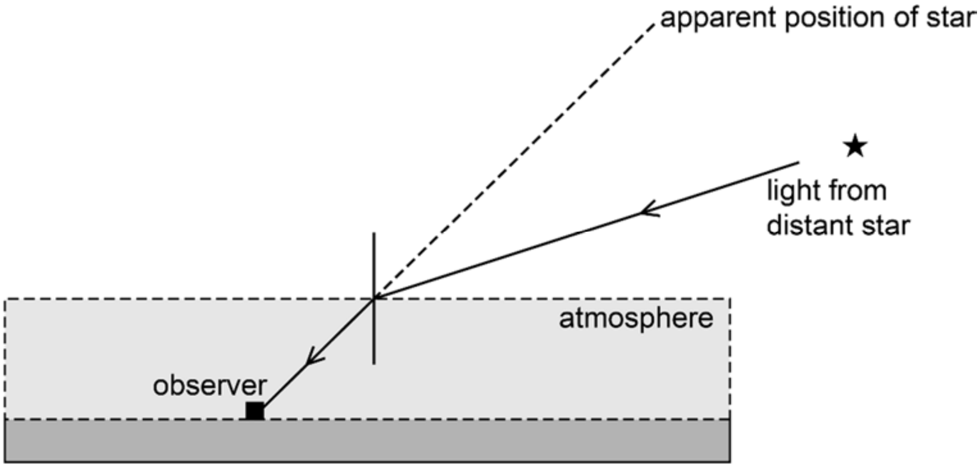
2 If the apparent position of the star is close to the observer's horizon then its actual position is higher in the sky.

3 If the star can be seen when the telescope is at 45° to the horizontal, then the actual position of the star in the sky must be at an altitude less than 45° to the horizontal.

Light slows down when it enters the atmosphere so if it is initially at an angle to the normal it will refract toward the normal.

If the star is vertically above the observer, then the light will travel along the normal and the apparent position of the star will be the same as its actual position, so statement 1) is correct.

If the star is between the horizon and the observer's vertical then the ray refracts in such a way that the star appears higher in the sky, as shown below (simplified and much exaggerated!).

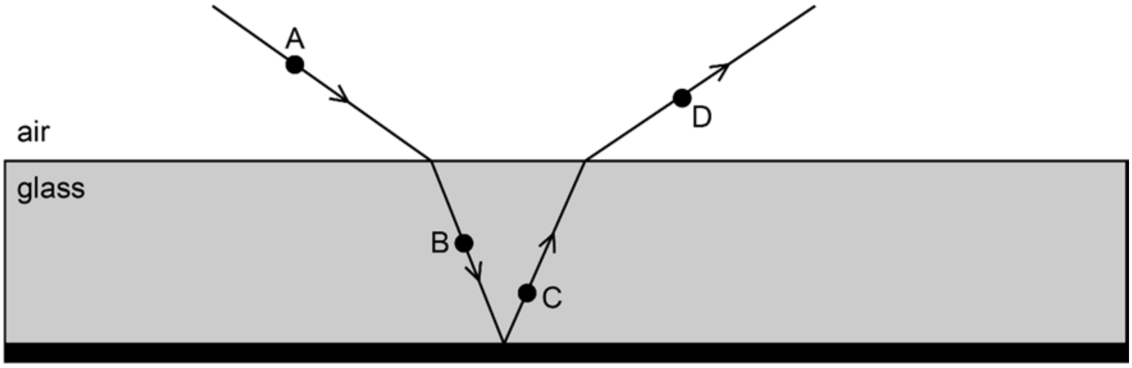


(Note: In reality the density of the atmosphere increases toward the surface so the waves slow down gradually and the refracted ray is actually curved.)

Statement 2) is therefore incorrect because the actual position would be lower in the sky, not higher.

Statement 3) is correct because refraction would make the apparent position higher than the actual position so the actual position is at an altitude less than 45°.

Mirrors are usually silvered on their back surface, so light reflecting from a mirror is also refracted as it enters and leaves the glass. The diagram below shows light of wavelength 500 nm reflecting from a mirror.



Which of the following statements is/are correct?

1 The speed of the wave is equal at A, B, C and D.

2 The wavelength at D is greater than the wavelength at B.

3 The frequency at A is equal to the frequency at B.

4 The wavelength at A is equal to the wavelength at D.

5 The frequency at D is greater than the frequency at C.

Frequency is unaffected by reflection or refraction so statement 3) is correct and statement 5) is incorrect.

The speed of light is lower in glass than in air so the speeds at A and D will be equal to one another and greater than the speeds at B and C, which are also equal to one another. Statement 1) is therefore incorrect.

Wavelength is reduced when the wave slows down so it is less at B and C than at A and D. Therefore statement 2) is correct.

Waves at A and D are both travelling through air so they have the same wavelength and statement 4) is correct.

Thus statements 2), 3) and 4) only are correct.

Know and understand the analogy of reflection and refraction of light with that of water waves

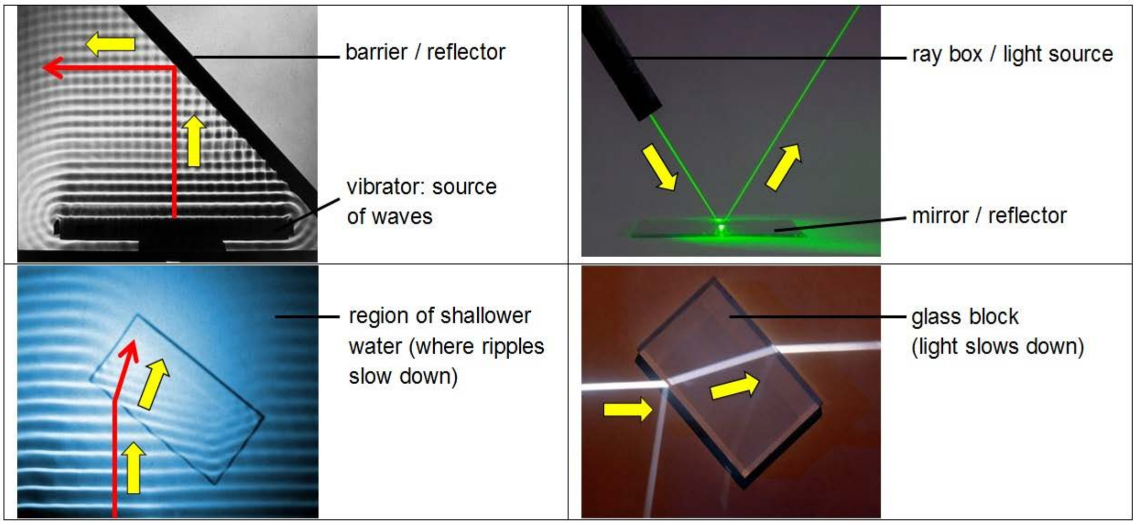
A ripple tank can be used to demonstrate how water waves reflect and refract.

- The ripple tank shows the behaviour of wave crests.

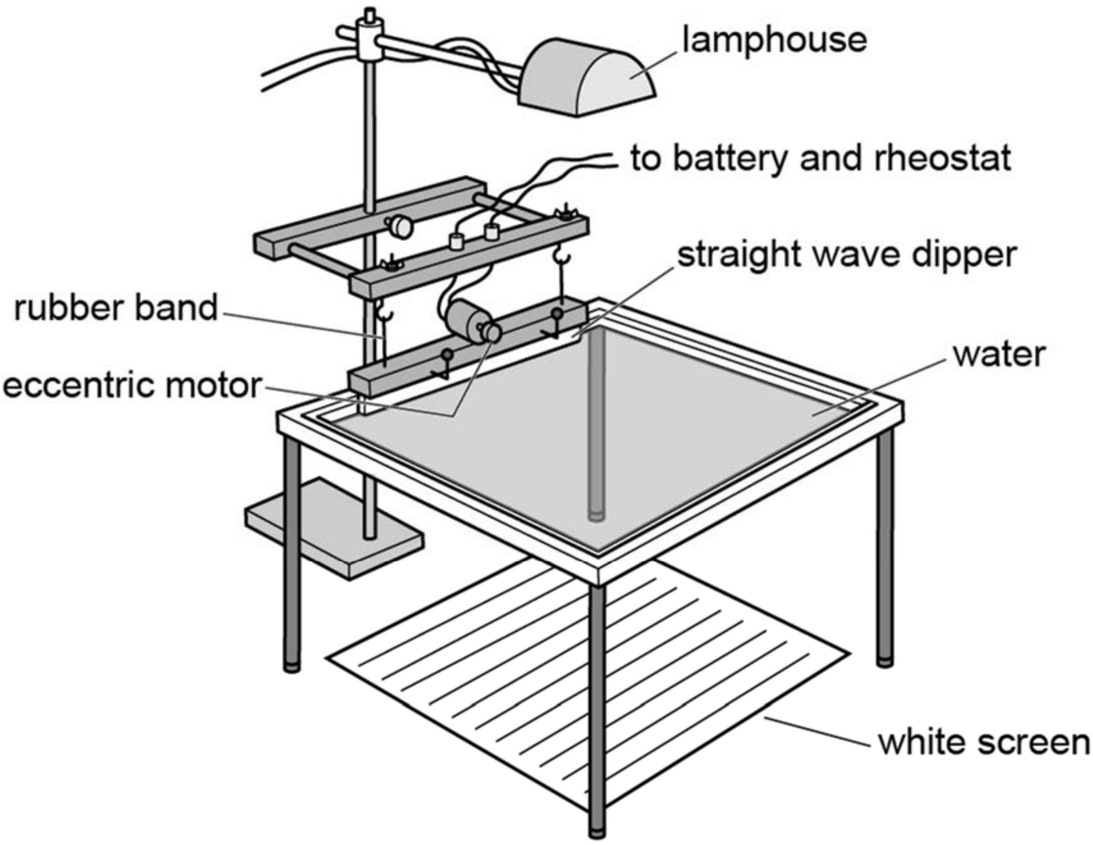
In a similar way a ray box can be used to show how light waves reflect and refract.

- Ray box experiments show the behaviour of rays.

Water waves and light waves reflect and refract in a similar way so we can use the behaviour of water waves to see how light waves behave and vice versa.



Here is a diagram showing how a ripple tank can be set up to display wave crests:



Light from the lamphouse casts shadows onto the screen. Dark lines, parallel to wave crests, are seen on the screen.

Differences between water waves and light waves

The analogy between water waves and light waves cannot be taken too far. There are important differences between the two types of wave:

- Light waves are electromagnetic and water waves are mechanical.
- Light can travel through a vacuum and water waves cannot.
- Light waves consist of vibrations of electric and magnetic fields and water waves consist of vibrations of particles.

Which of the following statements about the analogy between water waves and light is/are correct?

1 Water waves refract when they cross a boundary and their speed changes. Light waves refract when they move from water into glass so the speed of light in glass must differ from the speed of light in water.

2 When light waves strike a plane mirror the angle of reflection (measured from the normal) is equal to the angle of incidence (measured from the same normal). Therefore, when water waves strike a plane barrier the angle of reflection (measured from the normal) is equal to the angle of incidence (measured from the same normal).

3 The wavelength of water waves is not changed by reflection. Therefore, the wavelength of light waves will not be changed by reflection.

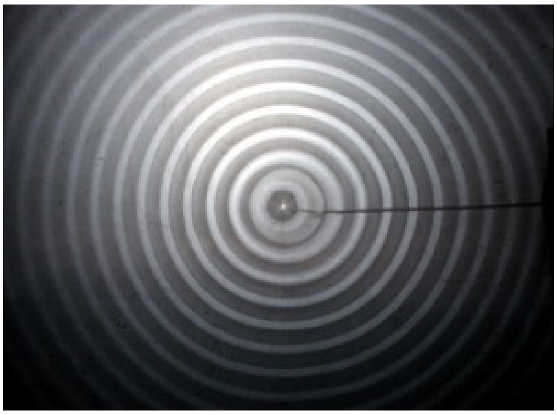
Refraction is a property of all types of wave and is caused by a change of speed at a boundary, so this statement is correct.

The law of reflection states that the angle of incidence is equal to the angle of reflection and is a property of all types of wave, so this statement is correct.

When a wave reflects it stays in the same medium and its wavelength is unchanged by reflection, so this statement is also correct.

Hence all three statements are correct.

The image below shows ripples spreading out from a point. The source of the waves is a small ball vibrating perpendicular to the surface and staying in contact with it. The water in the ripple tank has constant depth (i.e. the medium does not change as the waves move outwards).



Which of the following statements is correct?

1 The ball is vibrating with constant frequency.

2 The waves are travelling at a constant speed.

3 This pattern (in 3D) could be for light for a single wavelength leaving a point source in a vacuum.

4 If rays were to be drawn onto the diagram, they would all point directly out from the centre.

The separation of the wave crests is constant, so the wavelength is constant and this statement is correct.

Wave speed depends on the medium and here the medium (i.e. depth of water) does not change, so the speed is constant and this statement is correct.

The waves would spread out in the same way, with constant wavelength and speed because they remain in the same medium, the vacuum, so this statement is correct.

Rays indicate the direction in which the wave is moving so they are perpendicular to the wave crests and point radially outwards, so this statement is also correct.

All four statements are correct.

Know and understand the Doppler effect

When there is relative motion between a source of waves and an observer, the wavelength and frequency of the waves detected by the observer is different from the wavelength and frequency of the waves received when there is no relative motion.

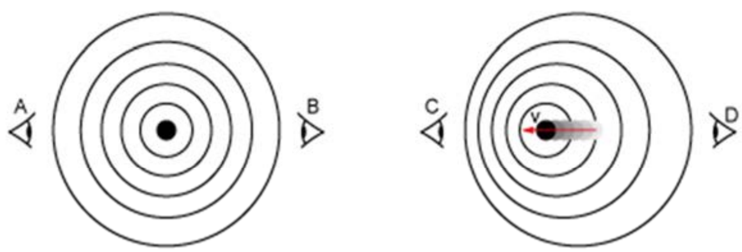
This change in wavelength and frequency is called the Doppler effect.

The effects are summarised below:

Relative motion	Effect on wavelength	Effect on frequency
Source and observer at rest relative to one another	No change	No change
Source and observer approaching one another	Shorter	Higher
Source and observer moving away from one another	Longer	Lower

The faster the observer approaches or recedes from the source, the greater the shift in frequency and wavelength.

The diagrams below show how wavelength is affected by the relative motion of the source of the waves.



Source and observer at rest relative to one another

Source and observer approaching one another

In the left-hand diagram, A and B are at rest with respect to the source. The waves spread out symmetrically in all directions so A and B receive waves of the same wavelength and frequency.

In the right-hand diagram, the source is moving toward C and away from D. Successive wave crests leave the source at equal time intervals but from different source positions, so they get ‘bunched up’ in front of the source (toward C) and 'stretched out' behind the source (toward D). C and D therefore detect Doppler shifts.

The change in wavelength also affects frequency. As the waves bunch together they arrive more frequently than when the source is stationary relative to the observer. When the waves are further apart their frequency drops.

Note that the Doppler effect depends on motion along the line from the source to the observer. The greater the speed along that line, the greater the shift in frequency and wavelength. Motion perpendicular to that line does not affect the received wavelength and frequency.

Compared to A and B:

- receives waves of shorter wavelength and higher frequency.
- receives waves of longer wavelength and lower frequency.

If we know the wavelength and frequency of a stationary source, then the Doppler effect can be used to work out if the source is approaching us or moving away from us and to calculate its speed.

Examples of the Doppler effect

- An ambulance siren has a higher pitch (frequency) as the ambulance approaches and then drops to a lower pitch (frequency) as the ambulance moves away.
- Police speed guns bounce pulses of radio waves from moving cars. The size of the Doppler shift on the reflected waves is used to calculate the speed of the car.
- Pulses of ultrasound can be reflected from blood cells and the Doppler shift on the reflected waves is used to measure and record the speed of blood flow.
- Astronomers can use the Doppler shift of known spectral lines from stars and galaxies to calculate their direction of motion and speed relative to the Earth.

A bee buzzes at a constant frequency of 180 Hz.

Copy and complete the table below to show the frequency of the sound heard when the bee moves in different ways.

Each entry should be one of: 180 Hz, <180 Hz, >180 Hz

Motion of bee relative to listener	**Frequency heard by listener**
Hovering stationary	
Flying away	
Flying toward	
Flying in a circle around the head	

Hovering stationary: there is no relative motion between the source and the observer so there is no Doppler shift and the heard frequency is the same as the emitted frequency of 180 Hz.

Flying away: the waves are stretched out as the source moves further from the listener so the wavelength increases and the frequency falls below 180 Hz.

Flying toward: the waves are bunched together as the source moves toward the listener so the wavelength decreases and the frequency rises above 180 Hz.

Circling: the distance between the bee and the listener is not changing so there is no relative motion along the line from the source to the observer. Therefore, the waves received by the listener are not Doppler shifted and the listener hears 180 Hz.

In the 1920s Hubble and Slipher measured the wavelengths of light in known spectral lines emitted by stars in distant galaxies. They discovered that:

the wavelengths were increased relative to the same spectral lines from sources at rest in the laboratory

the increase in wavelength was greater for more distant galaxies.

Assuming that these effects are caused by Doppler shifts, what two conclusions can be drawn about the motions of distant galaxies?

An increase in wavelength implies that the distance between the source (i.e. stars in the distant galaxies) and the observer (i.e. astronomers on Earth) is increasing. Therefore, distant galaxies must be moving away from us.

The larger the Doppler shift, the greater the relative velocity between the source and the observer. More distant galaxies have greater Doppler shifts so they must be moving away faster than nearby galaxies.

Photos: Alamy/Getty Images/Science Photo Library; Ripple tank photo courtesy of Harvard Natural Sciences Lecture Demonstrations

P6.3

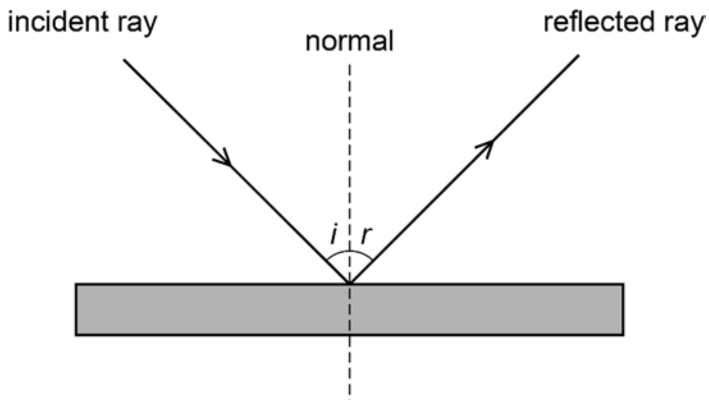
Optics:

a. Draw and interpret ray diagrams to describe reflection in plane mirrors.
b. Know and be able to apply: angle of incidence = angle of reflection
c. Draw and interpret ray diagrams for refraction at a planar boundary.
d. Know and be able to interpret angle of incidence and angle of refraction.
e. Know and understand the effect of refraction on wave direction (away from or towards the normal) and speed (increasing or decreasing).

Draw and interpret ray diagrams to describe reflection in plane mirrors

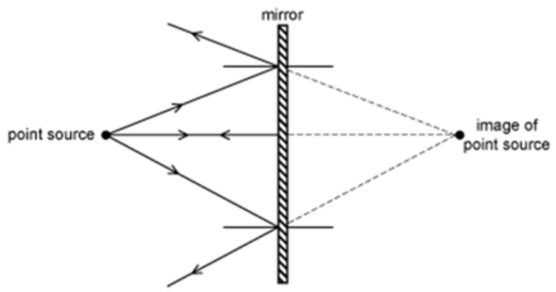
Know and be able to apply: angle of incidence = angle of reflection

- RAYS represent the direction of energy transfer in the wave.
- NORMALS are lines drawn perpendicular to the surface at a point where the ray hits the surface.
- ANGLE OF INCIDENCE (i) is the angle between the incident ray and the normal at the point where this ray hits the surface.
- ANGLE OF REFLECTION (r) is the angle between the reflected ray and the normal at the point where the ray hits angle of incidence = angle of reflection: **i = r**



Reflections in a plane mirror

Rays from a point source in front of a plane mirror spread out and reflect from different positions on the mirror. The law of reflection applies to each ray so they continue to spread out after reflection and appear to come from a point behind the mirror. This is the image of the point source.



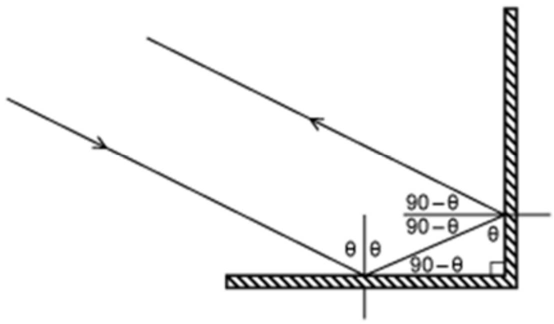
When an extended object is placed in front of the mirror, light from each point on the object reflects from the mirror.

An image of the extended object is formed behind the mirror.

The image in a plane mirror is formed at the same distance behind the mirror as the object is in front of it.

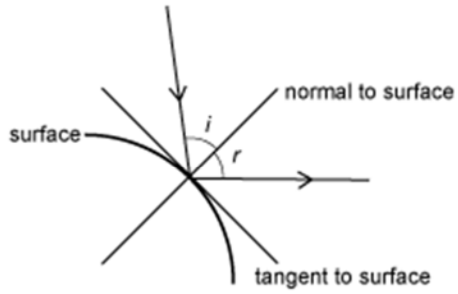
Multiple reflections

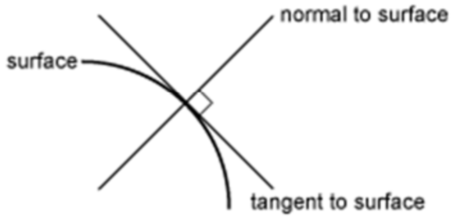
When a ray reflects from more than one surface, e.g. in a corner reflector (as shown below), simple trigonometry can be used to find the incident angle at the second reflector.



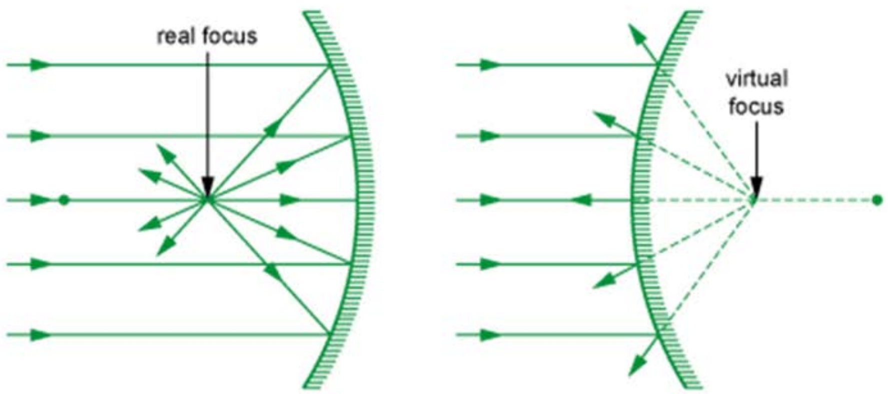
Reflection from a curved surface

The law of reflection applies whatever the shape of the surface but the normal must be perpendicular to the surface at the points where the rays hit it. On a curved surface the normal is perpendicular to the tangent to the surface:

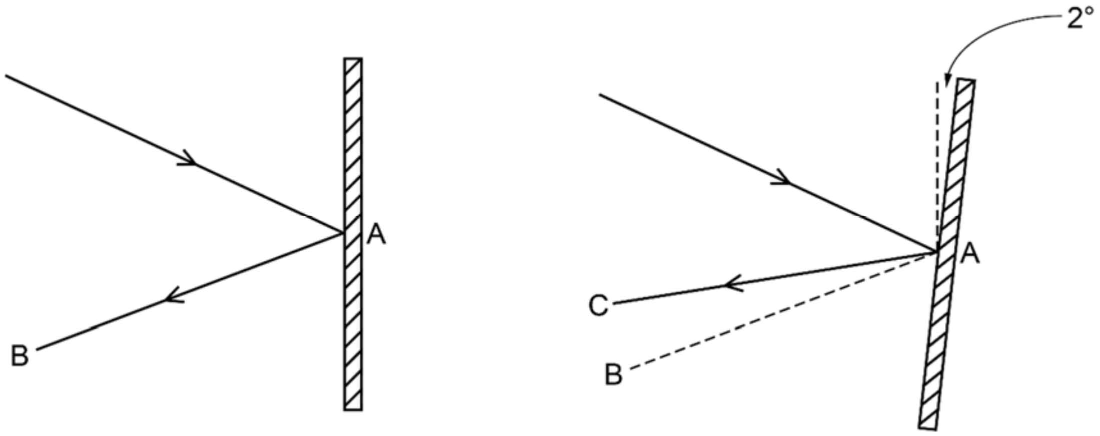




The diagram below shows how parallel rays reflect from curved mirrors:

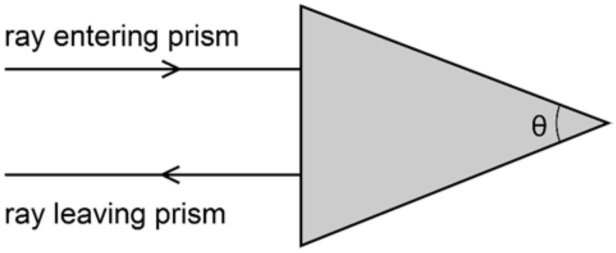


In an optics experiment a ray of light from a ray box is reflected from a plane mirror and leaves the mirror along line AB. The student accidentally disturbs the mirror so that it turns through 2° about point A. The reflected ray now travels along line AC. What is the angle between AB and AC?

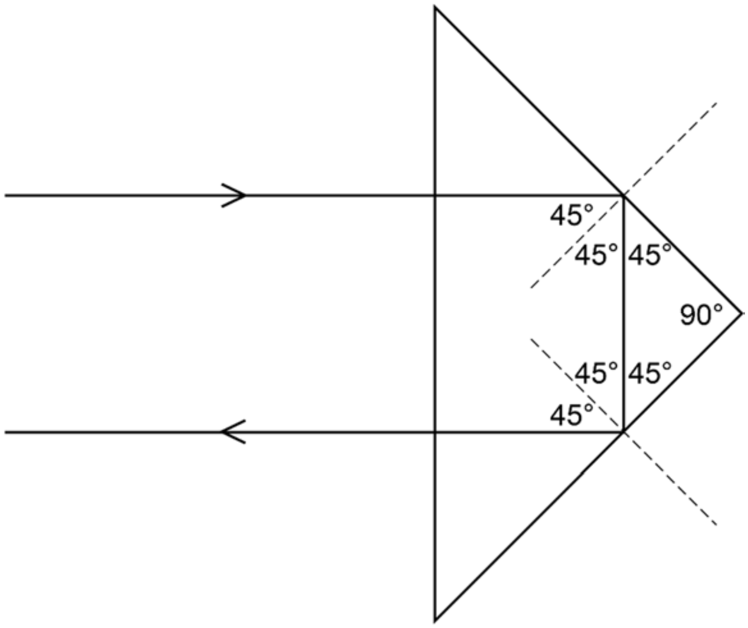


Consider the normal to the mirror at A. This moves with the mirror so it rotates clockwise by 2°. The incident ray direction does not change so the incident angle changes by 2°. Since the reflected angle is equal to the incident angle, this will also change in the same way (i.e. both decrease as in the diagram or increase if the mirror rotates in the other direction) by 2°. The angle between the incident and reflected angles is the sum of incident and reflected angles, so this changes by 2 + 2 = 4°. The incident ray direction is unchanged so the reflected ray turns by 4°. This is the angle between AB and AC.

Binoculars use triangular glass prisms to reflect light back in the direction from which it came. Light entering along a normal to one face of the prism reflects symmetrically off two internal faces of the prism and returns along a line parallel to the incident ray. What is the angle θ of the prism opposite to the face at which the light enters?



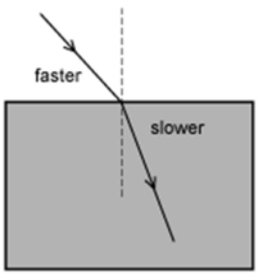
The ray must turn through 180° after two reflections and it travels symmetrically through the prism so it must turn through 90° on each reflection. The incident angle at both faces must therefore be 45°. The angles in the small triangle ABC below are therefore 45° at the base and 90° at the peak.

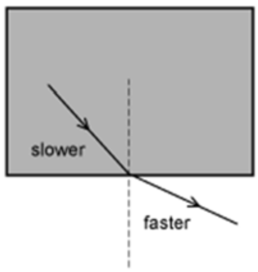


Draw and interpret ray diagrams for refraction at a planar boundary

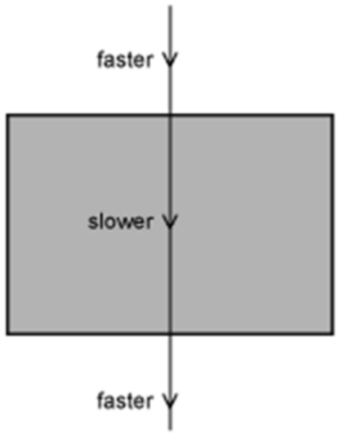
When light crosses a boundary between two different transparent media in which the light travels at different speeds it refracts. If the light is not travelling parallel to the normal then the wave direction changes as it crosses the boundary.

- If the wave slows down, it refracts towards the normal (e.g. air to glass).
- If the wave speeds up it refracts away from the normal (e.g. glass to air).





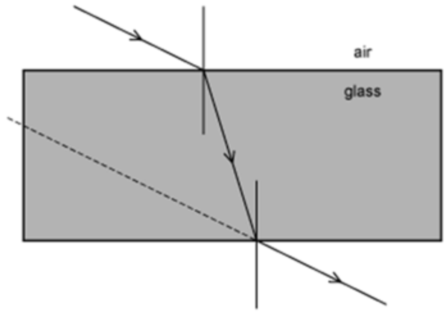
If light is travelling along the normal, it continues in the same direction (but still changes speed).



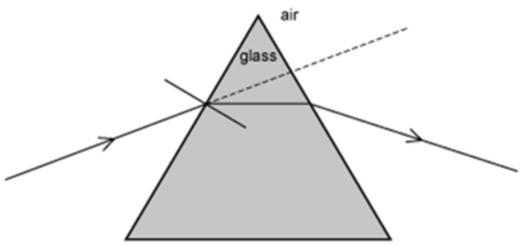
Multiple boundaries

When light travels through a glass prism it refracts at two planar surfaces, when it enters and when it leaves the prism.

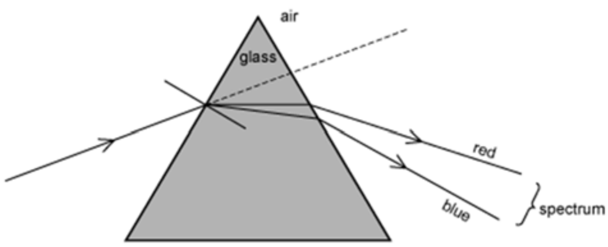
For a rectangular prism the emerging ray is parallel to the original incident ray. This is because the change of speed at both boundaries has the same ratio and the boundaries are parallel to one another:



For a triangular prism, the emerging ray has been deviated:



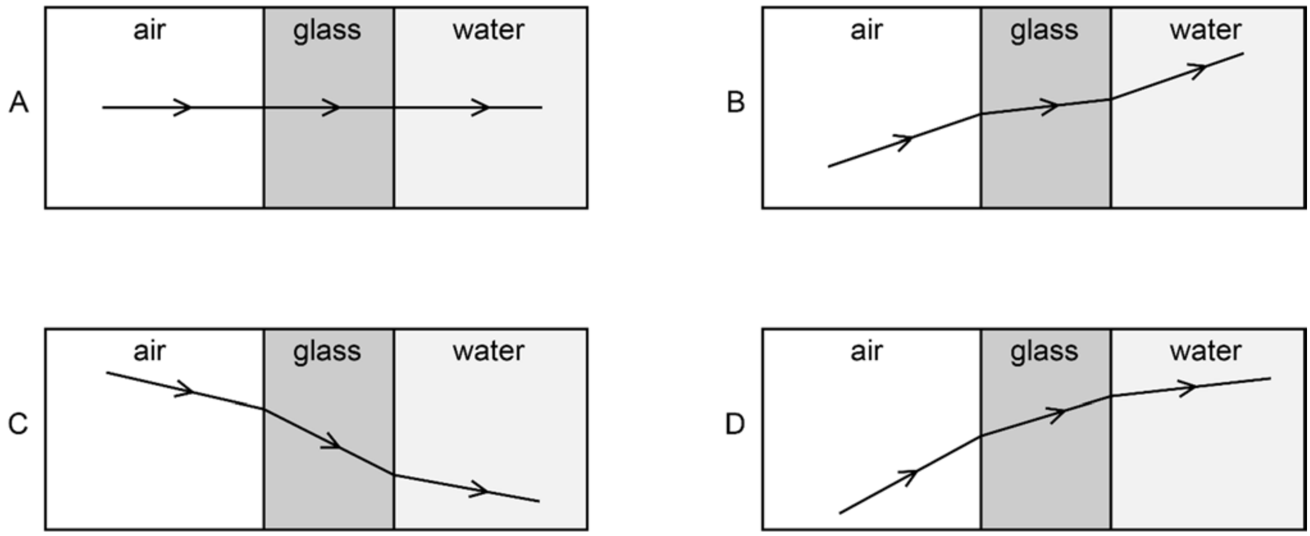
Different wavelengths of visible light travel at different speeds in glass so they are deviated by different amounts. If a ray of white light enters a triangular prism each wavelength deviates by a different amount so the emerging light is spread into a spectrum. Shorter wavelengths (the blue end of the spectrum) slows down more than longer wavelengths so blue light is deviated more than red light.



The speed of light in air is $3.0 \times 10^8\ ms^{-1}$, the speed of light in water is $2.3 \times 10^8\ ms^{-1}$, and the speed of light in glass is

$2.0 \times 10^8\ ms^{-1}$.

Which one or more of the diagrams below could show the path of a ray of light as it passes from air through glass and into water?



When light crosses a boundary, the direction of refraction depends on whether it increases or decreases in speed.

Going from air to glass to water involves a decrease at the first boundary and an increase at the second boundary. This causes the light to refract towards the normal at the first boundary and away from the normal at the second boundary. This is shown in B.

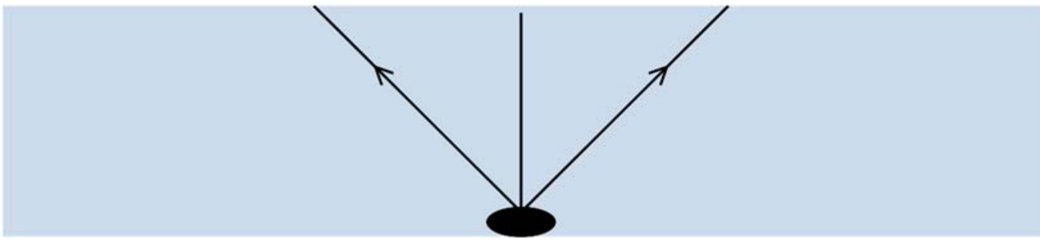
If the light is travelling along the normal its direction does not change. This is shown in A.

In C the refraction at the air/glass boundary is in the wrong direction.

In D the refraction at the glass/water boundary is in the wrong direction.

Hence the correct answer is A and B only.

The diagram below shows light rays leaving a point on a pebble at the bottom of a pool of water.



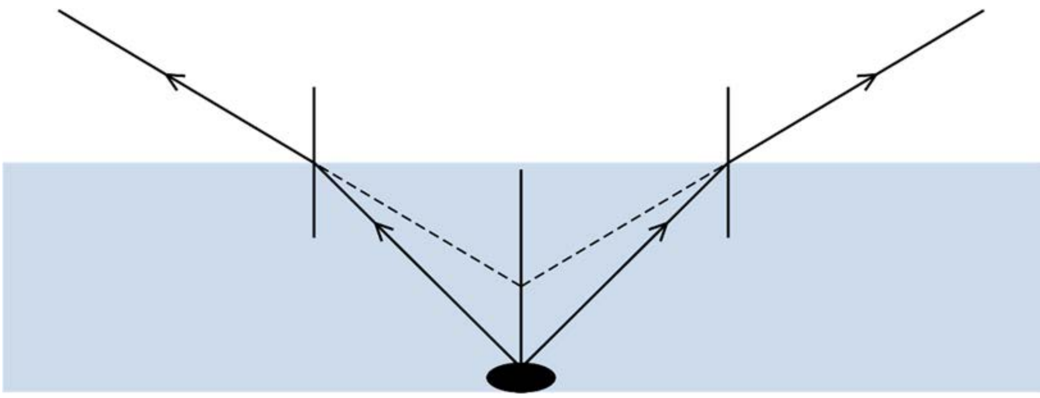
Which statement below is correct?

The rays will diverge even more when they enter the air and will appear to come from a point below the pebble.

The rays will diverge even more when they enter the air and will appear to come from a point above the pebble.

The rays will diverge less when they enter the air and will appear to come from a point below the pebble.

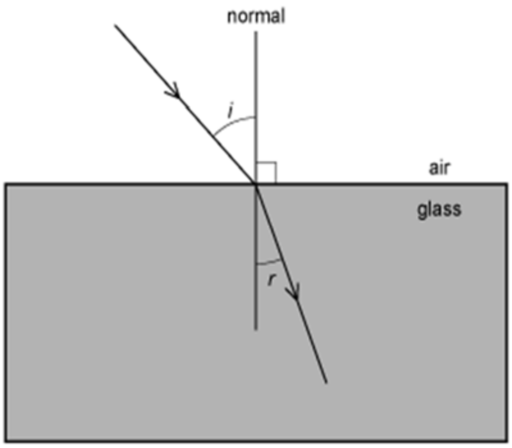
The speed of light in air is greater than the speed of light in water so the rays refract away from the normal and diverge more. When the refracted rays are traced back (dotted lines in the diagram below) they appear to come from a point above the pebble. This makes the pond appear to be shallower than it actually is. Therefore the correct statement is 2).



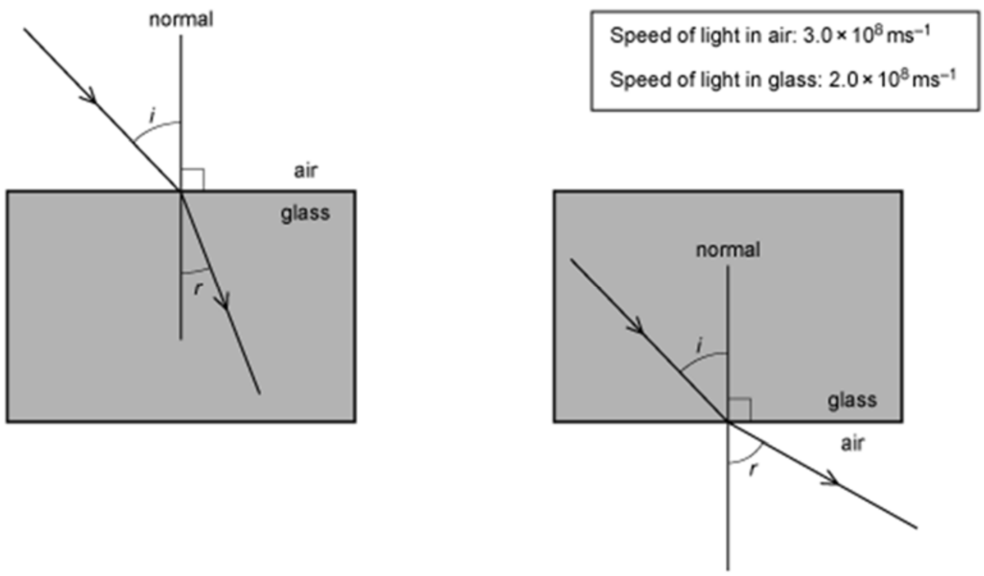
Know and be able to interpret angle of incidence and angle of refraction Ray diagrams for refraction:

- RAYS represent the direction of energy transfer in the wave.
- NORMALS are lines drawn perpendicular to the surface at a point where the ray hits the surface.
- ANGLE OF INCIDENCE (i) is the angle between the incident ray and the normal at the point where this ray hits the surface.
- ANGLE OF REFRACTION (r) is the angle between the refracted ray and the normal at the point where the ray hits

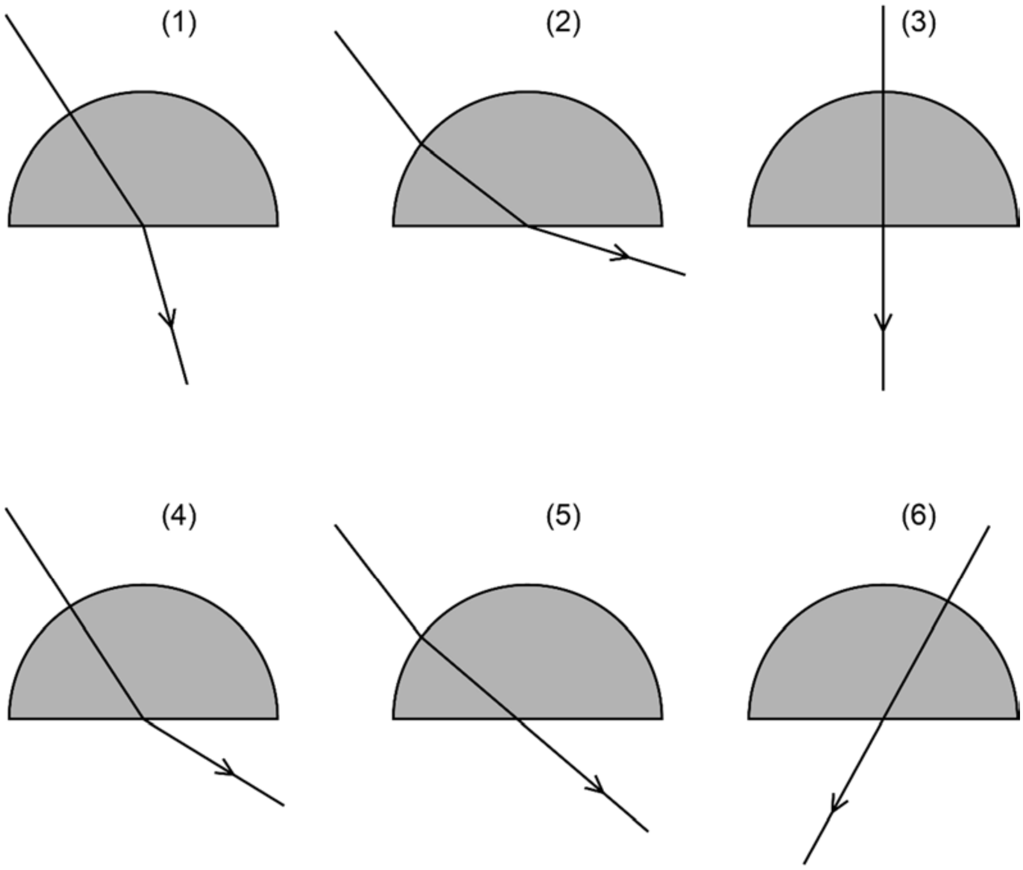
the surface.



The relationship between the angle of incidence and the angle of refraction depends on the speed of light on either side of the boundary. If the light is moving from a medium in which the speed of light is higher to a medium in which it is lower the ray refracts towards the normal. If it is moving from a medium in which the speed of light is lower to a medium in which the speed of light is higher it refracts away from the normal. This is shown below:



Which one or more of the following diagrams could show a light ray passing from air through a semi-circular glass block? Rays strike the flat face at the centre of the semi-circle.



The speed of light in glass is lower than in air, so rays should refract towards the normal on entering the glass and away from the normal on leaving it. The only exception to this is when the ray travels along a normal to the surface, in which case its direction does not change. Rays refract in the correct way at each surface in 2), 3) and 4) only.

In 1) the incident ray in air travels along the normal to the curved surface (i.e. radially) so its direction does not change on entering the block. However, the refracted ray is drawn closer to the normal as it leaves the block. It should refract away from the normal so this diagram is incorrect.

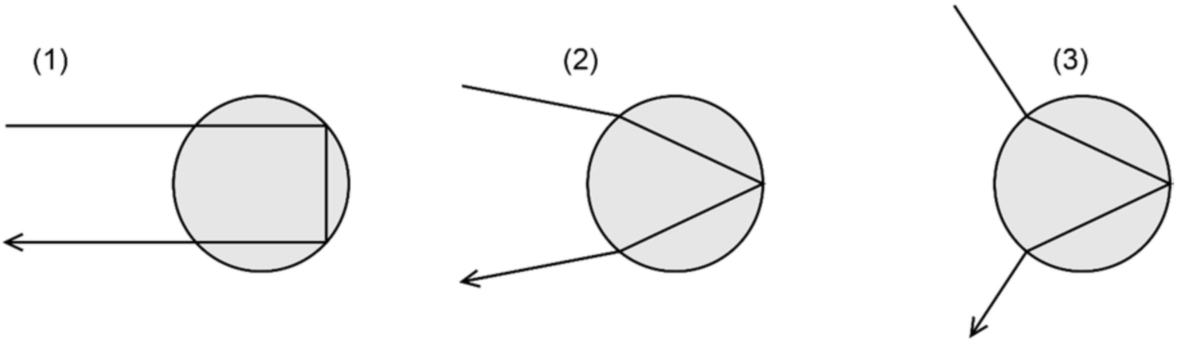
In 5) the refraction is in the correct direction on entering the block but the ray should refract away from the normal as it leaves the block. Instead it is shown with no change in direction, so this diagram is incorrect.

In 6) the ray enters the block along a normal so its direction does not change. However, it is shown leaving the block with no change in direction. It should refract away from the normal. So this diagram is incorrect.

Only diagrams 2), 3) and 4) could show the path of a light ray through a semi-circular block.

Rainbows are formed by light rays from the Sun that enter water droplets in the atmosphere and reflect inside the droplet before refracting out and reaching the observer's eye.

Which one or more of the diagrams below could show the path of such a light ray as it enters and leaves a raindrop?



The speed of light in water is less than in air so rays should refract towards the normal on entering the drop and away from the normal on leaving it. This occurs only in 2).

In 1) the rays should change direction on entering and leaving the drop, but they do not so this diagram is incorrect.

In 3) the rays refract in the wrong direction both on entering and on leaving the drop. So this diagram is incorrect.

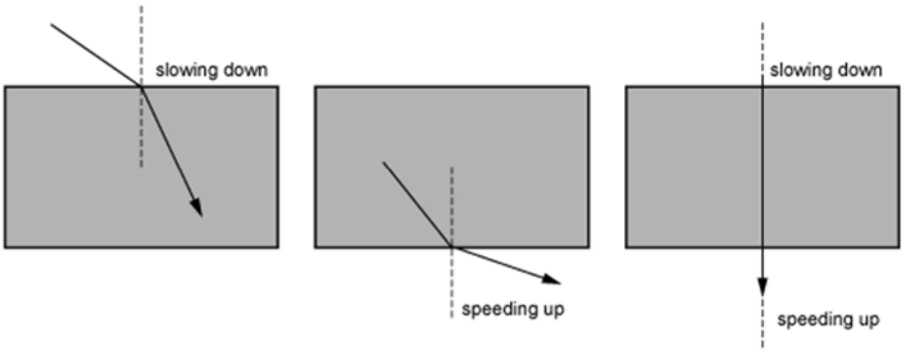
The only correct diagram is 2).

Know and understand the effect of refraction on wave direction (away from or towards the normal) and speed (increasing or decreasing)

When light crosses a boundary between two different transparent media in which the speed of light is different they refract.

If the rays are not parallel to the normal then the wave direction changes as they cross the boundary. If the rays are along the normal they continue in the same direction (but still change speed).

- If the wave slows down, it refracts towards the normal (e.g. air to glass).
- If the wave speeds up it refracts away from the normal (e.g. glass to air).



The greater the change in speed, the larger the change in direction of the waves.

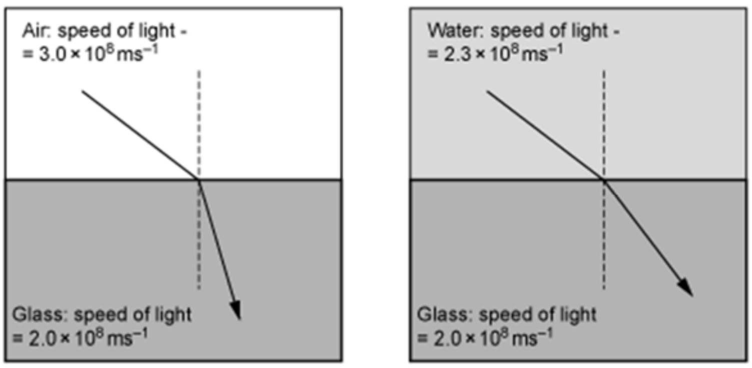
The speed of light in air, water and glass is:

Air: 3.0×10^8 ms^{-1}

Water: 2.3×10^8 ms^{-1}

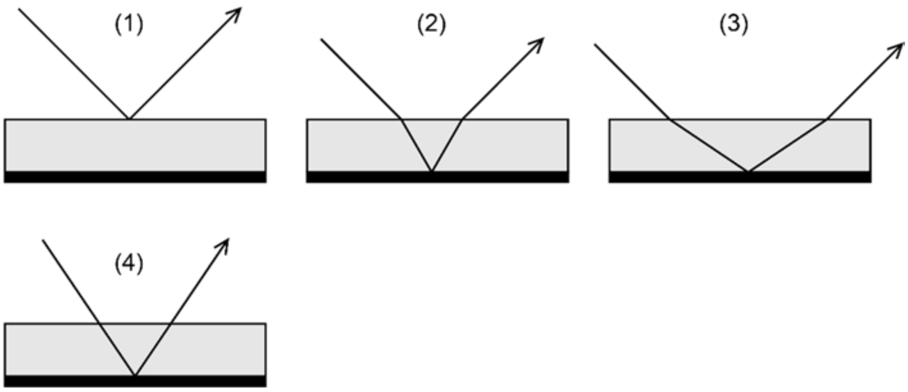
Glass: 2.0×10^8 ms^{-1}

A ray of light at an air/glass boundary will refract more than a ray of light at the same incident angle to an air/water boundary:



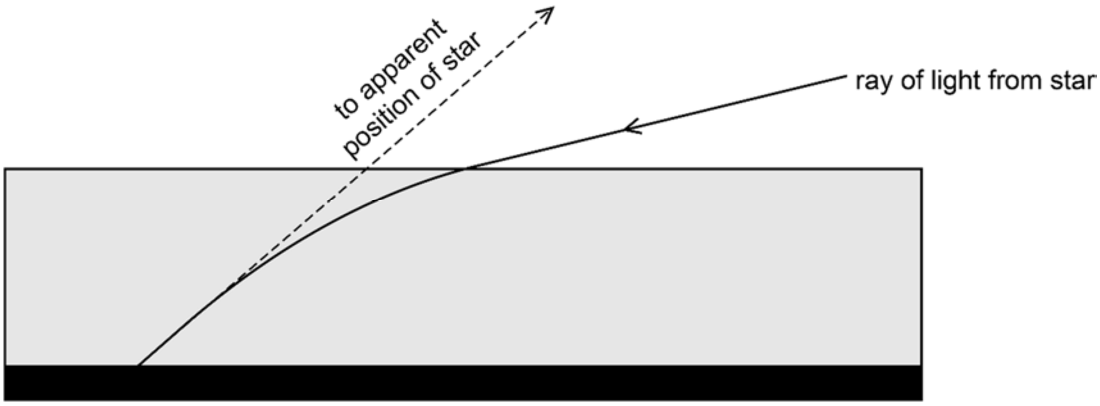
A mirror is made from a layer of glass with one surface silvered.

Which of the diagrams below shows how a ray of light reflects from such a mirror? (Ignore partial reflection.)



Glass is transparent so whilst it is true that there will be some partial reflection at the front surface, as in 1), most of the light will enter the glass and reflect from the back silvered surface. As it passes from air into the glass the light slows down and refracts towards the normal. This is shown in 2), so that is the correct diagram. In 3) the direction of refraction is away from the normal, so that is incorrect. In 4) the rays do not change direction at the air/glass boundaries so that is also incorrect.

Rays of light entering the atmosphere change direction, gradually following a curved path as they approach the surface. This makes the apparent positions of stars higher than they actually are. The effect is shown (exaggerated) in the diagram below.



What conclusions can be drawn from the fact that the light changes direction as it enters the atmosphere and the fact that its path is then curved towards the surface?

Light refracts as it enters the atmosphere, so its speed must change.

The direction of refraction is towards the normal at the surface of the atmosphere so light slows down in the atmosphere.

The amount of refraction continually increases so the speed of light must be getting progressively

P6.4

Sound waves:

a. Understand the production of sound waves by a vibrating source.
b. Understand the need for a medium.
c. Understand qualitatively the relation of loudness to amplitude and pitch to frequency.
d. Know and understand longitudinal waves.
e. Understand that reflection causes echoes.
f. Recall that the range of human hearing is 20 Hz to 20 kHz.
g. Know and understand ultrasound and its uses (sonar and medical scanning).

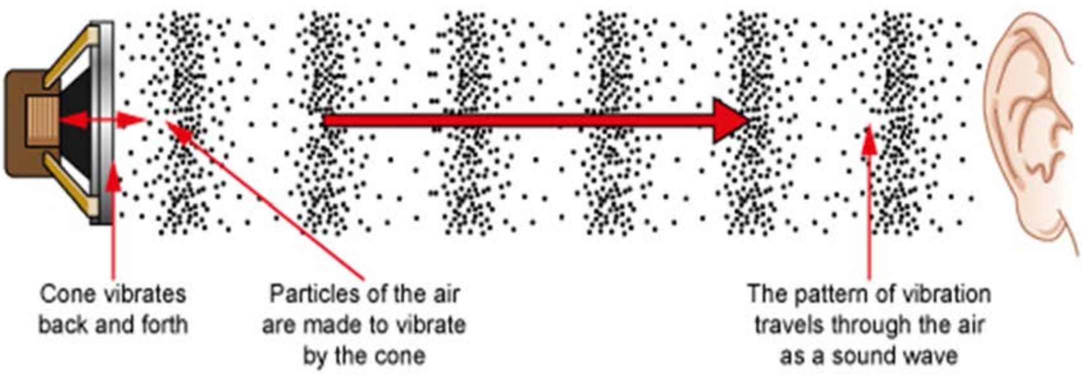
Understand the production of sound waves by a vibrating source

Sound waves are produced by a vibrating source.

The vibrating source causes the surrounding medium to vibrate, and this pattern of vibrations travels away from the source as sound waves.

E.g.

- The cone of the loudspeaker vibrates back and forth moving the air in front of it
- The pattern of vibrations travels through the air as sound waves



Relationship between vibrations of the source and sound produced:

- The sound waves have the same frequency (or frequencies) as the vibrations of the source.
- The amplitude of the sound waves depends on the amplitude of the vibrations of the source.
- The speed of the sound waves is determined by the medium through which they travel and NOT by the source.

Examples of sound sources:

Vibrating strings – e.g. guitar string, piano wire.

Vibrating surfaces – e.g. loudspeaker cone, drum skin, sounding box of a guitar.

Vibrating air columns – e.g. in musical instruments such as a recorder, saxophone, organ pipes.

Vibrating vocal chords – human speech.

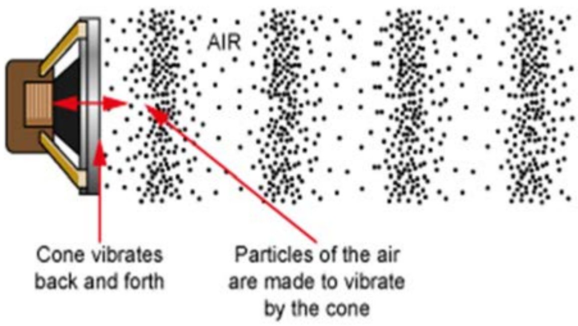
Production of ultrasound:

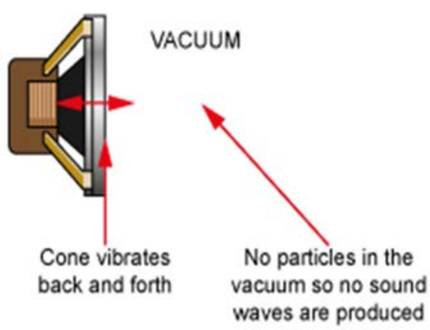
Ultrasound is produced by vibrating sources in a similar way to audible sounds but at frequencies above the limit of human hearing (> 20 kHZ).

Understand the need for a medium

Sound waves consist of vibrations of material particles so they can only travel through a material medium.

Sound cannot travel through a vacuum because there are no particles present to vibrate.





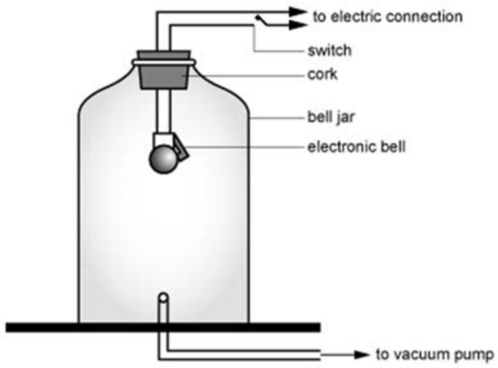
Different media

Sound waves consist of compressions and rarefactions of the medium so they can travel through solids, liquids and gases.

The speed of sound depends on the nature of the medium.

Bell and bell jar experiment

A simple experiment to show that sound needs a medium is shown below:



An electric bell is suspended inside a bell jar. The bell is switched on and can be heard clearly. A vacuum pump is then used to remove the air from inside the bell jar. The sound fades and when a vacuum is formed the bell can no longer be heard, even though it is clear that it is still being struck.

(The fact that the bell remains visible shows that light CAN travel through a vacuum.)

Through which of the following media can sound waves travel?

1 Air

2 Glass

3 Water

4 Mercury

5 Vacuum

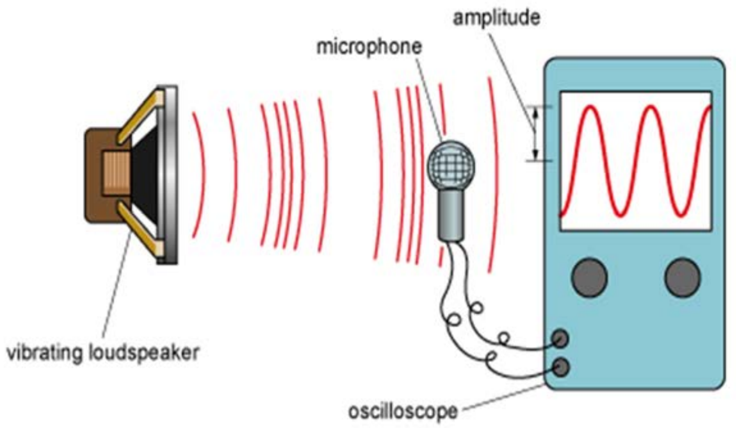
Sound waves consist of vibrations of material particles, so they can travel through any type of material medium. As the wave passes the particles get pushed closer together and further apart creating compressions and rarefactions in the medium. There are no particles in a vacuum so sound waves cannot travel through a vacuum. The answer is all of the media except 5).

Understand qualitatively the relation of loudness to amplitude and pitch to frequency

The loudness of a sound depends on the amplitude of the sound waves. The greater the amplitude, the louder the sound.

The pitch of a sound depends on the frequency of the sound waves. The higher the frequency, the higher the pitch.

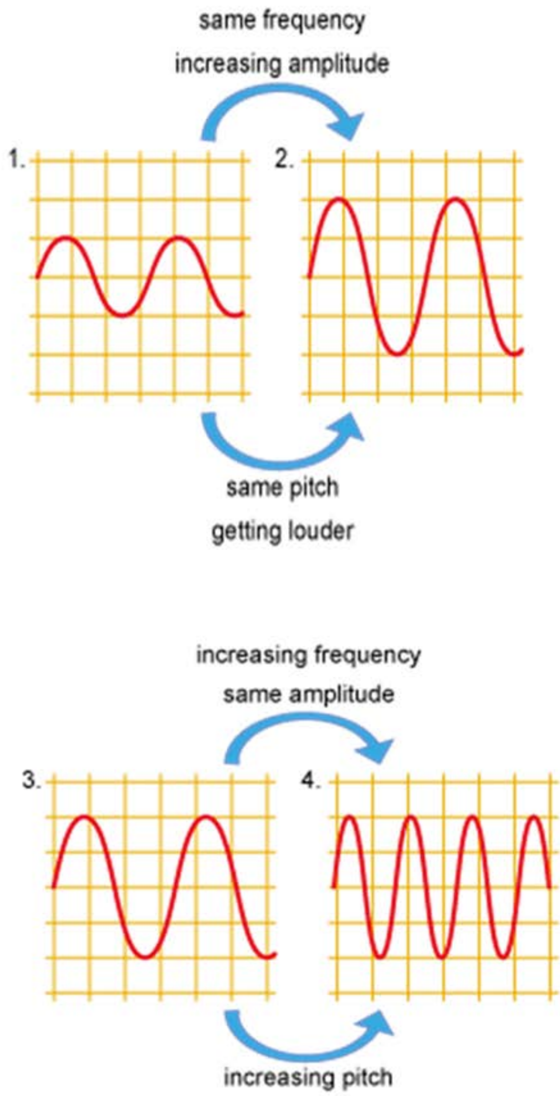
A microphone connected to an oscilloscope can be used to display the oscillations of sound waves.



The display can be used to compare the amplitude and frequency of different sounds.

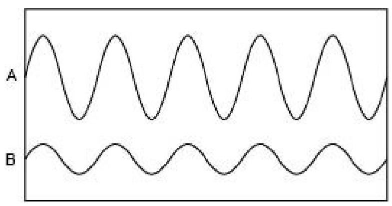
- The higher the peaks, the greater the amplitude / louder the sound.
- The closer together the peaks, the shorter the period of the sound and the higher its frequency / pitch.
- The effect on the wavelength of the sound can also be inferred from the period. If the period of the sound is shorter (frequency higher), then so is the wavelength.

Here are the oscilloscope displays for four sounds shown with the same display settings:



Two microphones located at different places are connected to the same oscilloscope. They both detect sounds and the displays are shown below.

How does the sound detected by microphone A compare to the sound detected by microphone B?



The sound at A is louder but has the same pitch (or frequency). Loudness is related to amplitude, and pitch is related to the time per oscillation; the shorter the time, the higher the pitch. On an oscilloscope screen, time is on the horizontal axis. In this case, both traces have the same period, but trace A has greater amplitude.

Know and understand longitudinal waves

A sound wave consists of oscillating particles in a medium (e.g. in air).

Particles in the medium vibrate along the same line as the wave is travelling, so sound is a longitudinal wave.

Longitudinal waves: the vibration direction is parallel to the wave direction.

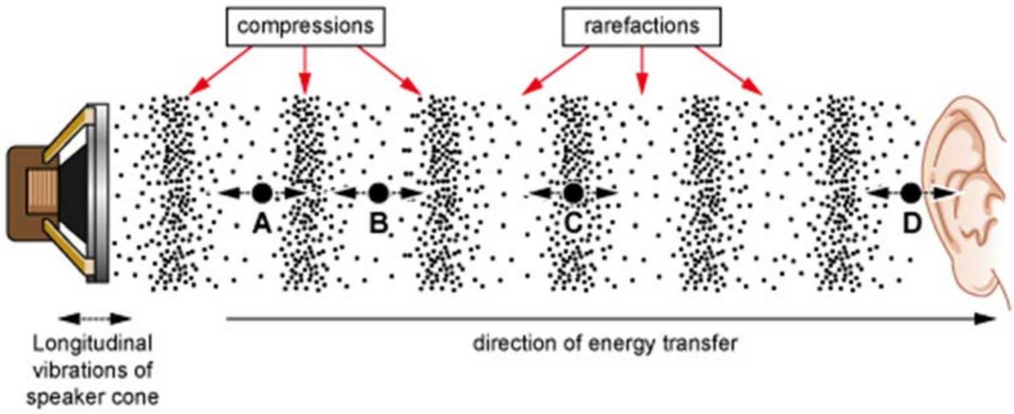
Key features of all longitudinal waves:

- particles in the waves vibrate along the same line as the direction of wave travel
- the waves consist of regions of compression (particles pushed closer together) and rarefaction (particles pulled further apart)
- the distance between two adjacent compressions (or two adjacent rarefactions) is equal to one wavelength.

Particle vibration in a sound wave

The diagram below shows the positions of four particles, A, B, C, and D as a sound wave passes. As the speaker cone moves forward it compresses the air in front of it, creating a region of higher pressure, a compression. As it moves back it rarefies the air, creating a region of lower pressure, a rarefaction. This pattern of compressions and rarefactions travels away from the speaker as sound waves.

All particles in the wave oscillate parallel to the direction of energy transfer. Their vibrations are shown as horizontal dashed arrows. This makes sound a LONGITUDINAL wave.



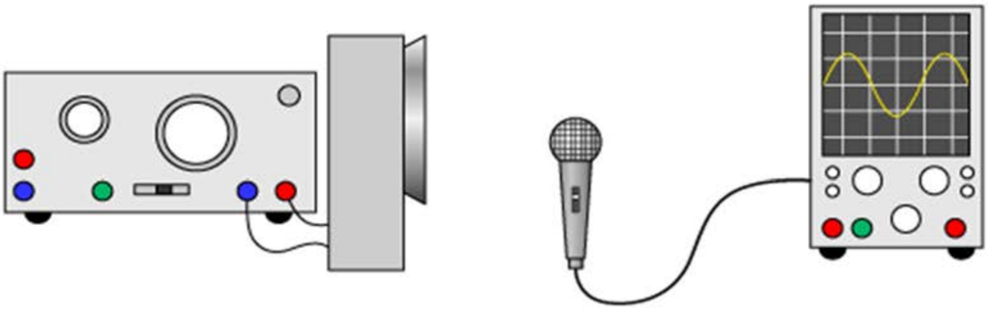
The distance between two adjacent compressions (or between two adjacent rarefactions) is equal to one wavelength.

Detecting sounds

When sound arrives at a detector (e.g. a microphone or the human ear), the sequence of compressions and rarefactions causes the pressure at the detector to vary. This exerts a varying force on the detector and this is what is detected (e.g. the ear drum is moved by this force).

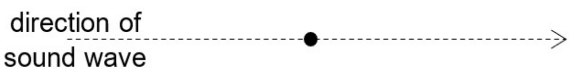
Displaying sound waves

When sound is detected by a microphone, the varying pressure is used to create a signal of varying voltage. This can be connected to an oscilloscope so that the trace on the oscilloscope screen varies in the same way as the pressure in the sound wave.



Note: The trace looks like a transverse wave but it is really a graph of voltage against time. The sound it represents is a longitudinal wave.

A tiny particle of dust is suspended in still air. A sound wave passes in the direction shown:

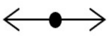


Which of the statements describes how the dust particle moves as the sound passes? (Ignore the effects of gravity.)

1 It is unaffected by the sound wave.

2 It moves steadily to the right.

3 It vibrates about its original position in this direction:



4 It vibrates about its original position in this direction:

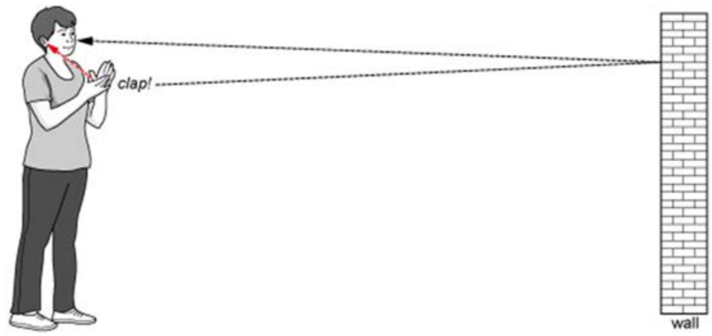
Sound is a longitudinal wave, so the particle vibrations in the air are parallel to the direction of the wave. The dust particle is moved by the air so it also vibrates parallel to the wave direction. Statement 3) is correct.

Understand that reflection causes echoes

Sound waves obey the law of reflection.

An echo is a sound heard after sound waves reflect from one or more surfaces.

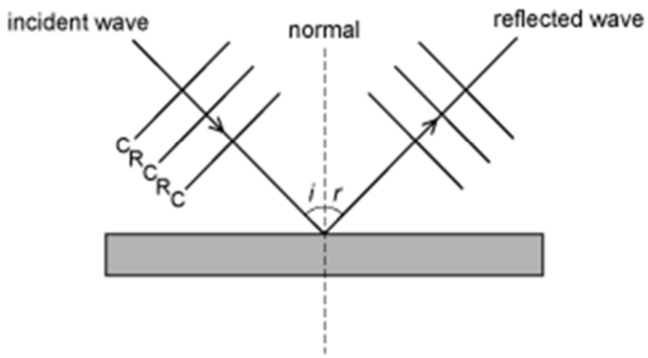
When a person hears sound directly from the source and also hears the echo of that sound, the echo is heard later. This is because the direct and reflected waves travel at the same speed through the air but the distance travelled by the reflected waves is greater.



In the example above, a girl stands at a distance from a vertical wall. When she claps her hands, sound travels out in all directions. Sound that travels directly to her ear (red dashed line) is heard almost immediately but sound that travels to the wall and reflects back (black dashed lines) arrives later. This is heard as an echo of the original sound.

Reflection of sound waves

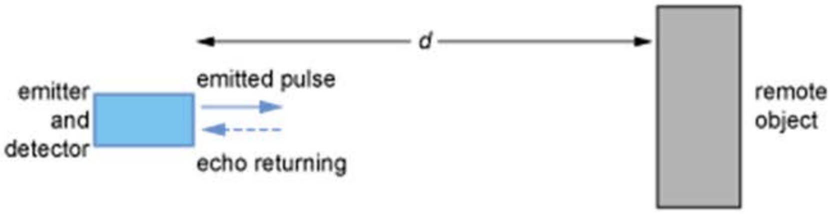
The angle between the incidence direction and the normal is equal to the angle between the reflected direction and the normal. Sound waves consist of compressions (C) and rarefactions (R) perpendicular to the wave direction as shown below.



The wavelength of the reflected waves is the same as the wavelength of the incident waves.

Using reflection to measure distances

Distances to remote objects can be determined by reflecting pulses of sound (or ultrasound) from them and measuring the time between emitting the pulse and detecting the echo of the pulse.



If the time for the pulse to travel out and return is t then the distance can be found from:

$$d = \frac{vt}{2}$$

where v is the speed of sound in the medium.

The factor of 2 is because the sound travels to the remote object and back, a distance of $2d$.

This principle is used in:

- sonar (measuring distances to objects under water)
- ultrasound scanning and imaging (e.g. prenatal scanning)
- echolocation (bats).

Recall that the range of human hearing is 20 Hz to 20 kHz

- The human ear cannot detect sounds with frequencies lower than about 20 Hz or higher than about 20 kHz (20000 Hz).
- Sounds with frequencies between about 20 Hz and 20 kHz are detected by the human ear and can be heard.

The range of human hearing is taken to be 20 Hz to 20 kHz.

In reality, the range of frequencies that can be heard by different people varies from individual to individual and the high frequency limit reduces significantly for older people.

Frequency and wavelength

There is a corresponding range of wavelengths of sounds that are audible to humans. We can calculate the wavelength range using the equation: v = fλ rearranged to $\lambda = \frac{v}{f}$

For f = 20 Hz and v = 330 ms^{-1} λ = 330 / 20 = 16.5 m

For f = 20 kHz and v = 330 ms^{-1} λ = 330 / 20 000 = 0.0165 m = 1.65 cm

Note that the low frequency limit corresponds to the long wavelength limit and the high frequency limit corresponds to the short wavelength limit.

Range of hearing in other animals

Some animals can detect sounds outside the range of human hearing, usually at higher frequencies than 20 kHz.

Dogs can hear sounds up to 60 kHz and bats can hear sounds up to 90 kHz.

Ultrasound

Sound with a frequency higher than 20 kHz is called ultrasound. This is above the upper limit for of hearing for most humans.

The table shows the typical range of frequencies that can be heard by dogs, bats and elephants.

Animal	Hearing range / Hz
dog	70 – 45 000
bat	2000 – 100 000
elephant	10 – 12 000

Which of the following statements is/are correct?

1 The highest pitched sound that can be heard by a human cannot be heard by a dog.

2 The lowest pitched sound that can be heard by an elephant can also be heard by a human.

3 A sound of frequency 500 Hz can be heard by humans, dogs, bats and elephants.

4 Of these animals, elephants can hear sounds with the shortest wavelength.

The highest pitched sound that can be heard by a dog is 45 000 Hz but the highest that can be heard by a human is 20 000 Hz, so statement 1) is incorrect.

The lowest pitched sound that can be heard by a human is 20 Hz but elephants can hear lower pitched sounds down to 10 Hz, so statement 2) is incorrect.

The lowest frequency that can be heard by a bat is 2000 Hz, so bats cannot hear 500 Hz and statement 3) is incorrect.

Elephants can hear the sounds with the lowest frequency and longest wavelength, but bats can hear sounds with the highest frequency and shortest wavelength so statement 4) is also incorrect.

None of the statements is correct.

Know and understand ultrasound and its uses (sonar and medical scanning)

Ultrasound:

- consists of longitudinal waves
- consists of particle vibrations in a material medium, so it cannot travel through a vacuum
- consists of sound waves with frequencies above 20 kHz
- typically cannot be heard by humans because ultrasound frequencies are higher than the upper limit of human hearing for most humans.

Ultrasound pulses can be used to measure the distance to a remote object. This is important for:

SONAR (sound navigation and ranging)

A pulse of ultrasound (a 'ping') is emitted and the time t between emission and the detection of the reflected pulse is measured. The distance to the remote object is calculated using the equation:

$$d = \frac{vt}{2}$$

The factor of $\frac{1}{2}$ is there because the pulse travels to the object and back again (i.e. a total distance of 2d).

Medical scanning

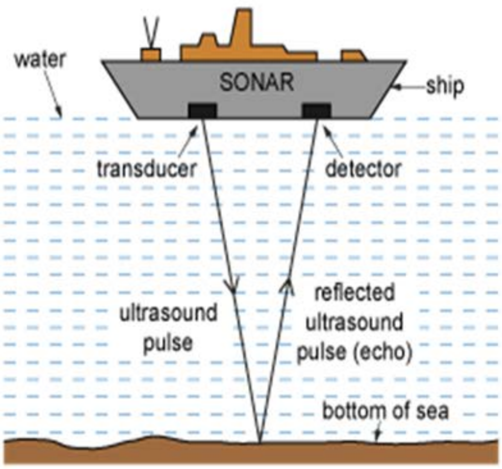
An ultrasound transmitter and detector is used to send pulses of ultrasound into the body (e.g. in prenatal scanning) and some of the ultrasound reflects from each internal boundary. This results in several returning pulses from which an image of the internal structures (e.g. an unborn baby) can be constructed.

Advantages of ultrasound:

- it can be used to image soft tissue
- it does not damage human tissue.

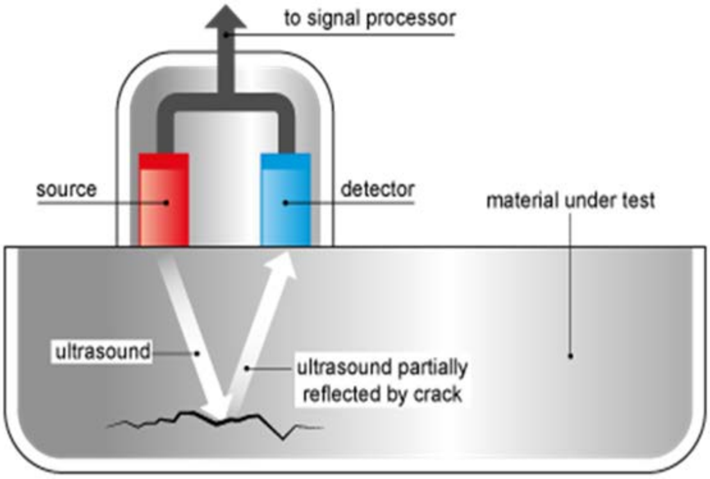
Measuring the depth of the sea

This diagram shows how a ship's SONAR can be used to measure the depth of the sea. SONAR can also be used to locate other vessels and shoals of fish.



Crack detection

Ultrasound can be used to detect cracks and flaws inside a solid material (e.g. a pipeline). Partial reflections occur at the crack. The time for these to return is used to determine the depth of the crack. This technique is shown in the diagram below.



Prenatal scanning

The principle of prenatal scanning is identical to that of crack detection. Each change of tissue type inside the body partially reflects the incident ultrasound pulse. This results in a series of returning pulses that can be used to locate the depths of each boundary. If an array of transmitters and detectors is used then a detailed image can be constructed.

A boat uses sonar to measure the depth of the sea. A pulse of ultrasound is emitted downwards from the bottom of the boat and its reflection is detected 0.40 s later.

How far is the seabed beneath the boat? (The speed of ultrasound waves in water is 1500 ms^{-1}.)

The distance travelled by the ultrasound in this time is given by distance = speed × time = 1500 × 0.40 = 600 m

However, the pulse has travelled to the seabed and back, so the distance to the seabed is half of this:

distance to seabed = 600 / 2 = 300 m

Photos: Science Photo Library

P6.5

Electromagnetic spectrum:

a. Know and understand the nature and properties of electromagnetic waves (they are transverse waves and travel at the speed of light in a vacuum).
b. Recall the component parts of the spectrum (radio waves, microwaves, IR, visible light, UV, X-rays, gamma).
c. Understand the distinction of the component parts by different wavelengths and/or frequencies.
d. Recall the order of the component parts by wavelength and/or frequency.
e. Understand applications and hazards of the component parts of the electromagnetic spectrum.

Know and understand the nature and properties of electromagnetic waves (they are transverse waves and travel at the speed of light in a vacuum)

Electromagnetic waves:

- include radio waves, microwaves, infra-red, visible light, ultraviolet, X-rays and gamma-rays
- transfer energy
- are all transverse waves
- do not need a material medium
- can travel through a vacuum
- all travel at the speed of light in a vacuum
- travel at lower speeds in other media.

The nature of EM waves

EM waves do not consist of vibrating particles, they consist of vibrating electric and magnetic fields. This is why they can travel through a vacuum.

Sources of EM waves

Anything that causes electric charges to vibrate will emit EM waves. For example, a warm body contains vibrating atoms which contain charged particles, so all bodies emit a spectrum of EM radiation (because all bodies are above the absolute zero of temperature). The spectrum of radiation emitted by a warm or hot body depends on the temperature of the body. The hotter the body, the more radiation and the more high-frequency radiation. This is why a warm stone emits plenty of infra-red radiation but no visible light, whereas a piece of metal heated in a Bunsen flame glows red hot. If heated further, the metal emits all of the visible spectrum and becomes white hot. Radiation from the surface of the Sun is mainly infra-red but includes visible light and ultraviolet. Its outer atmosphere even emits X-rays.

Absorption of EM waves

EM waves transfer energy from a source to an absorber. When EM waves are absorbed, they transfer energy to the matter that absorbs them and this can cause:

- heating – e.g. in a radiant electric heater or a microwave oven
- electrons in the surface to vibrate at the frequency of the wave – e.g. in radio reception
- ionisation – e.g. when X-rays or gamma-rays are absorbed by human tissues.

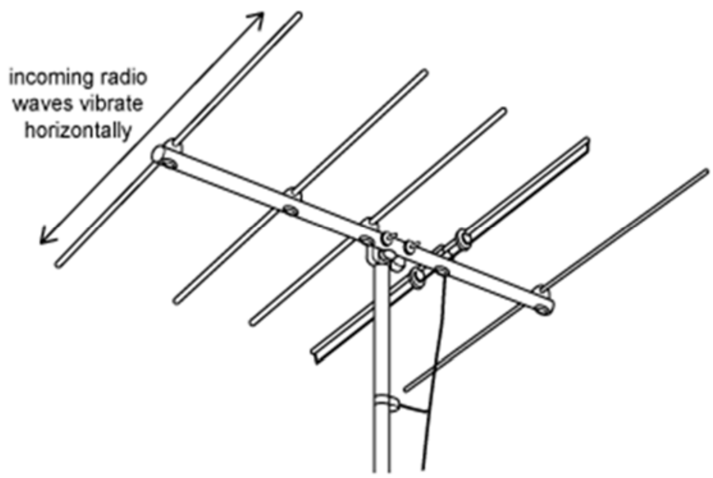
EM waves in a vacuum

The fact that we can see the Sun and stars is evidence that visible light can travel through a vacuum. Different types of telescope can detect radiation from space in all parts of the EM spectrum showing that all EM waves can travel through a vacuum. This can also be demonstrated in the school laboratory by evacuating a bell jar. Anything placed inside the bell jar remains visible.

All EM waves travel at the same speed in a vacuum; this is called the speed of light. The speed of light in a vacuum is 3.0×10^8 ms^{-1}.

EM waves are transverse

The image below shows a typical TV antenna. It is designed to detect radio waves that carry a TV signal. Note that the rods are all aligned. This is because the TV signal consists of EM waves that vibrate in the horizontal plane. When the waves meet the rods they cause electrons in the rods to vibrate at the same frequency as the wave and this electrical signal is used to create the TV pictures.



Which of the following statements about X-rays and ultraviolet radiation is/are correct?

1 X-rays are transverse waves but ultraviolet radiation is longitudinal.

2 They both travel at the speed of light in a vacuum.

3 They have different wavelengths and different frequencies.

4 X-rays are invisible but ultraviolet radiation is visible.

X-rays and ultraviolet (not to be confused with ultrasound) are both part of the electromagnetic spectrum and so are both transverse waves. Statement 1) is incorrect. And both travel at the speed of light in a vacuum. Statement 2) is correct.

X-rays are at the short wavelength high frequency end of the spectrum whilst ultraviolet radiation has a longer wavelength and lower frequency. Statement 3) is correct.

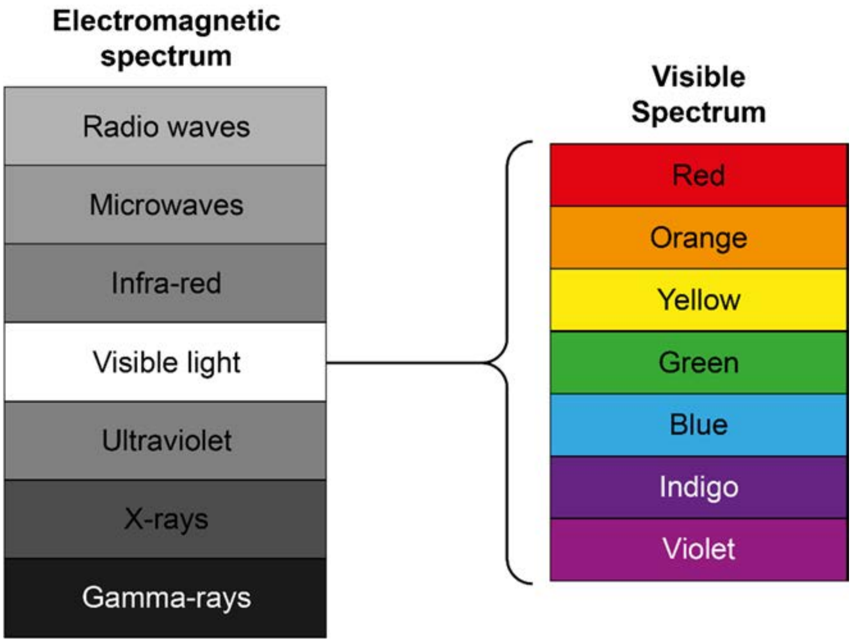
Both have higher frequencies than visible light so neither are visible. Statement 4) is incorrect.

So statements 2) and 3) are correct and statements 1) and 4) are incorrect.

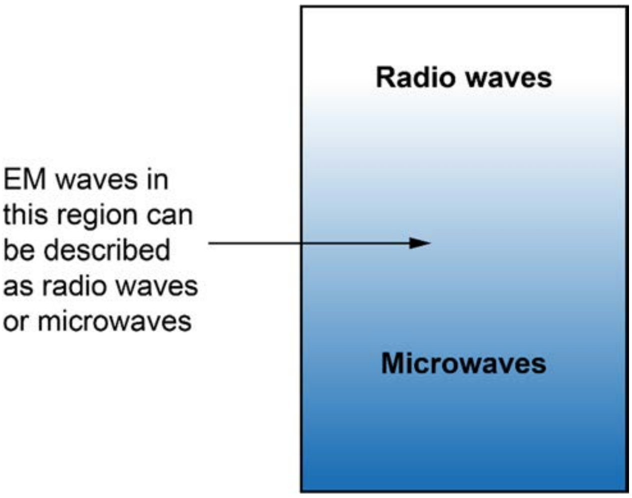
Recall the component parts of the spectrum (radio waves, microwaves, IR, visible light, UV, X-rays, gamma)

Electromagnetic waves form a spectrum that is divided into seven different regions.

Visible light is one small part of the electromagnetic spectrum and it is subdivided into another seven regions.



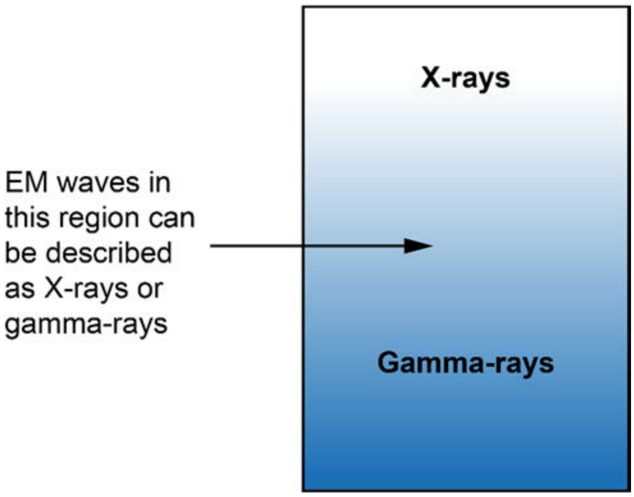
The divisions between each part of the electromagnetic spectrum are not sharp so that the regions overlap. For example, EM waves close to the boundary between microwaves and radio waves might be classified as either.



Microwaves and radio waves with the same wavelength are identical.

X-rays and gamma-rays

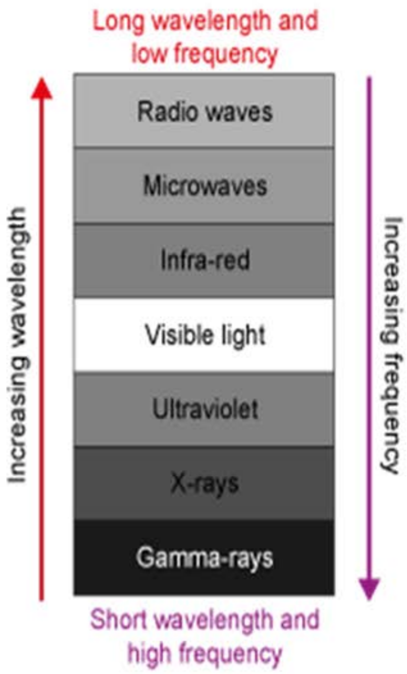
In the overlapping region between X-rays and gamma-rays the naming usually depends on how the waves have been produced. X-rays are usually produced when fast moving electrons crash into a metal target and stop. Gamma-rays are usually produced by radioactive decay. X-rays and gamma-rays with the same wavelength are identical.



X-rays and gamma-rays with the same wavelength are identical.

Understand the distinction of the component parts by different wavelengths and/or frequencies Recall the order of the component parts by wavelength and/or frequency

The different regions of the electromagnetic spectrum are distinguished by their wavelength and frequency. Radio waves have the longest wavelength and lowest frequency and gamma-rays have the shortest wavelength and highest frequency.



Wavelengths and frequencies in the EM spectrum

All electromagnetic waves travel at the same speed in a vacuum, the speed of light, c.

Frequency and wavelength are therefore linked by the equation:

$$c = f\lambda$$

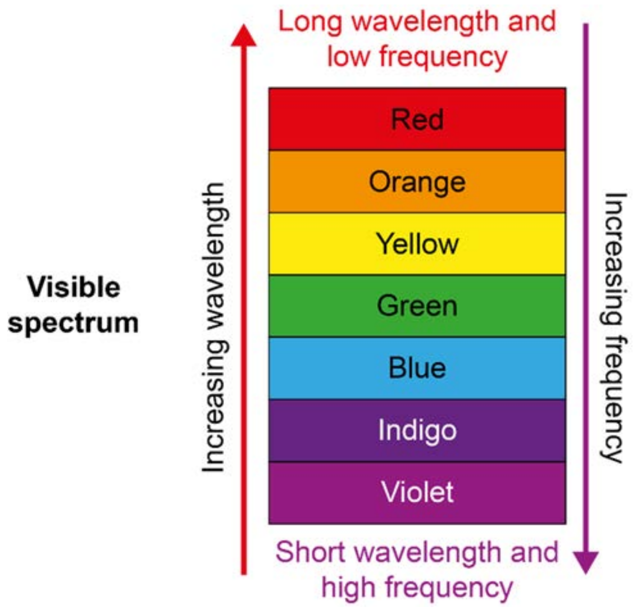
with c = 3.0×10^8 ms^{-1} (a constant value)

- Frequency is inversely proportional to wavelength: $f = \frac{c}{\lambda} = \frac{constant}{\lambda}$
- Wavelength is inversely proportional to frequency: $\lambda = \frac{c}{f} = \frac{constant}{f}$

These relationships explain why EM waves with longer wavelengths have lower frequencies.

Wavelengths and frequencies in the visible spectrum

Red light is at the long wavelength, low frequency end of the visible spectrum. Violet light is at the short wavelength, high frequency end of the visible spectrum.



Frequency, wavelength and speed of EM waves

The frequency of an electromagnetic wave is determined by the frequency of vibration of electrons in the source. However, the speed of the wave is determined by the medium through which the waves travel. All EM waves travel at the speed of light, c = 3.00×10^8 ms^{-1} in a vacuum, but at lower speeds in other media.

When an electromagnetic wave crosses a boundary from one medium to another its frequency stays constant but its speed changes. Since wavelength, frequency and speed are all related by the equation $v = f\lambda$, the wavelength must change in the same way as the speed (i.e. both reduce or both increase).

As the universe expands, the wavelength of electromagnetic waves is stretched. Which of the following changes could occur as a result of this expansion?

1 Gamma-rays become microwave radiation.

2 Blue light becomes ultraviolet radiation.

3 X-rays become infra-red radiation.

4 Microwaves become visible light.

1) and 3) are possible because the second named radiation has a longer wavelength than the first. In fact, the gamma rays that filled the early universe have been stretched to become microwaves today.

Understand applications and hazards of the component parts of the electromagnetic spectrum

Electromagnetic waves have a wide range of applications but are also potentially hazardous. The table below gives some important applications and hazards associated with different parts of the EM spectrum.

Region of EM spectrum	Applications	Hazards
Radio	Communications: radio and TV. Radar systems. Radio astronomy.	Only hazardous if extremely intense.
Microwave	Satellite and space communications. Radar systems. Mobile phones. Wifi systems. Microwave cookers.	Tissues can be damaged if too much microwave radiation is absorbed by living tissues so protective (reflective) suits must be worn if working near a powerful transmitter. They can also cause cataracts in the eye.
Infra-red	Radiant heaters. TV/DVD remote controls. Heat seeking missiles. Sensors on security lights. Optical fibre communications. Night sights. Thermal imaging. Weather satellites (IR photography).	Cell damage: burns.
Visible	Sight. Astronomical and terrestrial telescopes. Microscopes. Illumination. Optical fibre communications. LASERs.	Looking at an intense source of light can damage the retina of the eye. This is why you should not look directly at the Sun and never direct a telescope or binoculars towards it!
Ultraviolet	Causes some things to fluoresce – (e.g. washing powders in clothes so are often used at clubs and parties). Security marking. Can kill microbes so can be used to sterilise medical equipment. Insect control (UV attracts insects).	Can damage the retina of the eye. Can cause sunburn and skin cancer.

X-ray	X-ray images, CAT scans. Airport security. X-ray crystallography (investigating the structure of crystalline materials using X-rays). Detecting art forgeries. Xray telescopes in astronomy.	X-rays are a form of ionising radiation that can damage molecules, causing cell damage and various types of cancer.
Gamma-ray	Radiotherapy to kill cancer cells. Radioactive tracers. Food sterilisation. Locating cracks in pipes and turbines. Gamma-ray telescopes in astronomy.	Gamma-rays are a form of ionising radiation that can damage molecules causing cell damage and various types of cancer. They can also cause cells to mutate and if they affect sex cells or a developing embryo, the effects are seen in the next generation.

Useful properties

The applications of EM waves are linked to the properties of EM waves and the way they transmit and deliver energy. High frequency short wavelength EM waves such as X-rays and gamma-rays deliver energy in such a way that they can ionise atoms in the material that absorbs them. Lower frequency, longer wavelength EM waves cannot cause ionisation but do cause heating, making particles in the absorbing material vibrate more strongly. EM waves used for communication have extremely high frequencies and travel at the speed of light; this allows a great deal of information to be transmitted in a short time.

The table below links particular properties to particular types of application.

Properties of EM waves	Applications based on this property
Can transmit energy	Radiant heating, solar power, LASER cutting
Can travel through a vacuum	Satellite communications, astronomical telescopes
Can be reflected	Car headlamps, radio telescopes
Can pass through the Earth's atmosphere (radio)	Long distance communications, radio astronomy
Have a very high frequency	Can carry a great deal of analogue or digital information
Travel at the speed of light	Can transmit information very quickly
Can cause ionisation (X-rays and gamma-rays)	Can be used to prevent electrostatic charges building up (e.g. in an operating theatre)
Can penetrate smoke and dust (IR)	Locating people in smoke filled room
Emitted by all hot bodies (IR)	Can be detected and used to form thermal images (e.g. for weather satellites or night sights)
Can pass through human tissue (X-rays and gamma-rays)	X-ray imaging, irradiation of cancer cells, radioactive tracers
Can be guided inside an optical fibre (IR, visible)	Telephone systems, cable TV, endoscopes
Can be refracted when passing through a transparent medium (e.g. visible)	Focusing light, optical instruments
Can be made into extremely intense LASER beams	Cutting metal, reading bar codes, DVDs and CDs

Reducing risks with EM radiation

Type of EM radiation	How to reduce risk
Radio	Do not go too close to a powerful transmitter
Microwaves	Limit direct exposure, especially to eyes. If working with strong transmitters, maintain a safe distance and wear a reflective suit. Limit the time for which you use a mobile phone.
Infra-red	Maintain a safe distance from intense sources. Wear reflective clothing.
Visible	Do not look directly at bright sources. Never point a telescope directly at the Sun (unless it has a solar filter attached to the objective). Wear sunglasses. Place a shield in front of an intense source. Do not direct LASERs or LASER pens into the eye.
UV	Do not look directly at a UV lamp. Limit skin exposure to UV – e.g. sunbeds. Use a protective sun cream if outside on a sunny day. Wear sunglasses that block UV.
X-rays	Limit your exposure to X-rays, e.g. increase distance from source, wear protective clothing and/or stand behind a screen. Use lead shielding between an X-ray source and your body. Doctors should ensure that the potential benefit of X-ray treatment always outweighs the potential risk.
Gamma-rays	Maximise your distance from the source and minimise the time that you use the source. Use lead shielding to absorb some of the gamma radiation – e.g. by placing this between the source and your body. Never handle gamma-sources directly, always remotely.

P7. Radioactivity

P7.1

Atomic structure:

a. Understand the atom in terms of protons, neutrons and electrons.
b. Know and be able to apply the nuclear model of atomic structure.
c. Know the relative charges and masses of protons, neutrons and electrons.
d. Understand and be able to use the terms *atomic number* and *mass number*.
e. Know and understand the term *isotope*.
f. Know and understand the term *nuclide*, and use nuclide notation.
g. Understand that ionisation is caused by the gain/loss of electrons.

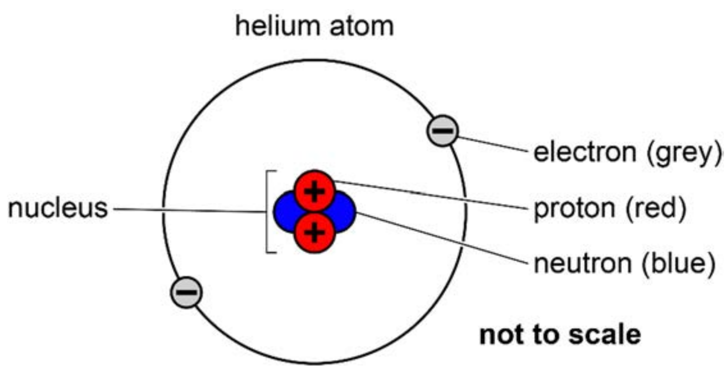
Understand the atom in terms of protons, neutrons and electrons

Know and be able to apply the nuclear model of atomic structure

The atom consists of three types of particle: protons, neutrons and electrons.

According to the nuclear model of atomic structure, an atom has a small, dense structure at its centre, called the nucleus. The nucleus consists of one or more protons, usually together with one or more neutrons. (The most common type of hydrogen nucleus consists of just a single proton, but all other types of nucleus contain neutrons as well as protons.) Protons and neutrons are known collectively as nucleons.

Outside the nucleus, the rest of the atom is empty space apart from one or more electrons. The diagram illustrates the nuclear model of an atom.



A nucleus typically has a radius around 100 000 times smaller than the radius of its atom, while containing over 99.9% of the atom's mass. The vast majority of an atom's volume is empty space, with the atom's overall size determined by the orbits of its electrons.

An atom of gold has 79 protons and 197 nucleons. How many neutrons does the atom have?

The atom has 79 protons, and 197 protons and neutrons altogether. So its number of neutrons is 197 − 79 = 118.

Know the relative charges and masses of sub-atomic particles

The particles which make up atoms – protons, neutrons and electrons – are all types of sub-atomic particle, meaning that they are particles which are smaller than the atom. The proton and neutron have very similar masses, while the electron's mass is far lower. Despite this, the proton and the electron carry the same amount of charge, with the proton being positive and the electron negative. The neutron, as its name suggests, is neutral (uncharged).

An atom has an equal number of protons and electrons, so it is neutral.

The relative charges and masses of the three types of sub-atomic particle are summarised in this table.

Type of particle	Relative mass	Relative charge
Proton	1	+1
Neutron	1	0
Electron	1/2000	−1

It may seem strange that protons can be packed closely within the nucleus, given that protons are positively charged and like charges repel. Protons and neutrons are held together in the nucleus by a force known as the 'strong nuclear force' or 'strong force', which balances the electrostatic repulsion between the protons.

Which of the following describes the charge on an atomic nucleus?

a Always positive

b Always negative

c Always neutral

d Varies depending on the type of atom

e Varies depending on whether the atom is chemically bonded to other atoms

The nucleus of an atom contains one or more protons and, usually, one or more neutrons. Protons are positively charged and neutrons are neutral, so the nucleus is always positive, a).

An atom has 19 nucleons, and it has 10 neutrons. How many electrons does the atom have?

The atom has 19 protons and neutrons altogether, and it has 10 neutrons, so it has $19 - 10 = 9$ protons.

An atom has an equal number of protons and electrons, so it has 9 electrons.

Understand and be able to use the terms atomic number and mass number

An atom's atomic number is the number of protons in its nucleus.

An atom's mass number is the total number of protons and neutrons in its nucleus.

Atomic number

The atomic number determines the element to which the atom belongs. If two atoms have different atomic numbers, they are atoms of different elements.

Atoms are neutral, so the number of electrons in an atom matches the number of protons, and is equal to the atomic number.

Atomic number is sometimes called 'proton number'.

Mass number

If the mass number and atomic number of an atom are both known, then the number of neutrons can be calculated as the difference between the two.

Mass number is sometimes called 'nucleon number', since protons and neutrons are collectively known as nucleons.

Write the atomic number and the mass number of:

an atom of hydrogen which has one proton and no neutrons

an atom of phosphorus which has 16 neutrons and 15 electrons

an atom of platinum which has 195 nucleons and 117 neutrons.

The number of protons is 1, so the atomic number is 1. The total number of protons and neutrons is 1, so the mass number is also 1.

The number of protons in an atom equals its number of electrons, so this atom has 15 protons and its atomic number is 15. The total number of protons and neutrons is 15 + 16 = 31, so the mass number is 31.

The number of nucleons (protons and neutrons) is 195, so the number of protons is 195 − 117 = 78, and this is the atomic number. The mass number is the number of nucleons, 195.

Know and understand the term isotope

Each element can exist as more than one possible nuclide. Nuclides of the same element are known as isotopes. For example, the three nuclides below are all isotopes of hydrogen:

$^{1}_{1}\mathrm{H} \quad ^{2}_{1}\mathrm{H} \quad ^{3}_{1}\mathrm{H}$

There are many more known nuclides than there are known elements, because every element has at least three known isotopes, and some have as many as thirty or more.

Some isotopes of an element may be more common than others. In some cases, an element in its naturally occurring form is made up of only one isotope (although other isotopes may have been created artificially or seen as products of radioactive decay). Some elements, however, have more than one isotope with significant abundance.

Which of the following must be isotopes of the same element?

1 A nuclide which has 19 protons and 20 neutrons.

2 A nuclide which has 20 protons and 20 neutrons.

3 A nuclide which has 19 protons and 19 neutrons.

4 A nuclide which has 40 nucleons.

Different isotopes of the same element have the same number of protons. The four nuclides have the following numbers of protons:

Nuclide	**Number of protons (atomic number)**
1)	19
2)	20
3)	19
4)	unknown

Nuclides 1) and 3) are definitely isotopes of the same element, since they have the same number of protons.

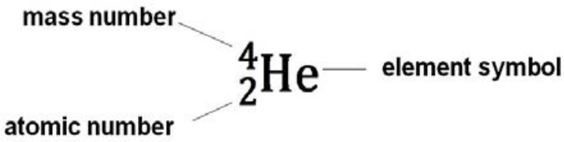
Nuclide 2) cannot be the same element as nuclides 1) and 3), because it has a different number of protons.

Nuclide 4) could be an isotope of the same element as nuclides 1) and 3), or alternatively, it could be exactly the same nuclide as nuclide 2). However, it could be a nuclide of a different, third element, if it has, for example, 21 protons and 19 neutrons. There is not enough information to know which of these possibilities is the right one, so we cannot say that nuclide 4) must belong to the same element as any of the others.

Know and understand the term nuclide, and use nuclide notation

A nuclide is any particular type (or 'species') of nucleus, characterised by the numbers of protons and neutrons it has.

The nuclear structure of an atom may be represented using the nuclide notation below.



If two nuclides have different atomic numbers, they are nuclides of different elements. For example, $^{3}_{1}\mathrm{H}$ and $^{3}_{2}\mathrm{He}$ have atomic numbers 1 and 2 and are nuclides of the elements hydrogen and helium respectively.

Examples of nuclide notation are shown below.

$^{4}_{2}\mathrm{He}$	$^{40}_{20}\mathrm{Ca}$	$^{64}_{30}\mathrm{Zn}$
This nuclide (a nucleus of helium) has 4 nucleons, 2 of which are protons. So it has 2 neutrons.	This nuclide (a nucleus of calcium) has 40 nucleons, 20 of which are protons. So it has 20 neutrons.	This nuclide (a nucleus of zinc) has 64 nucleons, 30 of which are protons. So it has 34 neutrons.

Nuclides are sometimes represented using only the element symbol and the mass number. He-4, for example, refers to a helium nucleus which has a mass number of 4. (This notation does not include the atomic number, so it gives no direct information about the number of protons.)

How many protons and how many neutrons does each of the following nuclides have?

$^{1}_{1}H$

$^{107}_{47}Ag$

$^{238}_{92}U$

In each case, the number of protons is given by the atomic number, which is in the lower position to the left of the element symbol.

The number of nucleons (protons and neutrons altogether) is given by the mass number, which is in the upper position to the left of the element symbol. The number of neutrons is the difference between the atomic number and the mass number.

Number of protons = atomic number = 1.

Number of neutrons = mass number − atomic number = 1 − 1 = 0.

Number of protons = atomic number = 47.

Number of neutrons = mass number − atomic number = 107 − 47 = 60.

Number of protons = atomic number = 92.

Number of neutrons = mass number − atomic number = 238 − 92 = 146.

Write the nuclide notation for each of the nuclides described below.

A nucleus of phosphorus (symbol P) which has 15 protons and 17 neutrons.

A nucleus of barium (symbol Ba) which has 82 neutrons and 138 nucleons.

The nuclide Ra-226, which has 138 neutrons.

The phosphorus nucleus has 15 protons so its atomic number is 15, and it has 17 neutrons so its mass number is 15 +17 = 32. So its nuclide notation is

$$^{32}_{15}\mathrm{P}$$

The barium nucleus has 138 nucleons, 82 of which are neutrons, so it has 138 − 82 = 56 protons. Its atomic number is therefore 56, and its mass number is 138. So its nuclide notation is

$$^{138}_{56}\mathrm{Ba}$$

Ra-226 has a mass number of 226. It has 138 neutrons, so it has 226 − 138 = 88 protons. Its atomic number is therefore 88. So its nuclide notation is

$$^{226}_{88}\mathrm{Ra}$$

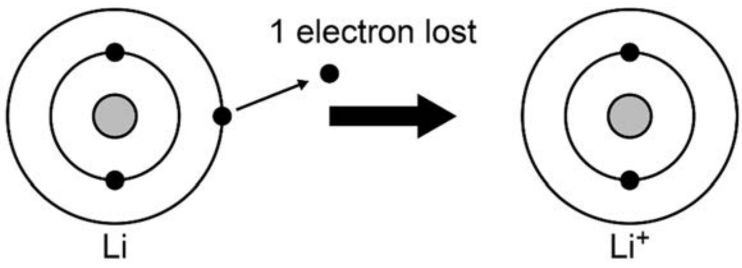
Understand that ionisation is caused by the gain/loss of electrons

In certain circumstances an atom may gain or lose one or more outer electrons, becoming negatively or positively charged (respectively) as a result. It is no longer referred to as an atom; it is now called an ion.

The process of gaining or losing one or more electrons is called ionisation.

Atoms do not readily gain or lose protons (or neutrons), and when an atom is ionised its number of protons is not affected. So if an atom loses electrons, the resulting ion has more protons than electrons and is positive overall; if it gains electrons, the resulting ion has more electrons than protons and is negative overall.

The diagram illustrates the loss of an electron by a lithium atom. This forms an ion which has one more proton than it has electrons, so the ion has an overall charge of +1. The ion can be represented by the symbol Li^+.



Li atom with 3 electrons Li+ ion with only 2 electrons

There are several circumstances in which an atom may gain or lose one or more electrons. In ionic bonding, atoms bond by exchanging electrons and becoming ions. Ionisation may also happen when two different materials are rubbed together (or even when they simply touch each other), causing some outer electrons to transfer from one material to the other. Ionisation can also be caused by ionising radiation.

An atom becomes an ion with a charge of +1. How does this happen?

An ion with a charge of +1 has one more proton than it has electrons. Atoms do not gain or lose protons during ionisation; they only gain or lose electrons. So the atom must lose an electron.

A particular ion has 45 nucleons, 19 electrons and a net charge of +2. The ion's number of electrons changes so that it becomes a neutral atom. How many protons, neutrons and electrons does this atom have?

A charge of +2 on the ion means that it has two more protons than it has electrons.

It therefore has 19 + 2 = 21 protons.

The atom has 45 protons and neutrons altogether, so it has 45 − 21 = 24 neutrons.

The neutral atom has the same number of electrons as protons, so it has 21 electrons.

An atom of ${}^{58}_{28}Ni$ gains two electrons.

What are the atomic number and the mass number of the ion formed?

The change in the number of electrons does not affect the nucleus. It still has the same numbers of protons and neutrons, so its atomic and mass numbers are unchanged. So the ion has an atomic number of 28 and a mass number of 58.

P7.2

Radioactive decay:

a. Know and understand that emissions arise from an unstable nucleus.
b. Know and understand the random nature of emissions.
c. Know and understand the differences between alpha, beta and gamma emission.
d. Know and understand the nature of alpha and beta particles, and gamma radiation.
e. Be able to use and interpret nuclear equations.
f. Know the effect of decay on atomic number and mass number.

Know and understand that emissions arise from an unstable nucleus

There are many naturally occurring nuclides. Some of these types of nucleus are stable and others are unstable. Nuclei that are stable will continue to exist indefinitely.

Nuclei that are unstable will, sooner or later, break down or 'decay', emitting radiation as they do so. The emitted radiation is one or more of three types: alpha, beta and gamma.

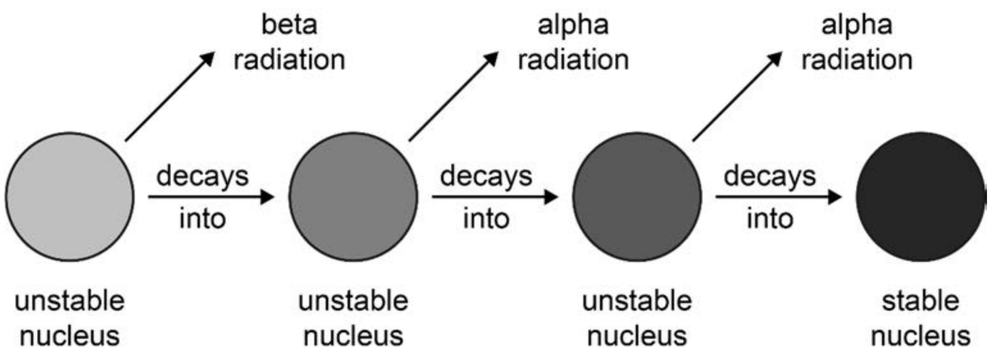
Nuclides that decay are described as radioactive.

All of the naturally occurring elements have more than one isotope, and some have as many as thirty or more, but most elements have only one or two isotopes that are stable. The rest are unstable, or radioactive. Below is a table showing the known isotopes of lithium as an example.

Isotope	${}^{4}_{3}\mathrm{Li}$	${}^{5}_{3}\mathrm{Li}$	${}^{6}_{3}\mathrm{Li}$	${}^{7}_{3}\mathrm{Li}$	${}^{8}_{3}\mathrm{Li}$	${}^{9}_{3}\mathrm{Li}$	${}^{10}_{3}\mathrm{Li}$	${}^{11}_{3}\mathrm{Li}$	${}^{12}_{3}\mathrm{Li}$
Stability	unstable	unstable	stable	stable	unstable	unstable	unstable	unstable	unstable

An unstable nucleus may decay into a stable nucleus, in which case there will be no further decay. Some types of radioactive nuclide, however, decay into a nuclide which is also radioactive. There may be a series of decays, called a decay chain, which ends when a stable nuclide is reached. The decays may involve more than one type of emission.

The diagram shows an example of a decay chain.



Know and understand the random nature of emissions

Radioactive decay is described as a 'random' process, because it is not possible to predict the time at which an unstable nucleus will decay.

There is no known cause that makes a nucleus decay at one particular moment rather than at a different moment. There is also no known way to make a nucleus decay sooner, or later. A process that occurs without a trigger or cause is described as 'spontaneous'.

Although it is impossible to predict when a particular nucleus will decay, it is possible to find out experimentally the probability that a nucleus of a particular type will decay during a given time period.

The random yet probabilistic nature of decay can be modelled by the repeated throwing of a dice, where the dice represents an unstable nucleus and each throw represents the passing of a period of time. If we use a 6-sided dice and the first throw of a 6 represents 'decay', we cannot predict at what time (i.e. on which throw) the dice will decay. However, we can say that the probability of decay on each throw is $\frac{1}{6}$.

The random nature of radioactive decay means that the measured count rate from a radioactive source cannot be predicted exactly, and will show some variation. When a count rate is measured experimentally, the mean value is taken over a period of time.

State whether each of the following statements is correct or incorrect.

Raising the temperature of a radioactive substance makes it decay faster.

An unstable nucleus can decay into another unstable nucleus.

The isotopes of an element are either all stable or all unstable.

You may know that an increase in temperature usually increases the rate of a chemical reaction. However, it has no effect on the rate of nuclear decay. There is no known way to change the rate of nuclear decay. This statement is incorrect.

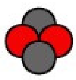
Depending on the nuclide, an unstable nucleus may decay into either a stable or an unstable nucleus. This statement is correct.

Most elements have both stable and unstable isotopes. This statement is incorrect.

Know and understand the differences between alpha, beta and gamma emission Know and understand the nature of alpha and beta particles, and gamma radiation

When an unstable nucleus decays, it emits one or more of the three types of ionising radiation: alpha, beta and gamma. These nuclear processes release energies far greater than the energies released by chemical reactions.

This table summarises the nature and characteristics of the three types of nuclear radiation.

Type of radiation	**alpha**	**beta**	**gamma**
Symbol	α or He	β^- or e^-	γ
Nuclide notation	${}^{4}_{2}\alpha$ or ${}^{4}_{2}\mathrm{He}$	${}^{0}_{-1}\beta$ or ${}^{0}_{-1}\mathrm{e}$	sometimes written as ${}^{0}_{0}\gamma$
Nature	two protons and two neutrons (identical to a helium nucleus) 	a fast-moving electron	a burst of electromagnetic radiation (of the gamma part of the spectrum) 
Mass	4 × proton mass	$\frac{1}{2000}$ × proton mass [approx.]	0
Relative charge (where the charge on a proton is +1)	+2	−1	0
Speed (compared with speed of light, c = 3×10^{8} ms^{-1})	typically around 0.1c	typically around 0.8 c	c

The type(s) of radiation an unstable nucleus emits depends on which nuclide it is.

Alpha and beta radiation may also be referred to as 'alpha particles' and 'beta particles'. ('Radiation' means the emission of energy via waves or particles, so the term can be correctly applied to alpha and beta as well as to gamma.)

The nuclide notation for an alpha particle, ${}_{-1}^{0}\beta$ or ${}_{2}^{4}\mathrm{He}$, indicates that an alpha particle has a mass number of 4 since it consists of four nucleons, and an atomic number of 2 since two of those are protons.

The nuclide notation for a beta particle, ${}_{-1}^{0}\beta$ or ${}_{-1}^{0}e$, indicates that a beta particle has a mass number of 0 since it has negligible mass compared with a nucleon, and an atomic number of −1 since its charge is the negative equivalent of the charge on a proton.

Which type of ionising radiation:

Is negatively charged? Beta

Has the greatest mass? alpha

Is a type of electromagnetic radiation? Gamma

Is the fastest moving? Gamma

Is made up of two protons and two neutrons? Alpha

Be able to use and interpret nuclear equations Know the effect of decay on atomic number and mass number

The emission of radiation affects a nucleus in different ways depending on the type of radiation emitted. The table below shows the origin of each type of radiation and the effect its emission has on the atomic number and mass number of the nucleus.

Type of radiation	**alpha**	**beta**	**gamma**
Origin of the radiation	The unstable nucleus emits two of its protons and two of its neutrons, bound together as a single particle.	One neutron of the unstable nucleus transforms into a proton (which remains in the nucleus) and an electron (which is emitted as the beta particle).	Excess energy of the unstable nucleus is ejected in the form of gamma radiation.
Effect on atomic number of nucleus	−2	+1	no change
Effect on mass number of nucleus	−4	no change	no change

Both alpha decay and beta decay cause a change in the number of protons, or atomic number, of the nucleus. This means that the resulting nucleus is of a different element than the original nucleus.

For alpha decay or beta decay, it is possible to write a decay equation.

A decay equation shows the initial nucleus, the final nucleus and the emitted particle. Each of these is written in nuclide notation.

Alpha decay equations

Here is an example of an alpha decay equation in which a nucleus of radium-226 emits an alpha particle and becomes a nucleus of radon-222:

$${}^{226}_{88}\mathrm{Ra} \rightarrow {}^{222}_{86}\mathrm{Rn} + {}^{4}_{2}\alpha$$

Because an alpha particle consists of four nucleons, two of which are protons, all alpha decay equations follow the pattern:

$${}^{A}_{Z}\mathrm{X} \rightarrow {}^{A-4}_{Z-2}\mathrm{Y} + {}^{4}_{2}\alpha$$

Beta decay equations

Here is an example of a beta decay equation in which a nucleus of carbon-14 emits a beta particle and becomes a nucleus of nitrogen-14:

$${}^{14}_{6}\mathrm{C} \rightarrow {}^{14}_{7}\mathrm{N} + {}^{0}_{-1}\beta$$

Because beta decay occurs when a neutron in the nucleus changes into a proton, all beta decay equations follow the pattern:

$${}^{A}_{Z}\mathrm{X} \rightarrow {}_{Z+1}^{A}\mathrm{Y} + {}^{0}_{-1}\beta$$

In a decay equation where the mass number of one of the nuclides is not written, it is possible to deduce its value using the fact that the sum of the mass numbers on the right-hand side equals the mass number of the original nucleus. This is because no nucleons are created or destroyed in the decay process.

Similarly, a missing atomic number can be deduced using the fact that the sum of the atomic numbers on the righthand side equals the atomic number of the original nucleus. This is because charge is conserved in the decay process: the charge on the original nucleus equals the sum of the charges of the products (taking into account their positive and negative signs).

Thorium-232 undergoes alpha decay to become an isotope of radium. Complete the decay equation for this process.

$$^{232}_{\square}\mathrm{Th} \rightarrow {}^{\square}_{88}\mathrm{Ra} + {}^{\square}_{\square}\square$$

The radium isotope produced in the above decay undergoes beta decay to become an isotope of actinium (element symbol Ac). Write a decay equation for this process.

The nuclide notation for an alpha particle is ${}^{4}_{2}\alpha$ (or ${}^{4}_{2}\mathrm{He}$). The thorium nucleus loses four nucleons, so the mass number of the radium nucleus is 232 − 4 = 288. The thorium nucleus loses two protons, so its mass number before the decay is 88 + 2 = 90.

$$^{232}_{90}\mathrm{Th} \rightarrow {}^{228}_{88}\mathrm{Ra} + {}^{4}_{2}\alpha$$

(As a check, you can see that the mass numbers add to the same total, 232, on each side of the equation, and the atomic numbers add to the same total, 90, on each side.)

The nuclide notation for a beta particle is

$${}^{0}_{-1}\beta \text{ or } {}^{0}_{-1}e^{-}$$

In beta decay, a nucleus keeps the same number of nucleons but loses a neutron and gains a proton, so the mass number of the actinium nucleus is the same as that of the radium nucleus, and the proton number is 88 + 1 = 89.

$$^{228}_{88}\mathrm{Ra} \rightarrow {}^{228}_{89}\mathrm{Ac} + {}^{0}_{-1}\beta$$

(As a check, you can see that the mass numbers add to the same total, 228, on each side of the equation, and the atomic numbers add to the same total, 88, on each side.)

Write, in nuclide notation, the nuclide $^{A}_{Z}X$ after:

a A beta decay.

b An alpha decay and a gamma decay.

c One alpha and two beta decays.

a)

$^{A}_{Z+1}Y$

In beta decay, a neutron transforms into a proton, which stays in the nucleus, and a beta particle. The number of nucleons (and therefore the mass number) stays the same but the number of protons (and therefore the atomic number) increases by 1.

b)

$^{A-4}_{Z-2}Y$

In alpha decay, two protons and two neutrons are ejected from the nucleus. The number of nucleons (and therefore the mass number) decreases by 4 and the number of protons (and therefore the atomic number) decreases by 2. In gamma decay, only energy is released, so the makeup of the nucleus is unchanged.

c)

$^{A-4}_{Z}X$

Four nucleons are ejected in alpha decay and none in beta decay. Two protons are ejected in the alpha decay and one proton is gained in each beta decay. So the net result is that the number of nucleons (and hence the mass number) decreases by 4 and the number of protons (and hence the atomic number) is unchanged.

P7.3

Ionising radiation:

a. Know the relative penetrating abilities of alpha, beta and gamma radiation.
b. Know the relative ionising abilities of alpha, beta and gamma radiation.
c. Understand qualitatively the deflection of alpha, beta and gamma radiation in electric or magnetic fields.
d. Know and appreciate the existence of background radiation.
e. Understand the applications and hazards of ionising radiation.

Know the relative penetrating abilities of alpha, beta and gamma radiation

The penetrating ability of a type of radiation refers to how easily it can pass through materials.

Gamma is the most penetrating of the three types of nuclear radiation, and beta is typically more penetrating than alpha. The table below lists examples that illustrate the typical relative penetrating abilities of the three types of radiation. (Emissions from different isotopes will vary in energy and hence penetrating ability, so what follows is very much a generalisation.)

Type of radiation	**Description of penetrating ability**
alpha	blocked by a sheet of paper blocked by human skin can typically penetrate a few centimetres in air
beta	typically blocked by thin metal typically not blocked by human skin can penetrate up to several metres in air
gamma	to block it to a large extent requires several centimetres of very dense material such as lead can penetrate up to hundreds of metres in air

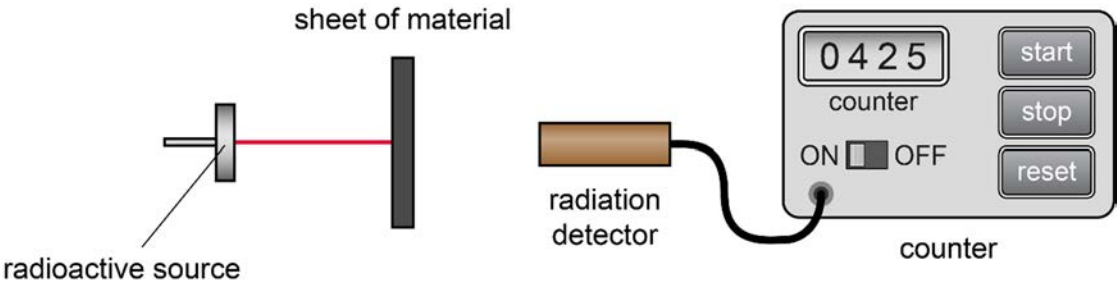
Each of the three types of nuclear radiation can have a range of different energies depending on the nuclide that emits it, and as a result there is a range of penetrating abilities for each; that is why descriptions in the table above include the wording ‘typically’ and ‘up to’.

Of the three types of radiation, beta has the most widely varying penetrating ability. As stated in the table, it can be blocked by thin metal, but depending on the type of metal and the energy of the beta particles the necessary thickness ranges from one to several millimetres thick. The lowest-energy beta particles cannot penetrate the skin.

One way to identify the type(s) of radiation emitted by a radioactive source is to investigate the penetrating ability of the radiation experimentally, using a radiation counter as illustrated in the diagram below.

The count rate (number of radiation impacts detected by the counter per second) is measured with different materials between the source and the counter, to find out how penetrating the radiation is.

Radiation counter



If a radioactive source emits both alpha and gamma radiation, then placing a barrier between the source and the detector which blocks the alpha but not the gamma will eliminate only the alpha contribution to the count rate. The same applies if the source emits beta and gamma radiation.

A radiation counter is placed near a radioactive source and measures a high count rate. When a sheet of paper is placed between the source and the counter, the measured count rate is not significantly affected. When a thin sheet of aluminium is placed between the source and the counter, the count rate is significantly reduced, but remains much higher than the count rate when the source is far away from the counter.

What is/are the most likely radiation type(s) emitted by the source?

Beta and gamma. A sheet of paper would absorb alpha radiation, and since the paper has no effect, the source is not emitting alpha radiation. A thin sheet of aluminium would not significantly absorb gamma radiation, but it could absorb beta radiation. So the source is emitting beta radiation. The count rate with the metal sheet in place is still higher than the level without the source present, so the source is also likely to be emitting gamma radiation.

Know the relative ionising abilities of alpha, beta and gamma radiation

Alpha, beta and gamma radiation are all types of ionising radiation. This means that when they collide with matter, they can knock electrons out of atoms, forming positive ions. These electrons may attach to other atoms, which then become negative ions.

The ionising abilities of the radiation types are related to their penetrating abilities, as shown in the table below.

Alpha	**Beta**	**Gamma**
Least penetrating	Intermediate	Most penetrating
Most ionising	Intermediate	Least ionising

Alpha radiation is the most ionising because, although alpha particles travel at lower speeds than beta particles, alpha particles' much greater mass means that they have more momentum; this, along with their double charge, gives them a strong tendency to interact with atoms and cause ionisation. Beta particles have lower momentum but have a single charge, so they interact with matter more strongly than gamma radiation, which is uncharged.

Radiation that is very ionising quickly loses its kinetic energy as it travels through matter, and so it is not very penetrating. Conversely, radiation that is not very ionising is much more penetrating as it tends to travel further before losing its kinetic energy.

Alpha, beta and gamma radiation are not the only known types of ionising radiation. X-rays and the higher frequencies of ultraviolet radiation can also ionise atoms.

Understand qualitatively the deflection of alpha, beta and gamma radiation in electric or magnetic fields

The table below describes the effects of electric fields and magnetic fields on alpha particles, beta particles and gamma rays.

Effects of electric fields on:	**Effects of magnetic fields on:**
alpha and beta particles	**alpha and beta particles**
Both alpha and beta particles are deflected (made to change direction) in the presence of electric fields, since charged bodies experience forces when they are in an electric field. Because of their opposite charges, alpha and beta particles are deflected in opposite directions. The extent of the deflection depends on a particle's mass, charge and speed. Because of their much lower mass, beta particles are deflected more than alpha particles. (More details below.)	Both alpha and beta particles are deflected in the presence of magnetic fields, since moving charges experience forces when they are in a magnetic field. Because of their opposite charges, alpha and beta particles are deflected in opposite directions. Beta particles are deflected more than alpha particles, again because of their much lower mass.
gamma rays	**gamma rays**
Since they have no charge, gamma rays are not deflected by electric fields.	Since they have no charge, gamma rays are not deflected by magnetic fields.

The table below summarises the factors affecting deflection in an electric field, and how these result in greater deflection for beta particles than for alpha particles.

Factor	Effect
charge	Alpha particle has twice as much charge as beta, so experiences twice as much force in an electric field.
mass	Alpha particle is about 8000 times heavier than beta, so acceleration $a = \frac{F}{m}$ is about $\frac{2}{8000}$ × or $\frac{1}{4000}$× that of a beta particle in the same field.
speed	Alpha particle typically has 10–20% of speed of beta – therefore it takes longer to move through the field so there is more time for the force to act.
overall effect	The effect of the speed difference is less than the opposing effect of the charge and mass differences (with the mass difference being the dominant factor). So an alpha particle is less deflected than a beta particle in the same electric field.

Three types of ionising radiation are P, Q and R.

In a magnetic field, P and R are deflected in opposite directions, and Q is not deflected.

R is the most ionising of the three types of radiation.

Identify the three types of radiation.

Alpha and beta radiation are both charged, so both experience a force in a magnetic field. Because they have opposite charges, they are deflected in opposite directions. Gamma is uncharged so is not deflected, so Q is gamma radiation. Alpha radiation is the most ionising type of nuclear radiation, so R is alpha and P is beta.

Know and appreciate the existence of background radiation

Sources of background radiation

There is always ionising radiation present at a low level in the human environment. This is called background radiation, and most of it is naturally occurring, although a small percentage is due to human activities.

The table below summarises the main contributions to the total background radiation from different sources.

Natural sources of background radiation (typically over 80% of the total)	**Artificial sources of background radiation** (typically under 20% of the total)
radon gas from the ground rocks and buildings cosmic rays food and drink	medical procedures (typically around 99% of the total radiation from artificial sources) nuclear power and nuclear weapons testing

Background radiation is taken to include all types of ionising radiation that are present in our environment, including radiation that is not nuclear in origin. Hence it includes X-rays which, though not a type of nuclear radiation, are ionising.

Exposure to background radiation from different sources will vary from one person to another, depending on where they live (since different rock and soil types emit radiation, and release radon gas, at different rates) and on the medical procedures they undergo. (The effect of a person's job is extremely small: radiation exposure is hardly any higher than average for people who work in hospital X-ray or radiotherapy departments or at nuclear power stations.)

Correcting for background radiation in count rate measurements

When the count rate from a radioactive source is measured using a radiation detector, the detector readings should be corrected for background radiation. This is done by measuring the average background count rate (the mean count rate in the absence of the radioactive source) and subtracting it from the average count rate in the presence of the source:

mean count rate from source = mean count rate (measured with source present) − mean background count rate (measured with source absent)

A scientist measures the count rate from a radioactive source as 200 counts per second (cps).

The scientist removes the source but does not move the radiation counter. The measured count rate is now 10 cps.

Explain why the count rate is not zero when the radioactive source is not present.

Estimate the count rate from the source.

State one reason why the background radiation rate may be higher in one region than in another.

Background radiation. There is always a low level of ionising radiation in the environment, mainly from natural sources such as rocks and cosmic rays.

Count rate from source = total count rate – background count rate = 200 – 10 = 190 cps.

For example, different types of rock and soil emit radiation at different rates (because they contain different levels of radioactive isotopes).

Understand the applications and hazards of ionising radiation

Hazards

Nuclear radiation can damage human tissue because it ionises atoms and molecules, which can affect the chemical reactions that take place in cells. In particular, ionisation can damage the delicate strands of DNA within the nuclei of cells. This can cause cells to malfunction, and in some cases this leads to uncontrolled cell division, or cancer.

High doses of nuclear radiation destroy living cells, which can cause a variety of symptoms sometimes described as 'radiation sickness' and, in extreme cases, death.

The table below summarises the risks of being exposed to an alpha, beta or gamma source inside or outside the body. Remember that a radioactive nuclide may emit more than one type of radiation.

Type of radiation source	Inside the body	Outside the body
alpha	the most hazardous because alpha radiation is highly ionising and will damage nearby cells	not hazardous, because alpha particles cannot penetrate the skin
beta	less hazardous than alpha because it is less ionising, but will still cause some cell damage	hazardous, because beta particles can typically penetrate the skin and damage the tissue underneath
gamma	generally less hazardous than alpha and beta because it is less ionising, and much of it will pass straight through cells without damaging them	hazardous, because gamma rays can easily penetrate the skin and cause damage anywhere in the body

Applications

Nuclear radiation has many practical applications, both in medicine and in industry. For each application, a suitable radioactive nuclide is selected, based on its half-life and the type(s) of radiation it emits.

Given details of an application, it is often possible to deduce which type of source and what half-life (whether hours, days, or years, for example) would be suitable. Examples are shown below.

Application	**Information given**	**Deductions about suitable type of source and half-life**
Sterilisation	Exposing an object to a high dose of ionising radiation kills any bacteria and other types of single-celled organism that are on or inside the object, including any pathogens. This technique is used to sterilise medical equipment (such as syringes) and some foods. It is particularly useful for plastic items such as syringes, which cannot be sterilised by heating because they would melt.	Since gamma radiation is the most penetrating, it is the best type of radiation for reaching all parts of the equipment or food being sterilised. (Although it is the least ionising nuclear radiation, it will kill pathogens if the radiation intensity is high enough.) The half-life should be long (years rather than days) so that the radiation source does not need to be replaced frequently.
Medical tracers	Medical tracers are radioactive nuclides that enable observation of the structure and function of internal organs. The tracer is swallowed or injected, and continues to emit radiation which is then detected from outside the body. This makes it possible to observe the movement of the tracer through the organ of interest, and so observe how the organ is functioning. The procedure typically takes no more than a few hours.	This application requires radiation which causes minimal ionisation and easily penetrates the body – so a gamma source is used. It is important that the source does not also emit alpha or beta, as these are highly ionising and therefore damaging to tissues. The half-life should be long enough for the test to be completed before the count rate drops too much for useful measurements to be possible. At the same time, the patient's exposure to radiation should be minimised by using a source with a half-life that is no longer than necessary. So the half-life needs to be short (hours rather than days or years).

Thickness gauge	When sheet material (such as paper or cooking foil) is made in a factory, the sheet thickness needs to be kept within an acceptable range of values. This can be done by positioning a radioactive source on one side of the moving sheet and a radiation counter on the other side. A change in sheet thickness causes a change in the measured count rate, and this information is used to make corrective adjustments to the machinery that controls the sheet thickness.	The radiation needs to be able to pass through the sheet but be significantly affected by changes in sheet thickness. Alpha radiation would be completely blocked, while gamma radiation would pass through unaffected by small changes in thickness. Beta radiation is ideal as it will partly penetrate the sheet material, with the degree of penetration strongly affected by the sheet's thickness. The half-life should be long (years rather than days) so that the radiation source does not need to be replaced frequently and so that variation in measured count rate is caused only by variation in thickness and not depletion of the source.

Exposure to nuclear radiation does not make an object become radioactive; hence it is safe to sterilise food and surgical equipment using gamma rays.

X-rays and the higher frequencies of ultraviolet radiation can also ionise atoms, and so can damage exposed cells in similar ways to nuclear radiation.

There is no specific amount of exposure to ionising radiation that guarantees that a person will develop cancer; but the greater the exposure, the higher the probability that the person will develop cancer at some point in their life.

People who are working with radioactive sources should take certain precautions – all of which help to minimise their exposure to nuclear radiation. These include:

- storing radioactive sources in lead-lined containers which are clearly labelled
- keeping sources as far as possible from the body while working with them, for example by handling them using long-handled tongs
- wearing gloves and other protective clothing.

An underground water pipe is tested for leaks by injecting a small amount of a soluble radioactive material (called a tracer) into water as it enters the pipe. A radiation counter is used to measure the count rate at the ground surface, along the pipe's route.

The measured count rate is higher above a leak in the pipe.

State the most suitable type of radiation for this application. Explain your choice.

Which is the most suitable half-life for the isotope: minutes, days or years?

Gamma radiation. Alpha radiation would not penetrate through the ground to the surface, nor would most beta radiation, so these would not be significantly detected by a counter at the surface. (Water leaking from the pipe soaks into the surrounding soil, carrying the tracer closer to the surface, but it does not necessarily carry the tracer all the way up to the surface.) Gamma radiation can be detected from the surface even where there is no leak, but will be detected at a significantly higher level above a leak.

Days. There needs to be enough time to transport the source to the location where it is injected, and for it to flow along the pipe and build up in the soil near a leak. However, the half-life should not be longer than necessary, to avoid unnecessary exposure to humans and other living organisms.

P7.4

Half-life:

a. Be able to interpret graphical representations of radioactive decay (including consideration of decay products).
b. Understand the meaning of the term *half-life*.
c. Understand and be able to apply half-life calculations.

Be able to interpret graphical representations of radioactive decay (including consideration of decay products)

The half-life of a radioactive source can be read from a graph of count rate against time (or from a graph of number of remaining undecayed nuclei against time).

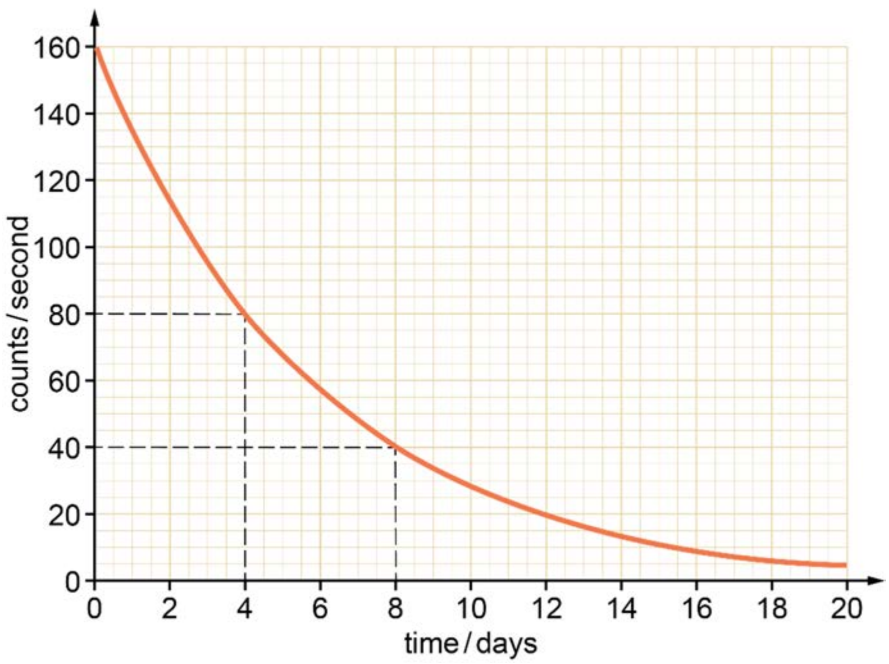
The graph below shows the count rate detected from a radioactive source versus time in days. (The count rates have been corrected for background radiation.)

The graph shows that the time taken for the count rate to drop from 160 cps (counts per second) to 80 cps is 4 days.

So the half-life of the source is 4 days.

If we instead start by reading the count rate of 80 cps at 4 days, we find that this has halved to 40 cps at 8 days. So again we find that the half-life of the source is 4 days.

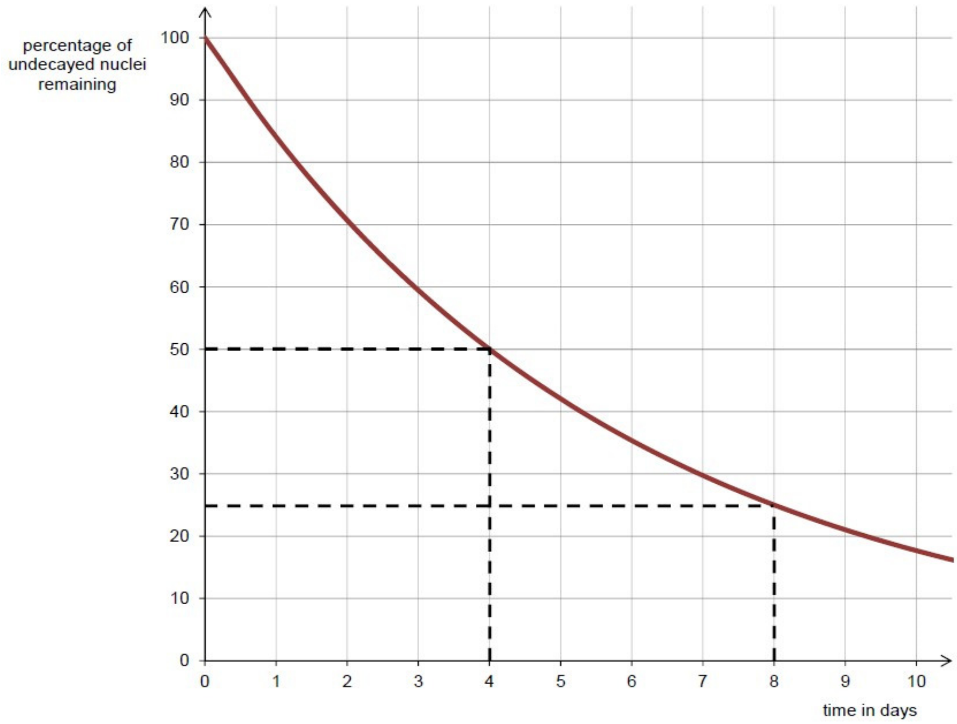
Count rate detected from a radioactive source



The half-life can be found by choosing any time and reading its corresponding count rate, and then reading how long it takes for that count rate to fall by half.

A graph of number of unstable nuclei remaining (often written as 'number of undecayed nuclei') against time can be used to find half-life in a similar way. However, the number of nuclei in a source at any given time is likely to be very large, so the percentage of nuclei remaining is commonly plotted instead as it is easier to read and interpret.

The percentage of undecayed nuclei remaining in a radioactive source versus time in days, for the same source as the graph above



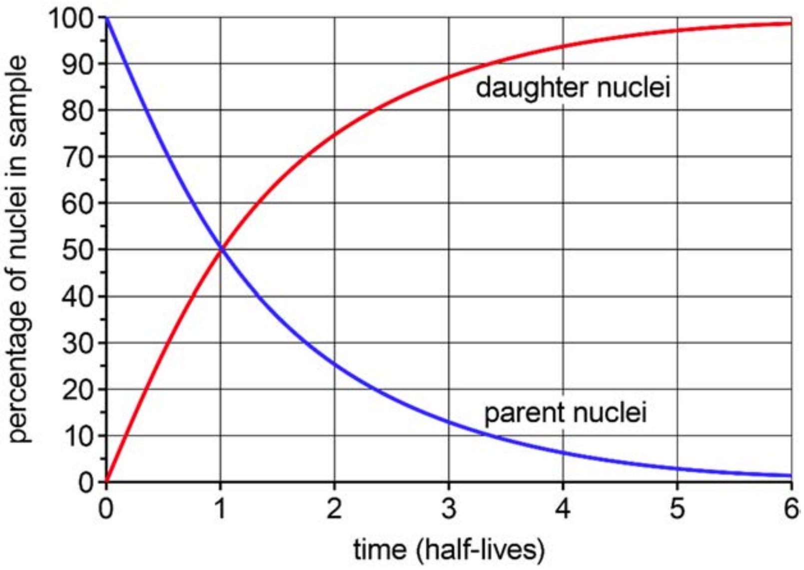
The percentage of undecayed nuclei drops from 100% to 50% in 4 days, so the half-life read from the graph is again 4 days.

Consider a sample of radioactive nuclide X which decays into stable nuclide Y. Initially, 100% of the sample's nuclei are of type X. As time passes, the percentage of type X nuclei ('parent nuclei') falls, as seen in the graphs above.

Meanwhile, the percentage of type Y nuclei ('daughter nuclei') increases.

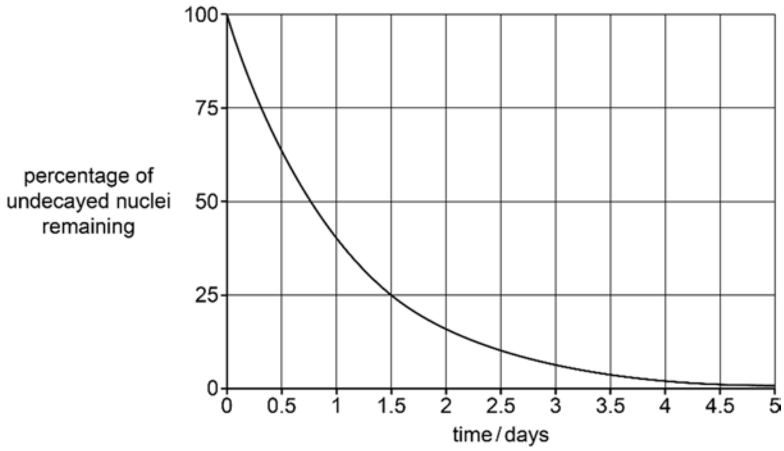
Percentages of each type of nucleus against time in the sample

(assuming that all of the atoms stay within the sample)

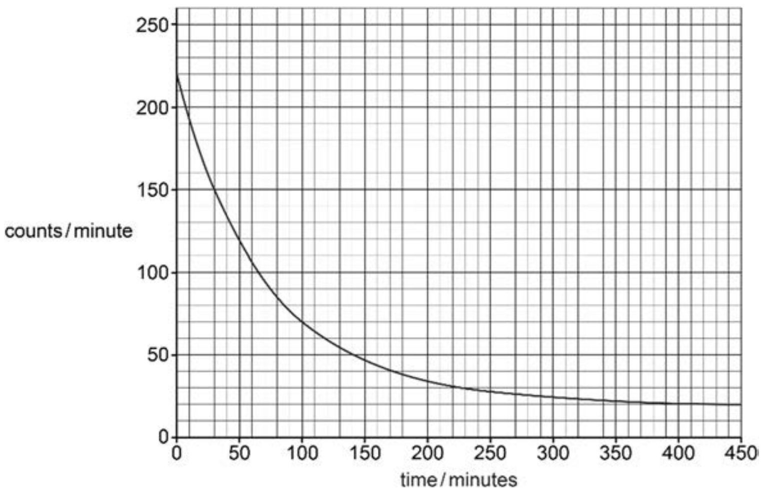


The percentages of parent and daughter nuclei at any given time add to 100%.

The graph below shows the percentage of undecayed nuclei remaining against time for a sample of a radioactive isotope. Find the half-life of the source.



The graph below shows measurements of count rate against time for a sample of a radioactive isotope, but the measurements have not been corrected for background radiation. Find the half-life of the source.



Choose a time at which to read the percentage of undecayed nuclei remaining, and then find the time it takes for this percentage to halve. For example, when the time is 0 the percentage of undecayed nuclei is 100%. This drops to 50% after 0.75 s. However, this is difficult to read exactly from this graph because there are no fine gridlines, so you could find the time it takes for the percentage to halve twice, to 25%. The percentage is 25% after 1.5 s, so the half life is half of this, or 0.75 s.

50 minutes. The background count rate is 20 cpm, since this is the rate after the radiation from the source has died away. Choose a time at which to find the source count rate and then find the time it takes for this to halve. For example, when the time is 0 the source count rate is 220 − 20 = 200 cpm. After 50 min the total count rate is 120 cpm and the source count rate is 100 cpm, so the source count rate halves in 50 minutes, which is the half-life. (Note that the background count rate is shown as constant in this graph, whereas in reality the random nature of radioactive decay would cause the background count rate to vary around an average value.)

Understand the meaning of the term half-life

The half-life of a radioactive isotope is the average time taken for half of the nuclei in a sample of that isotope to decay.

The half-life of an isotope is also the average time taken for the count rate of a sample of the isotope to halve (assuming the radiation counter is kept at a constant distance from the source).

The two definitions above both yield the same time, since the count rate from a radioactive source is proportional to the number of unstable nuclei remaining in it. If the number of unstable nuclei in a sample halves, for example, then the count rate from the sample also halves.

If a radioactive sample has a long half-life, its count rate will fall slowly. If it has a short half-life, its count rate will fall quickly.

Half-life is unaffected by the size of the sample, or by the percentage of nuclei that have already decayed – and this is what makes it such a useful measure for comparing how rapidly different isotopes decay, and for predicting count rates of radioactive samples over time.

Different radioactive isotopes have different half-lives, with measured half-lives ranging from tiny fractions of a second to much longer than the lifetime of the universe.

Which of the following affects the half-life of a sample of a radioactive nuclide?

The number of unstable atoms in the sample.

The length of time for which the sample has been decaying.

Which radioactive nuclide is present in the sample.

Half-life is unaffected by the size of the sample and by the amount of time that has passed. The sample count rate always halves during a time equal to one half-life, no matter at what time that half-life begins. However, the length of a half-life is different for different radioactive nuclides. The answer is 3) only.

Understand and be able to apply half-life calculations

Calculating count rates

If the half-life and initial count rate of a radioactive source are known, it is possible to predict the count rate at any whole number of half-lives later. The count rate will halve during each period of time equalling the half-life.

For example, consider a radioactive source with an initial count rate of 4000 counts per minute (cpm) and a half-life of 10 h.

Count rate after 10 h (= 1 half-life): $\frac{1}{2} \times 4000$ cpm = 2000 cpm

Count rate after 20 h (= 2 half-lives): $\frac{1}{2} \times \frac{1}{2} \times 4000$ cpm = 1000 cpm

Count rate after 50 h (= 5 half-lives): $(\frac{1}{2})^5 \times 4000$ cpm = 125 cpm

Similar calculations can be carried out using percentages of undecayed nuclei instead of count rates.

Consider a radioactive source in which no nuclei have yet decayed and the half-life is 5 years (yr).

Percentage of nuclei undecayed after 5 yr (= 1 half-life): $\frac{1}{2} \times 100\% = 50\%$

Percentage of nuclei undecayed after 20 yr (= 4 half-lives): $(\frac{1}{2})^4 \times 100\% = 6.25\%$

Calculating half-life

Conversely, it is possible to deduce the half-life of a radioactive source given its count rates at two different times (where the two times are a whole number of half-lives apart).

For example, consider a source with an initial count rate of 600 counts per second (cps) and a count rate of 75 cps at a time 12 h later.

Number of times count rate halves is 3: 600 → 300 → 150 → 75

So 12 h is equivalent to 3 half-lives, and one half-life is 12 h ÷ 3 = 4 h

A similar calculation can be done to find a half-life given percentages of undecayed nuclei remaining at two different times.

If any measured count rate includes background radiation, then before it is used in a calculation it should have the the mean background count rate subtracted from it, to give the source count rate.

A radioactive source has half-life 10 seconds. Its measured count rate now is 64 000 counts per second (cps). a) What was the count rate 20 s ago?

What will the count rate be after 30 s?

After how much time will the count rate be 500 cps?

20 s equals two half-lives. Working backwards in time, the count rate doubles for each half-life. So the count rate two half-lives ago was 64 000 → 128 000 → 256 000 cps.

30 s equals three half-lives. Halve the present count rate three times: 64 000 → 32 000 → 16 000 → 8000 cps.

Find how many times 64 000 halves to become 500: 64 000 → 32 000 → 16 000 → 8000 → 4000 → 2000 → 1000 → 500. 500 is 64 000 after halving seven times. Seven half-lives is 70 s.