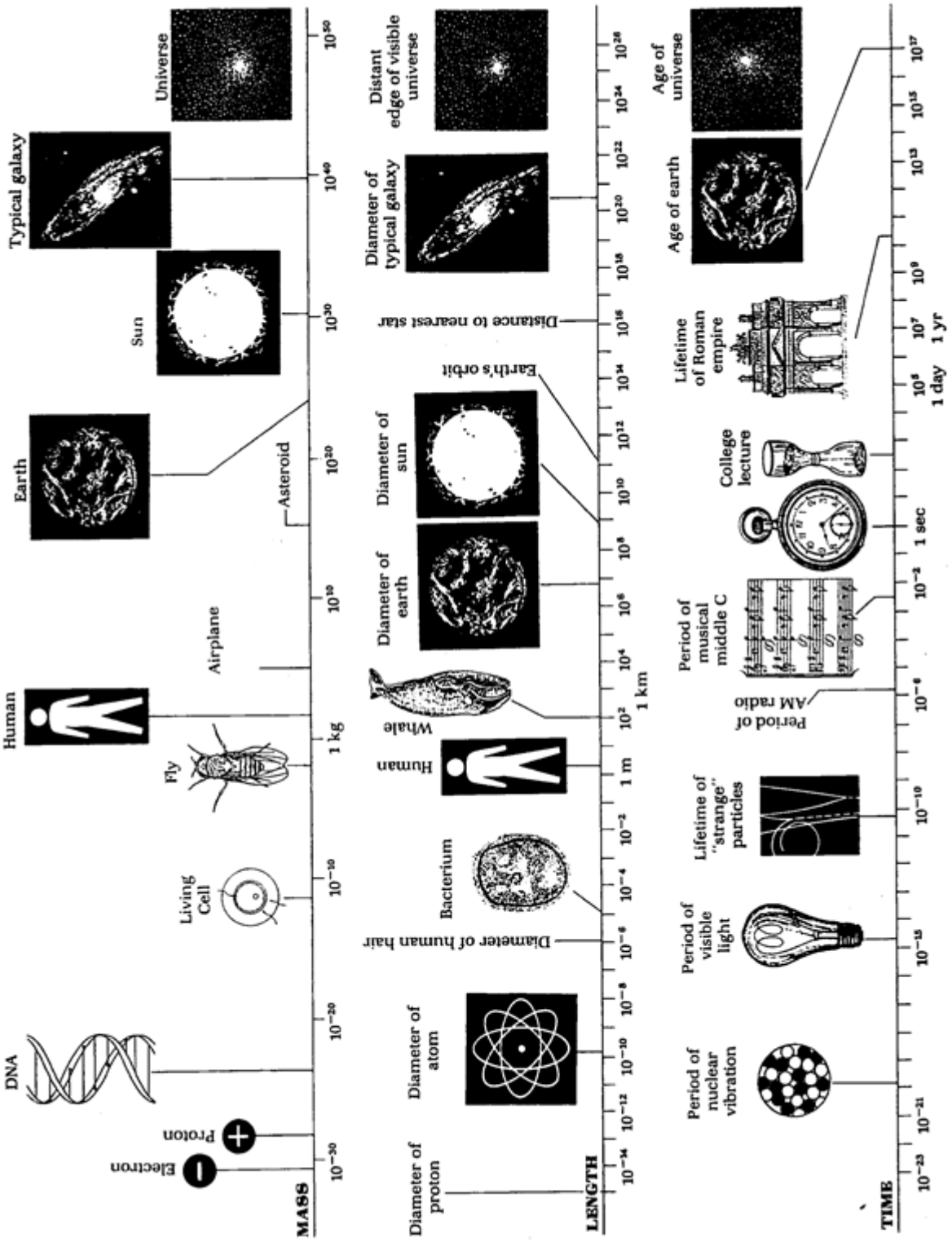
Chapter 1

Measurement



An indication of the range over which physical measurements are made.

TOPIC 1: MEASUREMENT

H2 Physics Syllabus 9749

SECTION I: Measurement

Measurement	**Learning Outcomes** Students should be able to:
Physical quantities and SI units	(a) recall the following base quantities and their SI units: mass (kg), length (m), time (s), current (A), temperature (K), amount of substance (mol). (b) express derived units as products or quotients of the base units and use the named units listed in 'Summary of Key Quantities, Symbols and Units' as appropriate. (c) use SI base units to check the homogeneity of physical equations. (d) show an understanding of and use the conventions for labelling graph axes and table columns as set out in the ASE publication *Signs, Symbols and Systematics* (*The ASE Companion to 16–19 Science*, 2000). (e) use the following prefixes and their symbols to indicate decimal sub-multiples or multiples of both base and derived units: pico (p), nano (n), micro (μ), milli (m), centi (c), deci (d), kilo (k), mega (M), giga (G), tera (T). (f) make reasonable estimates of physical quantities included within the syllabus.
Scalars and vectors	(g) distinguish between scalar and vector quantities, and give examples of each. (h) add and subtract coplanar vectors. (i) represent a vector as two perpendicular components.
Errors and uncertainties	(j) show an understanding of the distinction between systematic errors (including zero error) and random errors. (k) show an understanding of the distinction between precision and accuracy. (l) assess the uncertainty in a derived quantity by addition of actual, fractional, percentage uncertainties or by numerical substitution (a rigorous statistical treatment is not required).

Playlist of teaching videos and lecture examples:
https://youtube.com/playlist?list=PL_b5cjrUKDlY0aQFdGFfOKuFJQTiXzIeY

Contents

TOPIC 1: MEASUREMENT 2
1.1 Introduction 4
1.2 Physical Quantities and Units 4
1.2.1 Base Quantities and Units 5
1.2.2 Derived Quantities and Units 5
1.2.3 Homogeneity of equations 7
1.2.4 Prefixes 9
1.3 Estimation 10
1.4 Errors and Uncertainties 12
1.4.1 Significant Figures and Decimal Places 14
1.4.2 Combining Uncertainties 15
1.4.3 Precision and Accuracy 19
1.4.4 Systematic and Random Errors 21
1.5 Scalars and Vectors 26
1.5.1 Adding and Subtracting Coplanar Vectors 26
1.5.2 Resolution of Vectors 322
Appendix I International System of Units (SI) 35
Appendix II Summary of Key Quantity, Symbols and Units 37
Appendix III SI Prefixes 39
Tutorial 1 Measurement 40
Self-Review Questions 40
Discussion Questions 42

1.1 Introduction

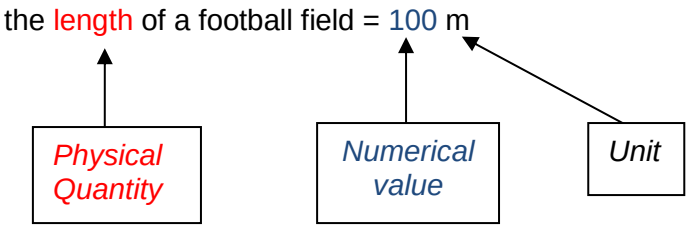
Measuring our observations objectively is an important task that differentiates science from the arts. The scientific method involves making observations that are repeatable by another person. For this reason, the study of physical quantities and their units is the starting point of our learning of science.

1.2 Physical Quantities and Units

A physical quantity is a physical property that can be measured and/or calculated and expressed in numbers. For example, "length" is a physical quantity that can be expressed by stating a number and a measurement unit such as metres or inches.

Hence ***the value of a physical quantity*** is expressed as ***a numerical value and a unit of measurement.***

For example,



This has a few implications.

1. Using the unit “metre” helps us to define a particular property of the football field – its length. Using other units could allow us to define other properties (for example, square metres describe the area) of a football field.
2. We can only compare properties of the same unit. Using the “metre” allows us to compare the length of one object with another. We cannot compare a “metre” with a “kilogram”.
3. The numerical value of the physical quantity is dependent on the unit it is expressed in. The length of the same football field above will have a value of 330 when expressed in the unit “foot”.

1.2.1 Base Quantities and Units

Through the centuries, scientists in different parts of the world have used different units at different times to measure the same physical quantities. However standardised definitions and units are needed in the modern world.

The current International System of Units[a], universally abbreviated SI (from the French *Le Système International d'Unités*), is the modern metric system of measurement.

Seven well-defined and independent base quantities are chosen. Their corresponding **SI base units** are indicated:

Base Quantity		**SI Base Unit**	
Name	Symbol	Name	Symbol
Mass	m	kilogram	kg
Length	l	metre	m
Time	t	second	s
Electric current	I	ampere	A
Thermodynamic temperature	T	kelvin	K
Amount of substance	n	mole	mol
Luminous intensity	I_v	candela	cd

Note:

1. The **units**, when **written in full**, are **in small letters**. Ampere and Kelvin are names of scientists.
2. The kilogram is the only SI unit with a prefix as part of its name and symbol.
3. Luminous intensity is not in the Learning Objectives(a) of the Syllabus.
4. Definitions of the units are not required in the Syllabus but are provided in Appendix I at the end of the lecture notes for your reference.

IMPORTANT TO KNOW

SI base units were chosen based on their mutual independence. They are the simplest units that cannot be expressed in terms of other SI base units.

1.2.2 Derived Quantities and Units

There are a lot of other physical properties that we come across in our daily lives. These quantities and units used scientifically are included in the SI as derived quantities and units.[b] Note that these derived units, such as newton and joule, are also SI units.

IMPORTANT TO KNOW

Derived quantities are related to the base quantities through mathematical and scientific equations. Derived units (the units of derived quantities) are defined in terms of the base units using the same equations.

One simple example is the derived quantity of area. It is defined using the mathematical equation area = length × breadth, where both length and breadth are defined in terms of the base quantity, length. Hence the unit for area can be written in terms of base units as m^2.

[a] The SI was established in 1960 by the 11th General Conference on Weights and Measures (CGPM, Conférence Générale des Poids et Mesures). (http://www.bipm.org/en/si/).
[b] There are some commonly used units that are not in the SI. Some examples are the minute, hour, day, degree, litre, tonne, electron-volt, unified atomic mass unit, nautical mile, knot, hectare, bar and angstrom.

Some derived units have special names and symbols. An example is the newton (symbol N), a unit for force. We can define this derived quantity - force, using formulae such as $F = ma$, where F is force, m is mass and a is acceleration. You may note that acceleration itself is also a derived quantity. Using such an equation, we can then find the representation of the newton (N) in terms of SI base units (kg m s^{-2}).

For more information on the common symbols and units that will be used in A level question papers, refer to Appendix II. As a practice, derive the SI base units for the quantities in the table below.

Derived quantities, symbols	**e.g. of mathematical relationship between quantities**	**SI base units**	**SI derived units**
Plane Angle	$s = r\theta$ s : arc length of circle r : radius of circle θ : angle subtended by arc		radian
Density, ρ	$\rho = \frac{M}{V}$ M : mass of body V : volume of body ρ : density	kg m^{-3}	
Force, F	$F = ma$ m : mass a : acceleration	kg m s^{-2}	N (newton)
Momentum, p	$p = mv$	kg m s^{-1}	N s (newton-second)
Pressure, p	$p = F/A$ A : area	kg m^{-1} s^{-2}	Pa (pascal)
Energy, E	$KE = ½ mv^2$ v : velocity	kg m^2 s^{-2}	J (joule)
Power, P	$P = E/t$ E : energy t : time	kg m^2 s^{-3}	W (watt)
Electric charge, Q	$Q = It$ I : current t : time	A s	C (coulomb)
Electric potential difference, V	$V = W/Q$ W : work done Q : electric charge	kg m^2 s^{-3} A^{-1}	V (volt)
Resistance, R	$R = V/I$ V : electric potential difference I : current	kg m^2 s^{-3} A^{-2}	Ω (ohm)
Frequency, f	$f = 1/T$ f : frequency T : period	s^{-1}	Hz (hertz)

Example 1

According to Newton's Law of Gravitation, the force F between two point masses M and m separated by the distance r is given by the formula $F = \frac{GMm}{r^2}$ where G is the universal gravitational constant.
Obtain the SI base units for G. [Ans: m^3 s^{-2} kg^{-1}]

Common quantities quoted with <u>no unit of measurement</u>

1. all pure numbers, e.g. 2, $\frac{1}{2}$, π, e
2. trigonometrical functions, e.g. sine, cosine, tangent
3. all logarithmic functions, e.g. $\log_{10}$, ln
4. powers, e.g. 10^n, e^n, n has no unit.
5. physical constants with no unit of measurement: e.g. refractive index of glass, relative density of a liquid.

1.2.3 Homogeneity of equations

For an equation to be physically plausible, the equation must be *<u>homogenous</u>*. This is to say the terms on both sides of the equations must have the same units when expressed in SI base units.[c] This is because only quantities with the same units may be equated, added or subtracted.

<u>Example 2</u>
Bernoulli's equation, which applies to fluid flow states that

$$P + h\rho g + \tfrac{1}{2}\rho v^2 = k$$

where P is pressure, h is height, ρ is density, g is acceleration due to gravity, v is velocity and k is a constant.

Show the LHS of the equation is homogeneous and state the SI unit for k. [Ans: kg m^{-1} s^{-2}]

[c] An equation where terms on both sides of the equation have the same SI base units also imply that they have the same dimensions. However, knowledge of dimensions is not necessary for 9749 syllabus. See Appendix I (Dimensions of physical quantities) if you wish to understand further.

Checking of homogeneity of an equation (by comparing the SI base units of the left and right hand side terms of the equation) is not only a powerful way of establishing if an equation is reasonable, but it also provides hints for guessing the actual equation.

Example 3
Consider the period *T* of a simple pendulum. The possible factors which may affect it are its length *l*, its mass *m* and the acceleration due to gravity *g*. Use unit analysis to arrive at a plausible relationship between *T* and these quantities.

A physically correct equation must be homogeneous. However, a homogeneous equation need not be physically correct. Ultimately, the validity of an equation can only be verified through experiments.

There are two basic reasons:

(1) The value of the coefficient may be incorrect.

e.g. $E = 3mv^2$ where E = kinetic energy

The coefficient 3 is incorrect. The value should be $\frac{1}{2}$ *instead.*

(2) Missing or extra terms that may have the same unit.

e.g. $E = \frac{1}{2}mv^2 + mgh$ where E = kinetic energy

There is an extra term mgh, which happens to have the same base unit as kinetic energy. This is an extra term.

1.2.4 Prefixes

Prefixes can be used with both base units and derived units. The rationale for prefixes is simple. While it is alright to write your height as 1.65 m, it will be quite cumbersome to write the width of your hair as 0.00005 m, or the size of an atom as 0.0000000001 m. In mathematics, you have learnt the use of standard form. The size of an atom can be rewritten more neatly as 1×10^{-10} m but scientists often use prefixes instead of standard form for values between 10^{-12} and 10^{12}.

Factor	Prefix	Symbol	Name	Decimal equivalent	Order of magnitude[d]
10^{-12}	pico	p	Trillionth	0.000,000,000,001	-12
10^{-9}	nano	n	Billionth	0.000,000,001	-9
10^{-6}	micro	μ	Millionth	0.000,001	-6
10^{-3}	milli	m	Thousandth	0.001	-3
10^{-2}	centi	c	Hundredth	0.01	-2
10^{-1}	deci	d	Tenth	0.1	-1
10^{0}	-	-	One	1	0
10^{3}	kilo	k	Thousand	1,000	3
10^{6}	mega	M	Million	1,000,000	6
10^{9}	giga	G	Billion	1,000,000,000	9
10^{12}	tera	T	Trillion	1,000,000,000,000	12

The accepted convention is to use a prefix such that the quantity can be written as a whole number and of least significant figures (s.f.). For example, the wavelength of red light would be written as 650 nm, rather than 0.65 μm. Values larger than the ranges listed in the prefixes above should be written in standard form. (The above list of prefixes are stipulated in LO(e). Other prefixes are listed in Appendix III).

You may notice that the symbol for the prefix milli (m) is the same as that for the base unit metre (m), leading to a confusion when we see, for example, "ms^{-1}". It could mean "metre per second", or "per millisecond". To distinguish the units especially in print, the A-level standard is to **use "ms^{-1}" for "per millisecond" and leave a space "$m\ s^{-1}$" for the derived unit "metre per second".**

Further reference:

Powers of Ten video
https://www.youtube.com/watch?v=0fKBhvDjuy0

The Scale of the Universe 2 video
https://www.youtube.com/watch?v=uaGEjrADGPA

[d] http://en.wikipedia.org/wiki/Order_of_magnitude

1.3 Estimation

When was the last time you estimated something?

Physicists frequently use “back-of-the-envelope” calculations or “Fermi” problems, named after Physicist Enrico Fermi who worked on the Manhattan Project during World War II. Fermi was known for making approximate calculations with little or no actual data. One well-documented example was his estimate of the strength of the atomic bomb based on the distance travelled by bits of paper dropped from his hand at the test blast.[e]

Estimation in Physics uses simple numbers (1, 2 or 5) with the correct order of magnitude (10^{3} or 10^{-4}). It is important to know whether a lecture theatre can sit a few students, tens, hundreds or thousands of students, but not so important to know that it has a capacity of 327. When the exact answer or values are not known, assumptions and estimation are used to find a rough answer.

Estimation is also very useful as a check to the answer that we would get from complex calculations.

Examples of “Fermi” problems:

- Estimate the total number of hairs on your head.
- How many bricks are needed to build a home?
- Estimate the efficiency of an electric kettle.
- How much paper is used by Hwa Chong Institution for lecture notes in 1 year?

Some important estimates that you should memorize:

- Typical wavelength of visible light: 10^{-7} m (range: 400 to 700 nm)
- Range of wavelengths of other electromagnetic regions
- Size of an atom: 10^{-10} m or 1 angstrom (Å)
- Size of a nucleus: 10^{-15} m

Strategy for estimation:

1. Identify the unknown: Define specifically what you need to estimate
2. Identify the known: Find your experience that may help you relate to the unknown
3. Find a relationship between the known and the unknown: make a connection between what you know and what needs to be estimated.

[e] http://en.wikipedia.org/wiki/Fermi_problem

Example 4 [N09/I/2]

Which estimate is realistic?

A The kinetic energy of a bus travelling on an expressway is 30 000 J.
B The power of a domestic light is 300 W.
C The temperature of a hot oven is 300 K.
D The volume of air in a tyre is 0.03 m^3.

Further reference:

MIT OpenCourseWare Video – the Art of Approximation
https://youtu.be/X8DlaW83HJc

1.4 Errors and Uncertainties

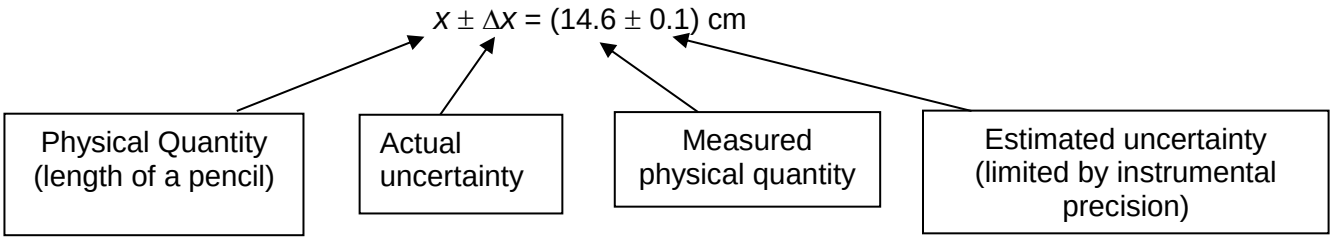
Whenever we attempt to make a measurement of a physical quantity, we are prone to all sorts of errors. As such, it is actually quite difficult, maybe impossible, to obtain a true value of the physical quantity. This can be a serious problem. It is essential in any experiment, in order to check the validity of a hypothesis, to be confident of our measurements, so that if the value of a physical quantity obtained in the experiment differs from the value predicted by the hypothesis, we can reject the hypothesis.

A measurement of a physical quantity X is reported in the form, $X = (x \pm \Delta x)$ where Δx refers to the uncertainty associated with the measurement of X.

It is to be interpreted as we are pretty confident that the true value of X lies in the interval $(x - \Delta x, x + \Delta x)$. If the predicted value falls within this interval, then we have no evidence to reject the hypothesis. However, if the predicted value falls outside this interval, we may claim that we have evidence to reject the hypothesis.

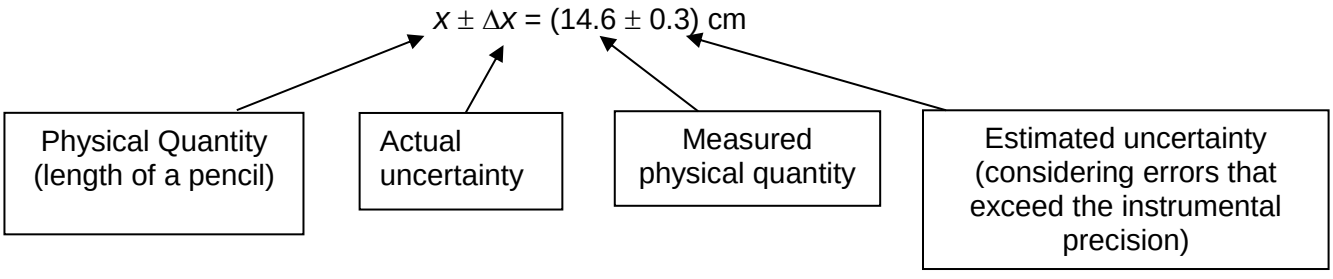
Two scientists who performed their experiments independently in an attempt to measure the quantity X will also say that their values of X are consistent with each other if the two intervals overlap. If the intervals established by each scientist do not overlap, then their values of X are inconsistent.

In attempting to measure X, we will naturally strive to do our best and make several measurements. But let's first suppose that we only make one measurement. In such a situation, the precision of the instrument we used limits our ability to obtain the true value. For example, when using a metre rule to measure the length of a pencil, we may at most be certain that our measured value is right up to the nearest mm, as the smallest division on the metre rule is 1 mm. We will hence report our measurement as $X = (14.6 \pm 0.1)$ cm.



However, there could be other sources of error that we did not realise. With only a single measurement, we will not be able to tell whether we might get a different reading the next time we tried and how far off that reading will be. As such, even though we have established an interval based on just one reading, it is entirely possible that the true value could fall outside this interval.

If we are aware of the presence of experimental errors other than instrumental errors, we could provide for it. For example, when we attempt to measure the length of the pencil, the butt of the pencil was broken with jagged edges, and it is difficult to ascertain the edge very precisely. As such, we could take that into consideration and report our measurement as $X = (14.6 \pm 0.3)$ cm with a reasonable subjective estimated uncertainty of 0.3 cm.



Ideally, we should make more attempts at measuring *X*. If we make multiple attempts, the spread of the data could give us a better idea of the impact of errors that could be present and help us to make a more realistic guess at the interval which could contain the true value. As a rule of thumb, a reasonable estimate would be to determine the uncertainty as $\Delta x = \frac{1}{2}(x_{max} - x_{min})$. We would also use the mean $\langle x \rangle$ of our readings as the true value of *X*. Hence we will report our measurement of *X* as

$$X = \langle x \rangle \pm \frac{x_{max} - x_{min}}{2}$$

	1st attempt	2nd attempt	3rd attempt
Length of pencil / cm	14.6	14.3	14.4

$$X = \left(\frac{14.6 + 14.3 + 14.4}{3}\right) \pm \left(\frac{14.6 - 14.3}{2}\right)$$
$$= (14.4 \pm 0.2) \text{ cm}$$

$x \pm \Delta x = (14.4 \pm 0.2)$ cm

- Physical Quantity (length of a pencil)
- Actual uncertainty
- Mean value of multiple readings
- Estimated uncertainty (considering max-min values)

You would realise by now that there is some level of arbitrariness in declaring the uncertainty associated with our measurement. Nevertheless, while declaring a larger uncertainty will result in a larger interval that is more likely to contain the true value, remember that too large an interval will also cause any subsequent conclusions derived from the experiment to be meaningless.

In practice, many attempts are made and we have a very large set of data and a statistical approach is used to establish the uncertainty but we will not go into that as it is not in our syllabus.

Fractional uncertainty

The **fractional uncertainty** is the ratio of the actual uncertainty to the measured value.

The fractional uncertainty in a quantity x is $\frac{\Delta x}{x}$.

Example: The fractional uncertainty of x, $\frac{\Delta x}{x} = \frac{0.2}{14.4} = 0.014$.

Percentage uncertainty

The **percentage uncertainty** is obtained by converting the fractional uncertainty into percentage form by multiplying by 100%.

The percentage uncertainty in a quantity x is $\frac{\Delta x}{x} \times 100\%$.

Example: The percentage uncertainty of x, $\frac{\Delta x}{x} \times 100\% = \frac{0.2}{14.4} \times 100\% = 1.4\%$.

1.4.1 Significant Figures and Decimal Places

The general rules are:

1. **Express uncertainties to 1 s.f.**
   If the uncertainty in a measurement is estimated to be 0.025 for example, it should be rounded to 0.03.

2. **Express the quantity to the same place value as its uncertainty.**
   Using the same example, 0.03 has its most significant digit in the second decimal place. Hence the quantity should be written to exactly two decimal places:

   12.10 ± 0.03

   and NOT 12.1 ± 0.03 or 12.102 ± 0.03.

Example 5

A student makes measurements from which he calculates the speed of sound to be 327.66 m s^{-1}. He estimates that the percentage uncertainty is 3%. Round off the speed to an appropriate number of significant figures. [Ans: 330 m s^{-1}]

1.4.2 Combining Uncertainties

There are established statistical rules for propagation of uncertainty from individual pieces of information. The A-level course only requires a simplified version of the statistical treatment.

(i) Adding or Subtracting Measured Quantities

When two measured quantities are added together or one subtracted from another, the actual uncertainty in the result is equal to the sum of the actual uncertainty of the two quantities.

We always add up the uncertainties even if the equation involves subtraction because we do not know the sign of the actual uncertainty and thus need to estimate the *largest* possible uncertainty in our quantities.

Rule 1: For $c = a + b$, the actual uncertainty in c, $\Delta c = \Delta a + \Delta b$
For $d = a - b$, the actual uncertainty in d, $\Delta d = \Delta a + \Delta b$

(ii) Multiplying or Dividing Measured Quantities

When two measured quantities are multiplied together or one divided by the other, the fractional uncertainty in the result is equal to the sum of the fractional uncertainties of the two quantities.

Rule 2: For $p = ab$, the fractional uncertainty in p, $\dfrac{\Delta p}{p} = \dfrac{\Delta a}{a} + \dfrac{\Delta b}{b}$

For $q = \dfrac{a}{b}$, the fractional uncertainty in q, $\dfrac{\Delta q}{q} = \dfrac{\Delta a}{a} + \dfrac{\Delta b}{b}$

Here, percentage uncertainty can be used interchangeably with fractional uncertainty as they differ only by a factor of 100 which is multiplied equally to every term in the equation. The percentage uncertainty in p is equal to the sum of the percentage uncertainties in a and b,

$$\frac{\Delta p}{p} \times 100\% = \frac{\Delta a}{a} \times 100\% + \frac{\Delta b}{b} \times 100\%.$$

Let us use a simple equation to illustrate the analysis of most mathematical or scientific equations we will encounter. Consider the linear equation $y = mx + c$, where m, x and c are all measurements with their associated uncertainties. Our experimental result, y, will also have an uncertainty that depends on that of m, x and c.

The addition is simple: $\Delta y = \Delta(mx) + \Delta c$ but for the product mx, $\dfrac{\Delta(mx)}{(mx)} = \dfrac{\Delta m}{m} + \dfrac{\Delta x}{x}$. Rearranging,

$\Delta(mx) = \left(\dfrac{\Delta m}{m} + \dfrac{\Delta x}{x}\right)(mx)$ and hence $\Delta y = \left(\dfrac{\Delta m}{m} + \dfrac{\Delta x}{x}\right)(mx) + \Delta c$.

Example 6

The measurements of the dimensions of a particular piece of rectangular cardboard are (18.5 ± 0.5) mm and (12.5 ± 0.5) mm. Determine the area of the cardboard with its associated uncertainty.
[Ans: $(2.3 \pm 0.2) \times 10^2$ mm^2]

(iii) Scaling

Actual uncertainty in a measured quantity is *scaled together* with the measured quantity.

Rule 3: If $r = ka$, then $\Delta r = k(\Delta a)$ where k is a constant

E.g. if $r = \frac{1}{4}b$ then $\Delta r = \frac{1}{4}(\Delta b)$

Think of the above as a special case of Rule 1. For instance, $r = 3a = a + a + a$
For the case of dividing measurement by a numerical constant k, think of it as multiplying the inverse, $1/k$.

Note: the fractional uncertainty $\frac{\Delta r}{r}$ is equal to $\frac{\Delta a}{a}$, as the constant k is cancelled out in the ratios.

Example 7

The radius of a circle is $r = (3.0 \pm 0.2)$ cm. Find the circumference with its uncertainty. [(19 ± 1) cm]

(iv) Powers

This applies to all exponents, n, both larger and smaller than 1, whether integer or fraction. However, if n is negative, the error is still considered as positive.

Rule 4: If $s = a^n$, then $\frac{\Delta s}{s} = |n| \frac{\Delta a}{a}$

E.g. if $s = b^{-½}$, then $\frac{\Delta s}{s} = \frac{1}{2} \frac{\Delta b}{b}$

Example 8

Given a sphere of radius r = (18.5 ± 0.5) mm, find the volume of the sphere with its associated uncertainty. [Ans: (2.7 ± 0.2) × 10^4 mm^3]

Example 9

In an experiment, the external diameter d_1 and internal diameter d_2 of a hollow tube are found to be (64 ± 2) mm and (47 ± 1) mm respectively. Calculate the thickness of the tube and the associated uncertainty. What is the corresponding percentage uncertainty? [Ans: (9 ± 2) mm; 22 %]

Example 10

The period of oscillation of a simple pendulum is given by $T = 2\pi\sqrt{\frac{l}{g}}$. A student conducts an experiment to find the acceleration of free fall, *g*. He measures the length of the pendulum, *l* = 0.23 ± 0.01 m, and the time for 20 oscillations, *t* = 19.24 ± 0.01 s. Find *g* and its associated uncertainty. [Ans: (9.8 ± 0.4) m s^{-2}]

(v) Special functions or complicated functions, X

If measurements are put together with trigonometric functions, exponential functions or other complicated formulae, use simple numerical substitutions to evaluate the uncertainty directly. The following rule may be applied.

Rule 5:
Actual uncertainty of X = ½ (maximum possible X – minimum possible X)
$\Delta X = \frac{1}{2}(X_{max} - X_{min})$

Example 11

Consider S = *x* cos θ for *x* = (2.0 ± 0.2) cm, θ = (53 ± 2)°. Find S with its uncertainty.

Summary of Rules for propagation of uncertainties derived from measured values A ± ΔA and B ± ΔB

1. If $R = mA + nB$, then $\Delta R = |m|\Delta A + |n|\Delta B$
2. If $R = kA^m \times B^n$, where k is a numerical constant, then $\frac{\Delta R}{R} = |m|\frac{\Delta A}{A} + |n|\frac{\Delta B}{B}$
3. Other functions: e.g. $R = \sin A$, $R = \ln B$

Use the general approach: $\Delta R = \frac{1}{2}(R_{max} - R_{min})$

1.4.3 Precision and Accuracy

These are two terms which we need to understand in our approach to measurements.

Precision is a measure of how well a result can be determined (without reference to a theoretical or true value). It is the degree of consistency and agreement among independent measurements of the same quantity.

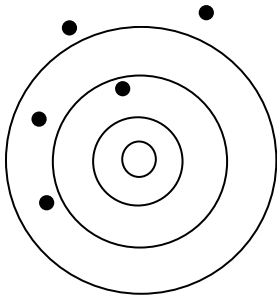
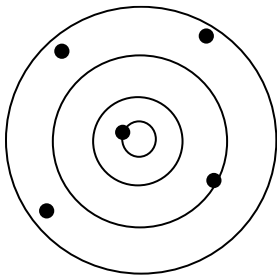
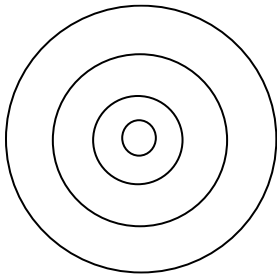
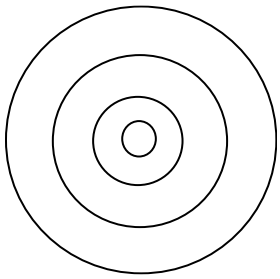
Measurements are said to be precise when *repeated* measurements remain very close to one another. If a certain measurement which is done, repeated several times, produces widely varying readings, the precision is poor.
For example, three measurements of my height: 1.75 m, 1.76 m, 1.75 m are precise but 1.72 m, 1.78 m, 1.76 m are not precise.

Accuracy is the closeness of agreement between a measured value and a true or accepted value.

For example, if we are told that an object has a true mass 500 g, a measured value of 400 g is inaccurate. Similarly, the measurement of the acceleration of free fall, g = 9.40 m s^{-2} is also inaccurate, compared to g = 9.70 m s^{-2}.

Example 12: Check your understanding

Complete the diagram below with appropriate ticks and crosses.

Target	Precise	Accurate
		
		
	✓	✗
	✓	✓

Further reference:

Rules for calculating uncertainties ~xmphysics0 (summary and examples)
https://www.youtube.com/watch?v=Ddfgq8bjb7E

Proof of product rule (beyond the syllabus) ~xmphysics0
https://www.youtube.com/watch?v=GP8r3E529kI

1.4.4 Systematic and Random Errors

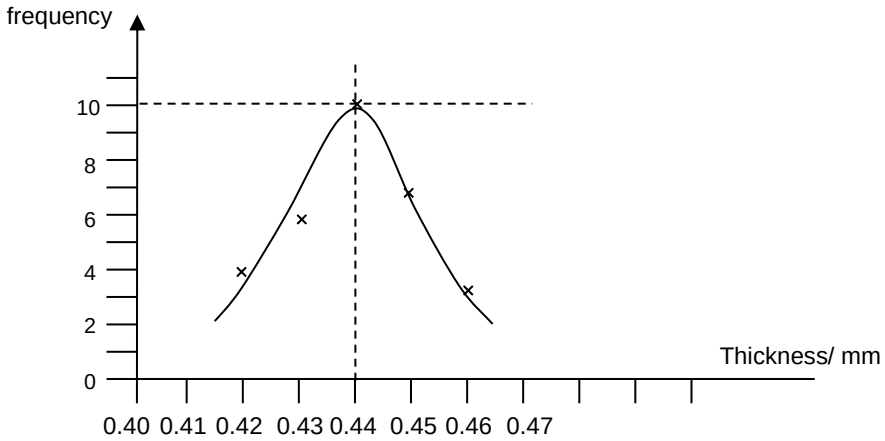
We have asserted that the true and exact value of a physical quantity can never be known. This is because our measurement of a physical quantity will always be limited by experimental errors. The causes of these experimental errors can be split into two broad categories – systematic errors and random errors.

1.4.4.1 Random Error

Suppose we take many measurements to determine the diameter of a thin piece of wire using a micrometer screw gauge and tabulate our measurements in a table.

Readings / mm	Frequency (no. of times of occurrence)
0.42	4
0.43	6
0.44	10
0.45	7
0.46	3
Total no. of readings	30

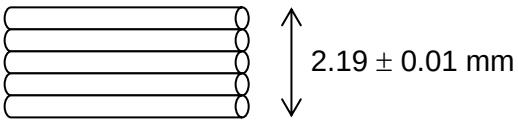
The distribution of the measurements can be represented in the plot as shown.



A *scattering* of readings about the **mean** value of the measurements suggests the presence of random errors (deviations in varying magnitude & direction). Random errors can occur despite repeating the experiments under the same conditions. Random errors are caused by environmental conditions, irregularity of quantity being measured, and limitations of the measurement equipment or the observer.

There are two ways to manage random errors.

1. Improving the procedure to minimize uncertainty due to random errors.
E.g., instead of measuring the diameter of only one thin wire, an improvement in the experimental procedure can be done by lining up a few thin wires and stick them together side-by-side using scotch tape. We will then measure the combined diameter instead.



Now if we use the same micrometer screw gauge, the actual uncertainty of our measurement is still the same as ± 0.01 mm. What about the diameter of one thin wire?

Since we measured the combined diameter of five thin wires, we can write $5d = (2.19 \pm 0.01)$ mm. Hence $d = \frac{1}{5}(2.19 \pm 0.01) = (0.438 \pm 0.002)$ mm. The actual uncertainty in our result has decreased tremendously and the diameter can now be expressed to one more significant figure than before.

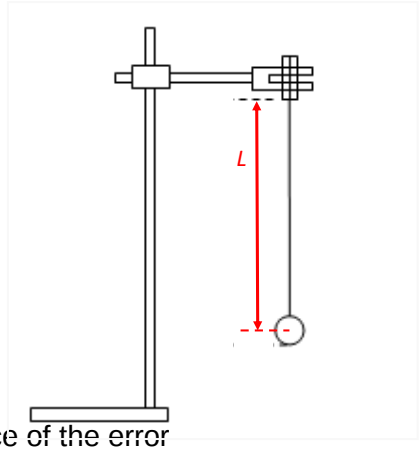
2. **A need for a large sample size in order for the mean value to be a good estimate of the true value**

When there are random errors present during measurement, due to the random nature of the errors, taking the mean of only a small number of repeated readings can result in a value that is still quite different from the true value. For the mean value to be a good estimate of the true value, a large number of repetitions is necessary. The more the better!

Example 13: Oscillations of a pendulum

The period of one oscillation of the pendulum T is related to the length of the pendulum L according to the equation $T = 2\pi\sqrt{\frac{L}{g}}$, where g = 9.81 m s^{-2}. A diagram of the experiment is shown on the right. T is measured for one oscillation with a stop watch and L is measured with a metre ruler held in hand.

L/cm	*T/s (1st reading)*	*T/s (2nd reading)*
22.0	1.16	0.98
23.5	1.07	1.23
25.1	1.15	1.19
25.6	1.30	1.10
27.3	1.42	1.20
28.1	1.38	1.25



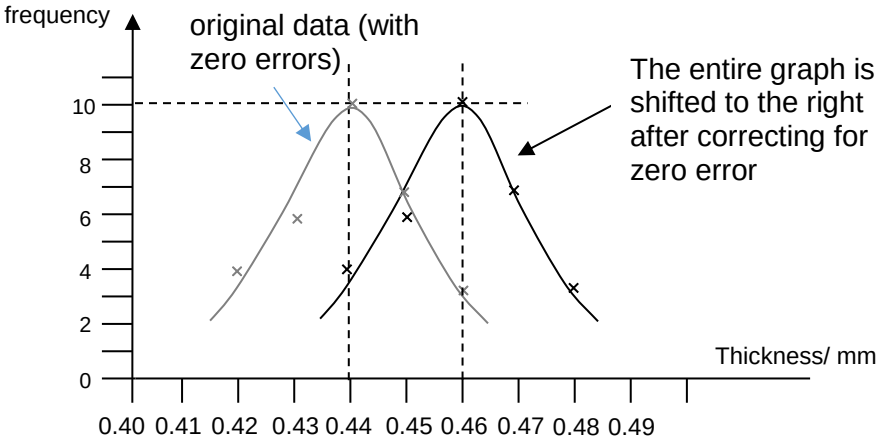
Using the given information,

a) identify a random error in this experiment and explain the source of the error
b) suggest an appropriate method to reduce the random error you have identified

1.4.4.2 Systematic Error

Suppose now we discover that the micrometer screw gauge used has a *zero error* of - 0.02 mm. Every single measurement is off by 0.02 mm.

The distribution of the readings would look like this:



This shift of values from the real values is the result of a *systematic error*. Systematic errors are reproducible errors which cause a set of readings to deviate in a fixed direction from the true value. Systematic errors are caused by instrumental errors, environmental conditions, and poor experimental techniques. The source of systematic errors can be determined and eliminated by corrective action.

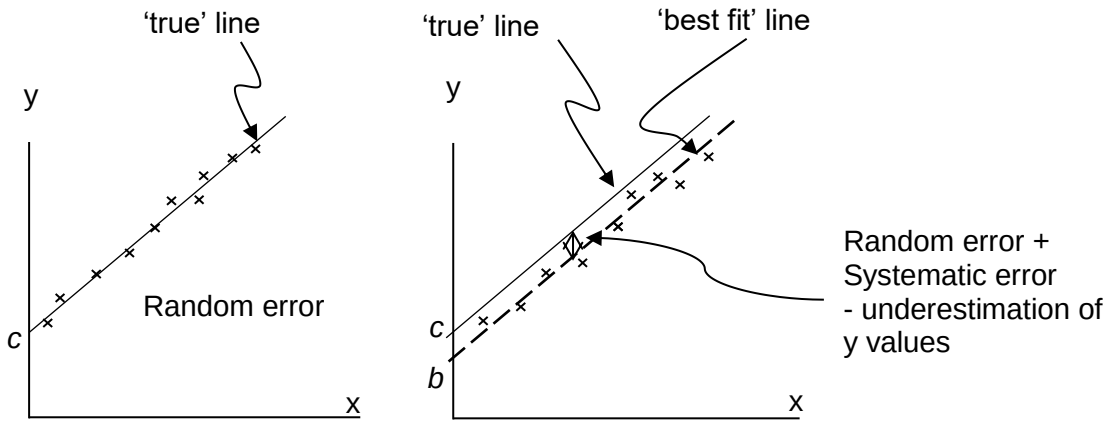
Question: Distinguish random errors from systematic errors.

Answer: Random errors are deviations of the measured value from the mean value, with varying signs and magnitudes. **Systematic errors** are deviations of the mean value from the true value, with same sign and similar magnitude.

Eliminating Systematic Error

In some cases, systematic error can be accounted for and corrected after it has been detected. The easiest way to detect systematic error is to note the vertical-intercept of the graph that is plotted. For example the graph of $y = mx + c$ when plotted gives a y-intercept of b instead of c as shown in the diagram below. The systematic error of $(c - b)$ results in an underestimation of the y- values.

The effects of random errors and of systematic errors appear in graphs as illustrated below.



Proper usage of measuring instruments is essential in eliminating systematic error. Measuring instruments need to be checked before usage. Always ensure balances, callipers, scale pointers, digital devices all read zero before being used. Calibration of instruments should be cross-checked if possible by comparing a measurement using two measuring devices.

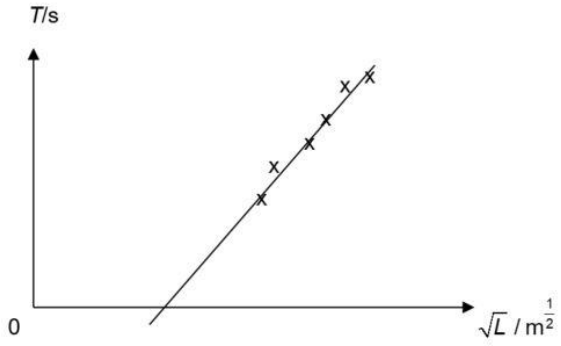
In many experiments, the most glaring error is due to the experimenter himself. Carelessness while taking measurements can result in severe systematic errors and random errors. However, blunders such as misreading a ruler or calculation mistakes should NOT be quoted as examples of errors in practical reports or exams!

Examples of Systematic Errors

Descriptions of errors	**Error Sources**	**Corrections**
• zero errors on the scales of instruments • poor calibration of instruments	Due to apparatus	• correct all measured readings by negating the error accordingly • calibrate the instrument properly before experiment
• consistent parallax error which affects all the readings in the same way, for instance, taking readings off a scale from a fixed angle	Due to poor experimental technique	• adopt the correct way to take reading: ensure that the *line of sight* is *perpendicular* to the measuring scale
• *background radiation* causes the count rate of your radioactive sample to be consistently higher than the true reading	Due to external factors	Take the external factor(s) into account and adjust all readings appropriately. For instance, • Measure the average background count rate and subtract it from the measured count rate.

Example 14: Oscillation of a pendulum

Refer to the Example 13 for the same setup. The graph of T against $\sqrt{L}$ is shown:



Using the given information,

(a) identify a systematic error in this experiment and explain the source of the error.

(b) suggest an appropriate method to eliminate the systematic error you have identified.

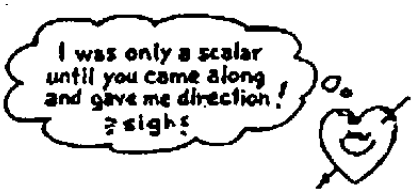
1.4.4.3 Summary Table

Type of error	Characteristics	If this error exists, will ***precision*** be affected?	If this error exists, will ***accuracy*** be affected?
random	varying in both magnitude and direction about a mean value, can be reduced by taking average of repeated readings, but not eliminated.	Yes	No
systematic	varying in fixed direction about a true value, CANNOT be reduced by taking average of repeated readings, but may be eliminated by good experimental techniques.	No	Yes

1.5 Scalars and Vectors

Physical quantities that have *no direction* associated with them are known as *scalars*. Scalars are specified completely by numerical values and units. Examples of scalars include distance, speed, mass, time, temperature, energy, gravitational and electric potentials.

Physical quantities that have both a magnitude and a *direction* are known as *vectors*. Examples of vectors include displacement, velocity, acceleration, force, gravitational and electric field strengths.

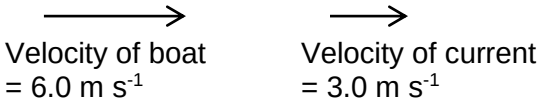


In books, a vector is often denoted in one of the following ways: $\boldsymbol{F}$, $\overrightarrow{F}$ or $\underset{\sim}{f}$. The magnitude is indicated as $|\boldsymbol{F}|$ or simply F.

The direction of vectors should always be clearly presented. In written form, we can express, for example, velocity as 5 m s^{-1} towards the east, or acceleration of free fall as 9.81 m s^{-2} downwards. This is somewhat troublesome to write when we come to diagonal motions.

Diagrams are always helpful in Physics when dealing with vector quantities. Each vector is represented by an arrow. The arrow is always drawn pointing in the direction of the vector quantity and the length of the arrow is proportional to the magnitude of the vector quantity. If more than one vector is drawn on a diagram, the lengths of the different arrows should be representative of the relative magnitudes of the vector quantities.

For example, a boat's engine can propel the boat to move at 6.0 m s^{-1} to the east and the current in the river is flowing at 3.0 m s^{-1} also to the east. The two vectors can be drawn as shown.

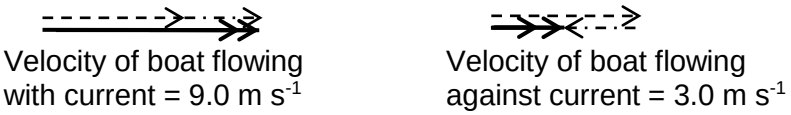


Note the relative lengths of the arrows: the velocity of the boat is twice that of the current, hence its arrow is also twice as long.

Multiplication of a vector by a scalar simply scales the length of the arrow. Multiplication of a vector by -1 (or a negative sign, "–") reverses the direction in which the vector points.

1.5.1 Adding and Subtracting Coplanar Vectors

It is obvious that the boat moves down the river faster with the presence of a current than without. In addition the boat will in fact move slower if it is moving upstream or against the flow of current. Hence there must be an additive and subtractive effect when there are two of more vector quantities present in one situation.



The sum of two or more vectors is called the *resultant* vector. The resultant velocity vector of the boat is 9.0 m s^{-1} when flowing with the current and 3.0 m s^{-1} when flowing against the current.
As demonstrated in the diagram above, resultant vectors are always found by connecting one arrow-head to the tail of the next arrow. The resultant vector is then drawn from the tail of the very first arrow to the arrow-head of the last arrow. This summation process can be repeated step-by-step for any number of vectors.

1.5.1.1 Vectors in 1D

Vectors in 1D can be added or subtracted via two methods.

The first method is via vector diagrams as we have described earlier. However, because in 1D, the directions of the vectors are limited to either forward or backward, we can actually add or subtract them just like how we add or subtract scalar quantities with the help of a sign convention.

By adopting a sign convention where the direction along the flow of river is taken as positive,
the velocity of the river current is then $v_{current}$ = +3 m s^{-1}.
The velocity of the boat traveling along the river is v_{boat} = +6 m s^{-1}.
Hence the velocity of the boat as seen by someone on the shore is given by
$v_{shore} = (+3) + (+6) = +9\ m s^{-1}$.

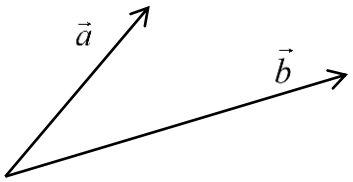
If the boat is traveling against the flow of river, using the same sign convention,
the velocity of the river current is then $v_{current}$ = +3 m s^{-1}.
the velocity of the boat traveling along the river is v_{boat}= - 6 m s^{-1}.
Hence the velocity of the boat as seen by someone on the shore is given by v_{shore}=(+3)+(-6) = -3 m s^{-1}.

Notice that we have managed to resolve the vector equation $\vec{v}_{shore} = \vec{v}_{boat} + \vec{v}_{current}$ without having to draw arrows.

1.5.1.2 Vectors in 2D

Two dimensional vectors lie on the same plane (coplanar) but point in different directions that are not along a single straight line.

In the diagram below, $\vec{a}$ and $\vec{b}$ are two coplanar vectors lying on the plane of the paper but not along a single straight line.



The tail of each arrow indicates the position where a vector quantity is acting upon. In the diagram above, if both $\vec{a}$ and $\vec{b}$ are forces, they would be acting on the same point, or the same object.

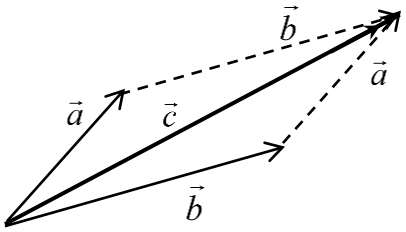
Imagine if you were pulled by two friends, one at each arm, in different directions. In which direction would you move? Can you possibly move in both directions you are pulled? There only exists one resultant force that will determine the direction you will move.

Similarly, the resultant vector $\vec{c}$ is obtained by adding vectors $\vec{a}$ and $\vec{b}$, or $\vec{c} = \vec{a} + \vec{b}$.

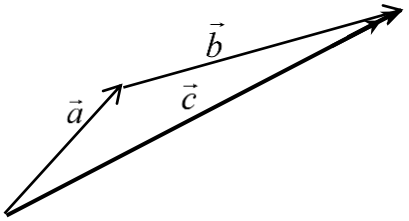
Addition of vectors: $\vec{c} = \vec{a} + \vec{b}$

Methods of Determining the Resultant Vector

a) Parallelogram Method: By putting the vectors to be added 'tail to tail'. Complete the parallelogram. The resultant vector is the diagonal from the tail of the two vectors to the other vertex of the parallelogram.



b) Triangle Method: By joining the end of the next arrow onto the tip of the previous one to form a chain of arrows, the resultant vector is a straight arrow that goes from the tail of the chain directly to the head. The order of adding the vectors does not affect the resultant.

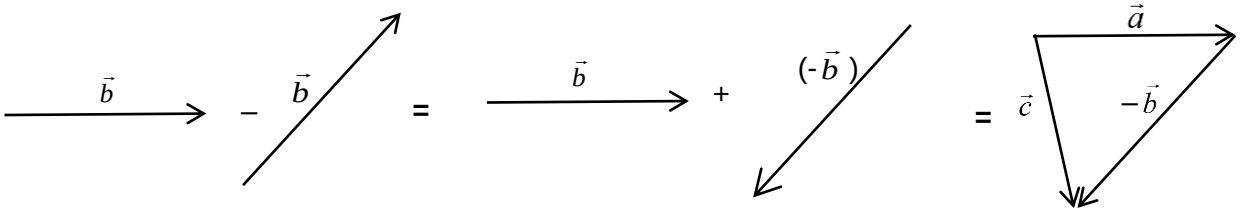


c) Component Method: The vectors are resolved into two perpendicular directions and then added. **This method is extremely important because it is the most convenient method to add three or more vectors.** (This method is described in section 1.5.2.)

Subtraction of vectors: $\vec{c} = \vec{a} - \vec{b}$

Subtraction of vectors can be evaluated by the same procedure as addition of two vectors since we can view subtraction of $\vec{b}$ from $\vec{a}$ as the summation of $-\vec{b}$ and $\vec{a}$.

$$\vec{c} = \vec{a} - \vec{b} = \vec{a} + (-\vec{b})$$



In Physics, vector subtraction can come about when we want to determine the ***change*** in a certain physical vector quantity. In general,

A change in a physical quantity = final value – initial value

$$\Delta\vec{x} = \vec{x}_{final} - \vec{x}_{initial}$$

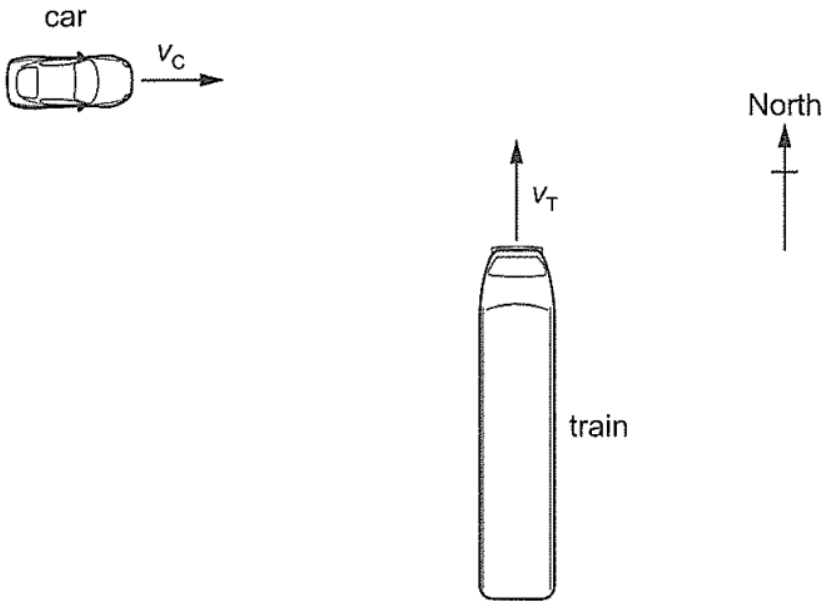
which is evaluated as: $\Delta\vec{x} = \vec{x}_{final} + (-\vec{x}_{initial})$

Further reference:

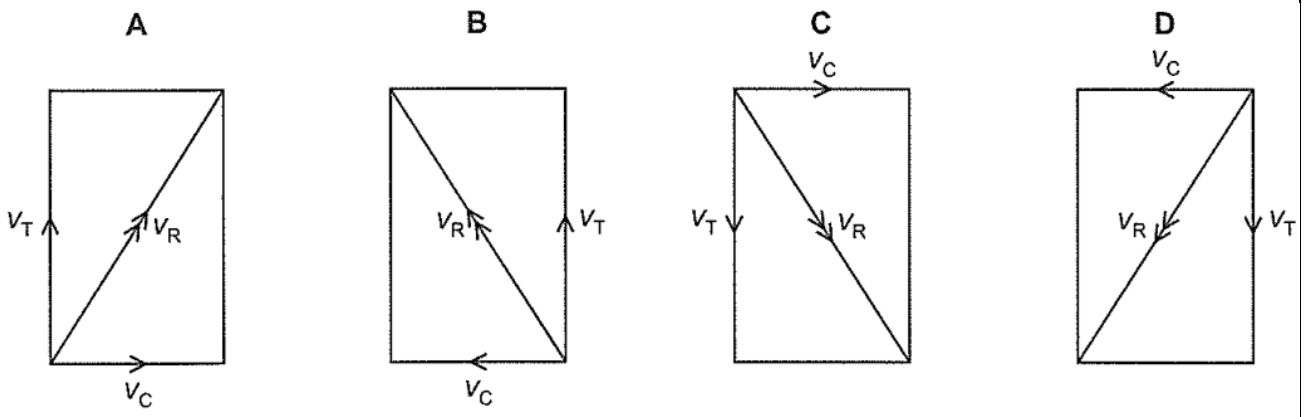
Vector addition and subtraction ~xmphysics0 (summary and examples)
https://youtu.be/FhoiORrgPFw

Example 15: H1/2016/P1/Q5

A passenger in a train travelling due north at speed v_T sees a car travelling due east at speed v_C.

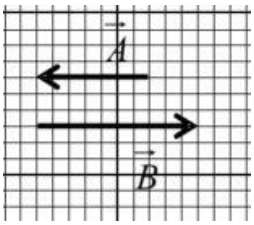


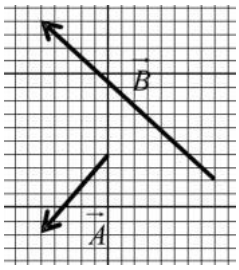
Which diagram shows the velocity v_R of the car relative to the passenger in the train?



Example 16 (a)

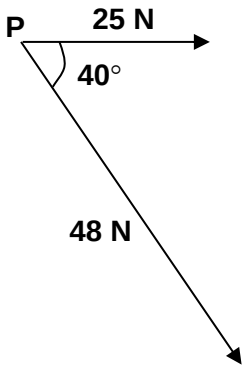
In the figure below, for each pair of vectors $\vec{A}$ and $\vec{B}$, draw the resultant vector $\vec{R}$ where $\vec{R} = \vec{A} + \vec{B}$.





Example 16 (b)

Two forces act at a point P as shown below. Determine (magnitude and direction of) the resultant, ***R***, of these two forces.



Example 16(c)

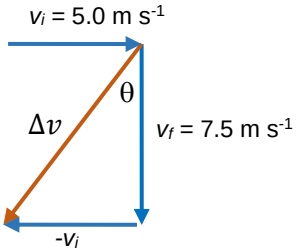
An object is moving at 5.0 m s^{-1} due east. Its direction changes to due south with a speed of 7.5 m s^{-1}. Determine (i) the change in speed) and (ii) the change in velocity.

Suggested solution

(i) change in speed $= 7.5 - 5.0 = 2.5\,\text{m s}^{-1}$

(ii) change in velocity

Component method



The x and y components of the initial and final velocities are

Initial velocity x-component: $v_{ix} = 5.0\,\text{m s}^{-1}$

Initial velocity y-component: $v_{iy} = 0.0\,\text{m s}^{-1}$

Final velocity x-component: $v_{fx} = 0.0\,\text{m s}^{-1}$

Final velocity y-component: $v_{fy} = -7.5\,\text{m s}^{-1}$

Change in velocity, $\Delta v = v_f - v_i$:

x-component: $\Delta v_x = v_{fx} - v_{ix} = 0.0 - 5.0 = -5.0\,\text{m s}^{-1}$

y-component: $\Delta v_y = v_{fy} - v_{iy} = -7.5 - 0.0 = -7.5\,\text{m s}^{-1}$

Magnitude of $\Delta v = \sqrt{\left(\Delta v_x^2 + \Delta v_y^2\right)} = \sqrt{\left((-5.0)^2 + (-7.5)^2\right)} = 9.0\,\text{m s}^{-1}$

The angle as show in in the diagram, $\theta = \tan^{-1}\left(\dfrac{\Delta v_x}{\Delta v_y}\right) = \tan^{-1}\left(\dfrac{5.0}{7.5}\right) = 33.7^o$

Trignometry method

From the diagram, we can find magnitude using pythagoras theorem

$\Delta v = \sqrt{\left((5.0)^2 + (7.5)^2\right)} = 9.0\,\text{m s}^{-1}$

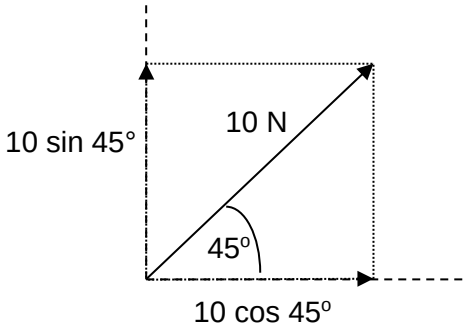
The angle as shown in the diagram, $\theta = \tan^{-1}\left(\dfrac{5.0}{7.5}\right) = 33.7^o$

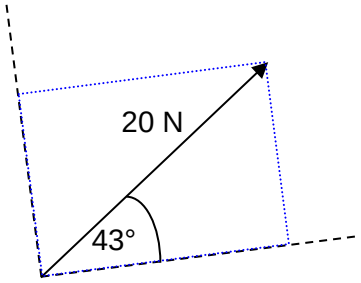
1.5.2 Resolution of Vectors

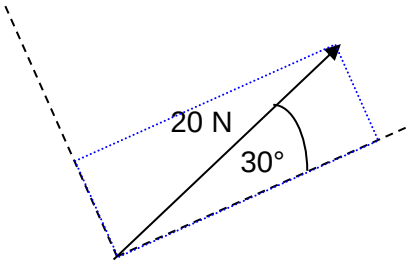
Since two vectors can be added to give a resultant vector, any vector can be broken up (or resolved) into two vectors or components. We will usually resolve a vector into two mutually-perpendicular components through the use of trigonometry and Pythagoras' theorem. Mutually-perpendicular vectors are independent of each other.

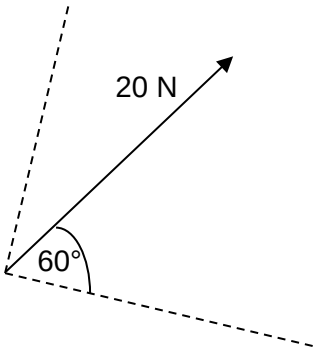
Practice
For each vector below, draw the two perpendicular components in the direction given by the dotted lines and state their magnitudes in terms of the respective values and angles.

The first example below has been done for you. Note that the dotted rectangle is an important working to account for the exact magnitudes of component vectors.









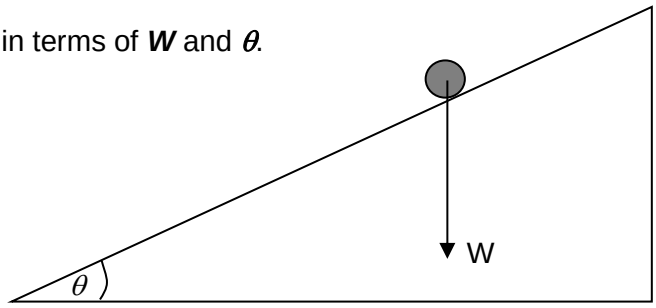
Each of the vectors above is resolved into two perpendicular components. A vector can be resolved into infinite pairs of perpendicular components. The choice of directions depends on the problem at hand.

Example 17

(a) An object rests on the plane of an inclined slope as shown. The weight ***W*** acts vertically down. Draw components of the weight

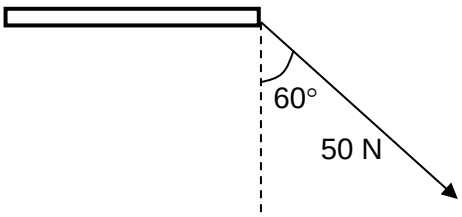
(i) parallel to the slope, W_P
(ii) perpendicular (normal), W_N, to the slope.

Label the magnitude of the two components in terms of ***W*** and θ.



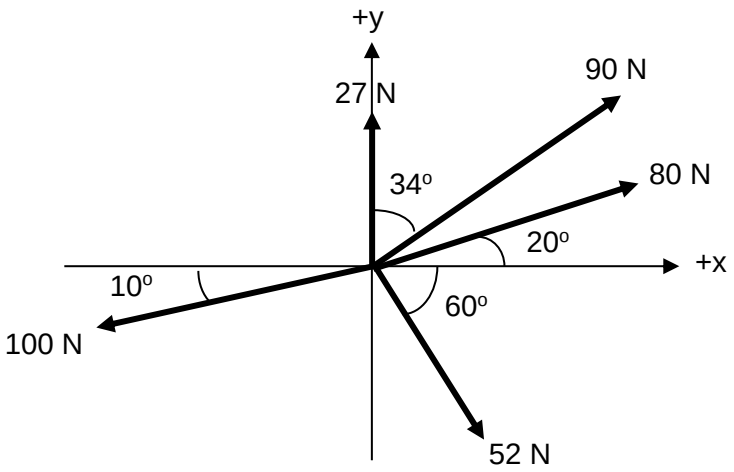
(b) A force of 50 N acts on a horizontal plank at angle of 60° to the vertical as shown. Draw components of this force (i) parallel to the plank, (ii) perpendicular to the plank.

Determine the magnitude of these components.



Example 18

Five forces shown act on an object. Find the resultant force.



- Resolve the vectors into two mutually perpendicular components:

Vector/N	x-component /N (+ →)	y-component /N (+↑)
80		
90		
27		
100		
52		
Resultant		

The magnitude of the resultant vector, $R =$

The direction of the resultant vector anti-clockwise from the positive x-direction,

$\theta =$

Further reference:

MIT OpenCourseWare Video – Vectors
https://youtu.be/mVQOmLTXLbQ

Further Reading and References

Appendix I International System of Units (SI)

The current International System of Units, universally abbreviated SI (from the French *Le Système International d'Unités*), is the modern metric system of measurement. This collection of units consists of seven defining constants[f], seven base units (from the seven defining constants), derived units (combinations of these seven base units) and a set of decimal-based multipliers used as prefixes. While there are units not included in the SI, the units in the SI are actually sufficient for use for all physical quantities known to us.

Base Quantity	SI Base Unit		
	Name	Symbol	Definition
length	metre	m	The metre is the length of the path travelled by light in vacuum during a time interval of 1/299 792 458 of a second.
mass	kilogram	kg	The kilogram, symbol kg, is the SI unit of mass. It is defined by taking the fixed numerical value of the Planck constant h to be $6.626\ 070\ 15 \times 10^{-34}$ when expressed in the unit J s, which is equal to kg m^2 s^{-1}, where the metre and the second are defined in terms of speed of light and Δv_{cs}.
time	second	s	The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium 133 atom. (Δv_{cs} = 9 192 631 770 hertz)
electric current	ampere	A	The ampere, symbol A, is the SI unit of electric current. It is defined by taking the fixed numerical value of the elementary charge e to be $1.602\ 176\ 634 \times 10^{-19}$ when expressed in the unit C, which is equal to A s, where the second is defined in terms of Δv_{cs}.
thermodynamic temperature	kelvin	K	The kelvin, symbol K, is the SI unit of thermodynamic temperature. It is defined by taking the fixed numerical value of the Boltzmann constant *k* to be $1.380\ 649 \times 10^{-23}$ when expressed in the unit J K^{-1}, which is equal to kg m^2 s^{-2} K^{-1}, where the kilogram, meter and second are defined in terms of *h*, *c* and Δv_{cs}.
amount of substance	mole	mol	The mole is the amount of substance of a system which contains exactly $6.022\ 140\ 76 \times 10^{23}$ elementary entities. This number is the fixed numerical value of the Avogadro constant, N_A, when expressed in the unit mol^{-1} and is called the Avogadro number. The amount of substance, symbol *n*, of a system is a measure of the number of specified elementary entities. An elementary entity may be an atom, a molecule, an ion, an electron, any other particle or specified group of particles.
luminous intensity	candela	cd	The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of 1/683 watt per steradian.

Adapted from the National Institute of Standards and Technology (NIST) Reference on Constants, Units and Uncertainty.

[f] CGPM announced the new definition of SI units from 7 defining constants on 16 Nov 2018. Details of the definitions are found here: https://www.bipm.org/en/measurement-units/base-units.html

Dimensions of physical quantities

The dimension of a quantity denotes the physical nature of the quantity. It only makes sense to add and subtract two quantities from each other when they are of the same nature.

In order for a mathematical equation that relates different physical quantities to be valid, the terms on both sides of an equation must have the same dimensions.

An equation where terms on both sides of the equation have the same dimensions is said to be homogeneous.

Dimensions of physical quantities can be treated as algebraic quantities. The dimensions of derived quantities can be determined using the same mathematical equations that relate the quantities.

In the SI, we have identified seven base quantities, each of them is of a different dimension (there are seven dimensions). All physical quantities of the same dimension have the same SI base unit. E.g., the length of a football field, a person's height, the thickness of a piece of paper, these physical quantities all have the same dimension of length. They all have the same SI base unit of metre.

To every dimension, we can associate an SI base unit. In this way, when we check the homogeneity of an equation by comparing the units of every term in the equation in terms of SI base units, we are also checking the dimensions of each term. If all the terms of the equation have the same unit in terms of SI base units, then the terms will have the same dimensions as well and hence the equation will be homogeneous.

Base Quantity		**SI Base Unit**		**Dimension**
Name	Symbol	Name	Symbol	Symbol
Mass	m	kilogram	kg	M
Length	L	metre	m	L
Time	t	second	s	T
Electric current	I	ampere	A	I
Thermodynamic temperature	T	kelvin	K	Θ
Amount of substance	n	mole	mol	N
Luminous intensity	I_v	candela	cd	J

Appendix II Summary of Key Quantity, Symbols and Units

(from GCE A-level syllabus)

SUMMARY OF KEY QUANTITIES, SYMBOLS AND UNITS

The following list illustrates the symbols and units that will be used in question papers.

Quantity	*Usual symbols*	*Usual unit*
Base Quantities		
mass	m	kg
length	l	m
time	t	s
electric current	I	A
thermodynamic temperature	T	K
amount of substance	n	mol
Other Quantities		
distance	d	m
displacement	s, x	m
area	A	m^2
volume	V, v	m^3
density	ρ	kg m^{-3}
speed	u, v, w, c	m s^{-1}
velocity	u, v, w, c	m s^{-1}
acceleration	a	m s^{-2}
acceleration of free fall	g	m s^{-2}
force	F	N
weight	W	N
momentum	p	N s
work	w, W	J
energy	E,U,W	J
potential energy	E_p	J
kinetic energy	E_k	J
heating	Q	J
change of internal energy	ΔU	J
power	P	W
pressure	p	Pa
torque	T	N m
gravitational constant	G	N kg^{-2} m^2
gravitational field strength	g	N kg^{-1}
gravitational potential	ϕ	J kg^{-1}
angle	θ	°, rad
angular displacement	θ	°, rad
angular speed	ω	rad s^{-1}
angular velocity	ω	rad s^{-1}
period	T	s
frequency	f	Hz
angular frequency	ω	rad s^{-1}
wavelength	λ	m
speed of electromagnetic waves	c	m s^{-1}
electric charge	Q	C
elementary charge	e	C
electric potential	V	V
electric potential difference	V	V
electromotive force	E	V
resistance	R	Ω
resistivity	ρ	Ω m
electric field strength	E	N C^{-1}, V m^{-1}
permittivity of free space	ε_o	F m^{-1}
magnetic flux	Φ	Wb

Quantity	*Usual symbols*	*Usual unit*
magnetic flux density	B	T
permeability of free space	μ_o	H m^{-1}
force constant	k	N m^{-1}
Celsius temperature	θ	°C
specific heat capacity	c	J K^{-1} kg^{-1}
molar gas constant	R	J K^{-1} mol^{-1}
Boltzmann constant	k	J K^{-1}
Avogadro constant	N_A	mol^{-1}
number	N, n, m	
number density (number per unit volume)	n	m^{-3}
Planck constant	h	J s
work function energy	Φ	J
activity of radioactive source	A	Bq
decay constant	λ	s^{-1}
half-life	$t_{1/2}$	s
relative atomic mass	A_r	
relative molecular mass	M_r	
atomic mass	m_a	kg, u
electron mass	m_e	kg, u
neutron mass	m_n	kg, u
proton mass	m_p	kg, u
molar mass	M	kg
proton number	Z	
nucleon number	A	
neutron number	N	

Appendix III SI Prefixes

Prefix	Symbol	Decimal	Power of ten	Order of magnitude
quecto-	Q	0.000000000000000000000000000001	10^{-30}	−30
ronto-	r	0.000000000000000000000000001	10^{-27}	−27
yocto-	y	0.000000000000000000000001	10^{-24}	−24
zepto-	z	0.000000000000000000001	10^{-21}	−21
atto-	a	0.000000000000000001	10^{-18}	−18
femto-	f	0.000000000000001	10^{-15}	−15
pico-	p	0.000000000001	10^{-12}	−12
nano-	n	0.000000001	10^{-9}	−9
micro-	μ	0.000001	10^{-6}	−6
milli-	m	0.001	10^{-3}	−3
centi-	c	0.01	10^{-2}	−2
deci-	d	0.1	10^{-1}	−1
-	-	1	10^{0}	0
deca-	da	10	10^{1}	1
hecto-	h	100	10^{2}	2
kilo-	k	1,000	10^{3}	3
mega-	M	1,000,000	10^{6}	6
giga-	G	1,000,000,000	10^{9}	9
tera-	T	1,000,000,000,000	10^{12}	12
peta-	P	1,000,000,000,000,000	10^{15}	15
exa-	E	1,000,000,000,000,000,000	10^{18}	18
zetta-	Z	1,000,000,000,000,000,000,000	10^{21}	21
yotta-	Y	1,000,000,000,000,000,000,000,000	10^{24}	24
ronna-	R	1,000,000,000,000,000,000,000,000,000	10^{27}	27
quetta-	Q	1,000,000,000,000,000,000,000,000,000,000	10^{30}	30

For use in Information Technology, prefixes for binary multiples have been adopted by the International Electrotechnical Commission (IEC).

Please refer to website for more information: http://physics.nist.gov/cuu/Units/binary.html

It appears that everyone has been wrong about the kilobyte so you can read it for yourself.

Tutorial 1 Measurement

Self-Review Questions

Use these questions to test your familiarity with the concepts for the topic. These questions should be sufficiently easy such that you can solve them on your own, with a little bit of thinking, without help from the tutors. The solutions to self-review questions are made available on Google Classroom for self-check. Thus your tutor may not go through these questions in class.

S1. The density of water is 1.00 g cm^{-3}. Express this value in kilograms per cubic metres (kg m^{-3}).

S2. [N17/I/1] Which list of SI units contains only base units?

A kelvin, metre, mole, ampere, kilogram
B kilogram, metre, second, ohm, mole
C kilogram, newton, metre, ampere, ohm
D newton, kelvin, second, volt, mole

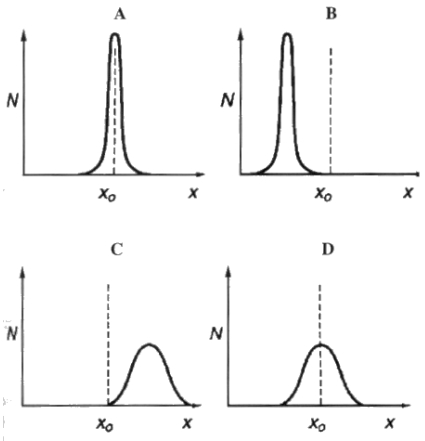
S3. The speed of a car cruising along PIE was 90.0 km h^{-1}. Express this value in metres per second.

S4. A light-year is a measure of length which is equal to the distance that light travels in 1 year. The distance from Earth to the star Proxima Centauri is 4.0 x 10^{16} m. Express this distance in light-years. (speed of light, c = 3.0 x 10^{8} m s^{-1})

S5. [modified N03/I/2] Errors in measurement may be either systematic or random.
Which one of the following involves random error?

A not allowing for zero error on a moving-coil voltmeter
B not subtracting background count rate when determining the count rate from a radioactive source
C stopping a stopwatch at the end of a sprint by a timekeeper
D using the value of g as 10 N kg^{-1} when calculating weight from mass

S6. [Adapted from N97/I/2] A quantity is measured many times and the number N of measurements giving a value x is plotted against x. The true value of the quantity is x_o. Fill in the table with ticks to describe the precision and accuracy of each graph.



Graph	**A**	**B**	**C**	**D**
Precise				
Not precise				
Accurate				
Not accurate				

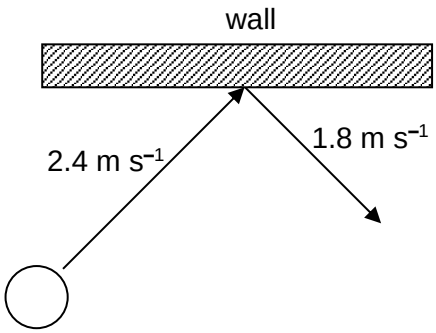
S7. [N20/II/1(b)] The drag force F acting on a cyclist travelling at speed v is given by $F = \frac{1}{2} C_D \rho A v^2$ where C_D is the drag coefficient, ρ is the density of air and A is the frontal area of the cyclist.

Data for the moving cyclist in shown in the table.

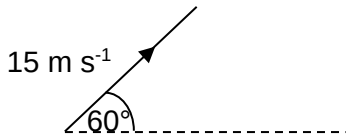
Quantity	**Magnitude**	**Uncertainty**
F / N	22	± 2
C_D	0.88	± 0.01
ρ / kg m^{-3}	1.2	± 0.1
A / m^2	0.32	± 0.02

Determine the speed of the cyclist, with its actual uncertainty. Give your answer to an appropriate number of significant figures.

S8. [2010 C1 LT1] Complete the vector diagram to show the vector representing the change in velocity of a billiard ball after rebounding from the edge of the table. Label the vector Δv.



S9. A stone is thrown with a velocity of 15 m s^{-1} at an angle of 60° to the horizontal as shown.



(a) (i) Explain why the diagram represents the velocity of the stone and not just its speed.

(ii) State whether the magnitude of the initial horizontal component of the velocity of the stone be greater, the same, or less than 15 m s^{-1}?

(b) Copy the diagram, and sketch the horizontal and vertical components of the velocity to correct proportion in magnitude.

(c) Calculate the magnitudes of

(i) the initial horizontal component of the velocity.

(ii) the initial vertical component of the velocity.

Discussion Questions

These questions are usually more challenging than the self-review questions and require more thinking and they are worth further discussions during tutorial sessions. You should attempt these questions with proper working on foolscap, to your best ability, and prepare to share your work as well as to learn from others their different approaches. Very often, in learning physics, the process of getting the answer is more important than the answer itself.

Physical Quantities and Units

D1. [SAJC 2007 Prelim] Which of the following could be the correct expression for the velocity *v* of ocean waves in terms of ρ the density of seawater, *g* the acceleration of free fall, *h* the depth of the ocean and λ the wavelength?

A $\sqrt{g\lambda}$ **B** $\sqrt{g/h}$ **C** $\sqrt{\rho g h}$ **D** $\sqrt{g/\rho}$

D2. [J97/II/29] The experimental measurement of the heat capacity C of a solid as a function of temperature T is to be fitted to the expression $C = \alpha T + \beta T^3$. (Recall heat capacity *C* is energy that is required to raise the temperature of the object by one unit of temperature, $C = \frac{Q}{\Delta T}$)
What are the possible units of α and β?

	α	β
A	J	J K^{-2}
B	J K^{2}	J
C	J K	J K^{3}
D	J K^{-2}	J K^{-4}
E	J	J

D3. [N99/I/1] Four physical quantities P, Q, R and S are related by the equation $P = Q - RS$.
Which statement must be correct for the equation to be homogeneous?

A P, Q, R and S all have the same units.
B P, Q, R and S are all scalar quantities.
C The product RS has the same units as P and Q.
D The product RS is numerically equal to $(Q-P)$.

D4. [09 C1 BT1 Q2]

The power *P* generated by an ideal wind turbine is given by $P = \frac{1}{2} k\, d\, (v - b)^3$

where *k* is a characteristic of the turbine,
d is density of the fluid,
v is the velocity of the fluid and
b is a characteristic of the fluid

The possible units of *k* and *b* are:

	Units of *k*	**Units of *b***
A	no unit	m^3
B	m^2	m^3
C	no unit	m s^{-1}
D	m^2	m s^{-1}

D5. [N00/II/1]

(a) The kilogram, the metre and the second are base units. Name two other base units.

(b) Explain why the unit of energy is said to be a *derived* unit.

(c) The density ρ and the pressure P of a gas are related by the expression

$$c = \sqrt{\frac{\gamma P}{\rho}}$$, where c and γ are constants.

(i) 1. Determine the base units of density ρ.

2. Show that the base units of pressure P are kg m^{-1} s^{-2}.

(ii) Given that the constant γ has no unit, determine the unit of c.

(iii) Using your answer to (ii), suggest what quantity may be represented by the symbol c.

D6. Which one of the following physical quantities, when given in SI unit, is likely to be of the same order of magnitude as the mass of a typical watermelon in SI unit?

A Power output of a domestic electric kettle

B Human reaction time

C Weight of a typical one year old baby

D Height of the overhead bridge outside HCI (College) from the road surface

D7. [N21/I/1] What is the best estimate of the weight of a smartphone?

A 1.5 cN **B** 1.5 dN **C** 150 cN **D** 150 dN

Errors and Uncertainties

D8. [N12/I/1] A student uses an analogue voltmeter to measure the potential difference across a lamp. The voltmeter is marked every 0.02 V and has a zero error of 0.08 V. The student is not aware of this zero error and writes down a reading of 2.16 V.

Is the reading accurate and is it precise?

	Accurate	Precise
A	no	no
B	no	yes
C	yes	no
D	yes	yes

D9. [N02/I/2] An object of mass 1.000 kg is placed on four different balances. For each balance, the reading is taken five times. The table shows the values obtained together with the means. Which balance has the smallest systematic error but is not very precise?

	balance reading / kg					
	1	2	3	4	5	mean / kg
A	1.000	1.000	1.002	1.001	1.002	1.001
B	1.011	0.999	1.001	0.989	0.995	0.999
C	1.012	1.013	1.012	1.014	1.014	1.013
D	0.993	0.987	1.002	1.000	0.983	0.993

D10. [N19/I/2] The manufacturers of a digital voltmeter give, as its specification,

'accuracy ± 1% with an additional uncertainty of ± 10 mV'

The meter reads 4.072 V.

How should this reading be recorded, together with its uncertainty?

A (4.07 ± 0.01) V **B** (4.07 ± 0.04) V

C (4.072 ± 0.052) V **D** (4.07 ± 0.05) V

D11. [HC Prelim03/P1/2] Given that the quantities L, x and y are related by the equation $Lx = y^2$, what is the percentage uncertainty in L if the percentage uncertainties in x and y are 1% and 3% respectively?

A 2 % **B** 4 % **C** 5 % **D** 7 %

D12. [N22/II/1(a) modified] A mass *m* is suspended from a vertical spring attached to a fixed support. The mass is pulled down and then released. Ten oscillations are timed using a stop-watch.

The data for the mass and the time, together with the uncertainties, are shown in the table below.

time for 10 oscillations / s	6.2 ± 0.2
***m* / g**	150 ± 1%

The period *T* of the oscillations of the mass is given by: $T = 2\pi\sqrt{\frac{m}{k}}$, where *T* is in s, *m* in kg and *k* is the spring constant in N m^{-1}.

Determine the value of *k* together with its actual uncertainty. Give your answer to an appropriate number of significant figures.

D13. [N10/I/3 modified] A wire of uniform circular cross-section has diameter *d* and length *L*. A potential difference *V* between the ends of the wire gives rise to a current I in the wire.

The resistivity ρ of the material of the wire is given by the expression

$$\rho = \frac{\pi d^2 V}{4LI}$$

In one particular experiment, the following measurements are made.

d = (1.20 ± 0.01) cm
I = (1.50 ± 0.05) A
L = (100 ± 1) cm
V = (5.0 ± 0.1) V

(a) Explain which of the four measurements gives rise to the least uncertainty in the value for the resistivity.

(b) Determine the value of the resistivity together with its actual uncertainty. Give your answer to an appropriate number of significant figures.

D14. [N12/I/2] The equation connecting object distance u, image distance v and focal length f for a lens is

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

A student measures values of u and v, with their associated uncertainties. These are

u = 50 mm ± 3 mm
v = 200 mm ± 5 mm

He calculates the value of f as 40 mm. What is the uncertainty in this value?

A ± 2.1 mm **B** ± 3.4 mm **C** ± 4.5 mm **D** ± 6.8 mm

Scalars and Vectors

D15. [N10/I/2] A boat changes its velocity from 8 m s^{-1} due north to 6 m s^{-1} due east. What is its change in velocity?

A 2 m s^{-1} at a direction of 37° east of north
B 2 m s^{-1} at a direction of 53° east of north
C 10 m s^{-1} at a direction of 37° east of south
D 10 m s^{-1} at a direction of 53° west of south

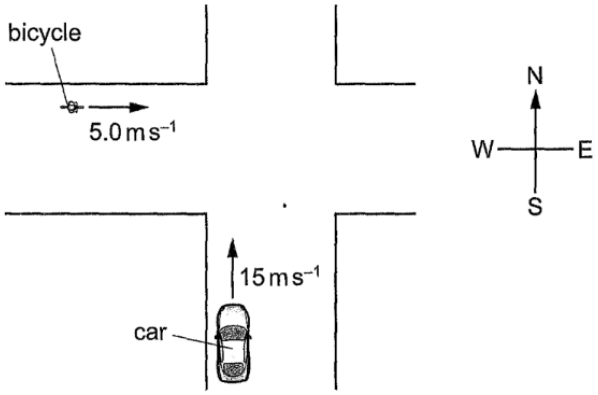
D16. [J89/II/8c] A car changes its velocity from 30 m s^{-1} due East to 25 m s^{-1} due South.
(i) Draw a vector diagram to show the initial and final velocities and the change in velocity.
(ii) Calculate the change in speed.
(iii) Calculate the change in velocity.

D17. [N16/I/3] An aircraft flies with an airspeed of 700 km/h through a 250 km/h jet-stream wind from the west. The pilot wishes to fly directly from Australia towards Changi airport in Singapore. To achieve this, the pilot points the aircraft away from the north direction.

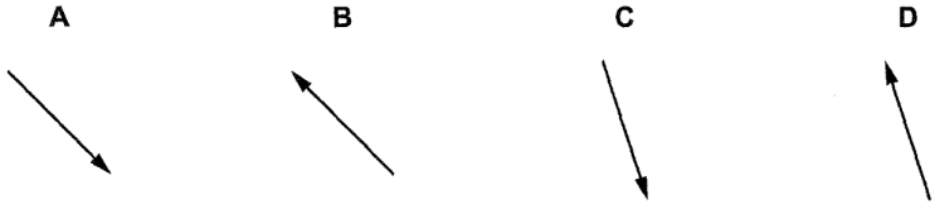
What is the speed of the aircraft in the direction of north relative to the ground?

A 450 km/h **B** 650 km/h **C** 740 km/h **D** 950 km/h

D18. [N20/I/5] A car and a bicycle are equal distances from a crossroads. The car is travelling north with a speed of 15 m s^{-1}. The bicycle is travelling east with a speed of 5.0 m s^{-1}.



At this instant, which arrow represents the velocity of the bicycle relative to the car?



Numerical answers

S1. 1.00 × 10^{3} kg m^{-3}
S3. 25.0 m s^{-1}
S4. 4.2 light years
S7. 11 ± 1 m s^{-1}
S9. (c)(i) 7.5 m s^{-1}; (c)(ii) 13 m s^{-1}
D12. 15 ± 1 N m^{-1}
D13. (b) (38 ± 3) × 10^{-3} Ω cm
D16. (ii). – 5 m s^{-1}; (iii) 39 m s^{-1} 140° clockwise with respect to the initial velocity

Chapter 2

Kinematics

Mike Powell flying through the air during the long jump finals at the 1992 Olympic Games in Barcelona, Spain. He holds the long jump world record of 8.95 m at the 1991 World Championships in Athletics (Tokyo). (Credits to pinterest.com)

"The most important factor for the distance travelled by an object is its velocity at takeoff - both the speed and angle. Elite jumpers usually leave the ground at an angle of twenty degrees or less; therefore, it is more beneficial for a jumper to focus on the speed component of the jump. The greater the speed at takeoff, the longer the trajectory of the centre of mass will be. The importance of a takeoff speed is a factor in the success of sprinters in this event."

(http://en.wikipedia.org/wiki/Long_jump)

TOPIC 2: KINEMATICS

H2 Physics Syllabus 9749

Kinematics	**Learning Outcomes** Students should be able to:
Rectilinear motion	(a) show an understanding of and use the terms distance, displacement, speed, velocity and acceleration. (b) use graphical methods to represent distance, displacement, speed, velocity and acceleration. (c) identify and use the physical quantities from the gradients of displacement-time graphs and areas under and gradients of velocity-time graphs, including cases of non-uniform acceleration. (d) derive, from the definitions of velocity and acceleration, equations which represent uniformly accelerated motion in a straight line. (e) solve problems using equations which represent uniformly accelerated motion in a straight line, including the motion of bodies falling in a uniform gravitational field without air resistance. (f) describe qualitatively the motion of bodies falling in a uniform gravitational field with air resistance.
Projectile motion	(g) describe and explain motion due to a uniform velocity in one direction and a uniform acceleration in a perpendicular direction.

Playlist of lecture examples:
https://youtube.com/playlist?list=PL_b5cjrUKDlZwgJjDB24gnT-yvQpEuV3c

Table of Contents

2.0 INTRODUCTION 5

2.1 TERMINOLOGY 5

2.1.1 Distance vs. Displacement 5

2.1.2 Speed vs. Velocity 5

2.1.3 Acceleration 7

2.2 DESCRIBING RECTILINEAR MOTION USING DIAGRAMS 9

2.2.1 Ticker tape diagrams 9

2.2.2 Vector diagrams 10

2.2.3 Recording motion using stroboscope 10

2.3 DESCRIBING RECTILINEAR MOTION USING GRAPHS 11

2.3.1 Displacement vs. Time Graph (*s-t* graph) 11

2.3.3 Summary of relations among *s-t*, *v-t* and *a-t* graphs 12

2.4 DESCRIBING RECTILINEAR MOTION USING EQUATIONS (FOR UNIFORM ACCELERATION) 14

2.4.1 Derivation of the Equations of Motion (for uniform acceleration) 14

2.5 FREE-FALLING BODIES 16

2.6 VISCOUS DRAG FORCE 17

2.7 NON FREE-FALLING BODIES 18

2.7.1 Velocity-time graph of object falling through air 18

2.7.2 Terminal Velocity 20

2.8.1 Projectile Motion 22

2.8.2 Projectile Motion: General Launch Angle 24

2.8.3 2D Motion with Air Resistance 30

FURTHER READINGS AND REFERENCES 32

Appendix I : Understanding the Calculus Notations 32

TUTORIAL 2A KINEMATICS - 1D MOTION 35

Self-Review Questions 35

Discussion Questions 38

TUTORIAL 2B KINEMATICS - 2D MOTION 45

Self-Review Questions 45

Discussion Questions 46

2.0 Introduction

Kinematics (from Greek κινεῖν, *kinein*, to move) is the branch of mechanics that describes the motion of objects without consideration of the causes leading to the motion. The study of why objects move falls under the branch of *Dynamics*. The first part of this chapter deals with one-dimensional motion of moving bodies in rectilinear motion.

2.1 Terminology

2.1.1 Distance vs. Displacement

Distance, *x*

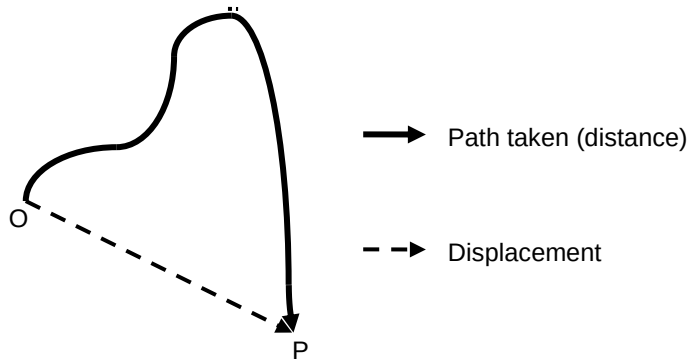
Distance is the total length travelled by a moving object irrespective of the direction of motion.

Displacement, *s*

The **displacement** of an object from a reference point O is the **linear distance** and **direction** of the object from O. The **linear distance** refers to the shortest distance from the **initial** to the **final position** of an object moved from one place to another.

Note that distance is a **scalar** quantity whereas displacement is a **vector**[1] quantity.

For illustration, in the diagram below, an object travels from point O to point P along the path represented by the solid line. The distance travelled is represented by the length of the solid line. The displacement is represented by the **length and direction** of the dashed line.



2.1.2 Speed vs. Velocity

Speed, *v*

The speed of an object is defined as the **rate of change of distance** travelled by an object with respect to **time**.

Velocity, *v*

The velocity of an object is its **rate of change of displacement** with respect to **time.**

[1] Generally, scalar quantities in textbooks are italicised and non-bold, e.g. *x*, while vector quantities in textbooks are either represented by an arrow above the quantity, a twiddle (~) below the quantity or italicised and bold, e.g. $\vec{A}$, $\underset{\sim}{a}$, or simply ***A***.

Note that speed is a **scalar** quantity whereas velocity is a **vector** quantity.

In past examinations, students were given zero marks for defining speed as distance travelled in one second. Quantities should always be defined in terms of other quantities and units should always be defined in terms of other units.

Mathematically, average and instantaneous speed[2] and velocity can be calculated as follow:

$$\text{Average speed, } \langle v \rangle = \frac{\text{Total distance travelled}}{\text{Total time taken}} = \frac{\Delta x}{\Delta t}$$

$$\text{Instantaneous speed, } v = \frac{dx}{dt}$$

$$\text{Average velocity, } \langle \boldsymbol{v} \rangle = \frac{\text{Total change in displacement}}{\text{Total time taken}} = \frac{\Delta \boldsymbol{s}}{\Delta t}$$

$$\text{Instantaneous velocity, } \boldsymbol{v} = \frac{d\boldsymbol{s}}{dt}$$

Example 1

A car travels 3.0 km due east from a point O to a point A for 2.0 minutes where it stops at a traffic light at A for 4.5 minutes. It then continues 4.0 km due north of A to point B for another 3.5 minutes. Find the

(a) average speed [Ans: 12 m s^{-1}]

(b) average velocity of the car. [Ans: 8.3 m s^{-1} , 53°]

[2] It is rare that the speed of an object is constant. When we refer to the speed of the object, we are usually referring to the **instantaneous speed**. Similarly when we refer to *velocity* and *acceleration*, it is the *instantaneous velocity* and *instantaneous acceleration* that is referred to, unless stated otherwise.

2.1.3 Acceleration

Acceleration, *a*

The **acceleration** of an object is its **rate of change of velocity** with respect to time.

Mathematically, average and instantaneous acceleration can be calculated as follow:

$$\text{Average acceleration, } \langle \boldsymbol{a} \rangle = \frac{\text{Total change in velocity}}{\text{Total time taken}} = \frac{\Delta \boldsymbol{v}}{\Delta t} = \frac{\boldsymbol{v}_{final} - \boldsymbol{v}_{initial}}{\Delta t}$$

$$\text{Instantaneous acceleration, } \boldsymbol{a} = \frac{d\boldsymbol{v}}{dt}$$

- S. I. units of acceleration: m s^{-2}
- Acceleration is a vector quantity. The direction of acceleration is in the same direction as the vector representing the **change** in velocity[3].

TIPS **Sign Convention**

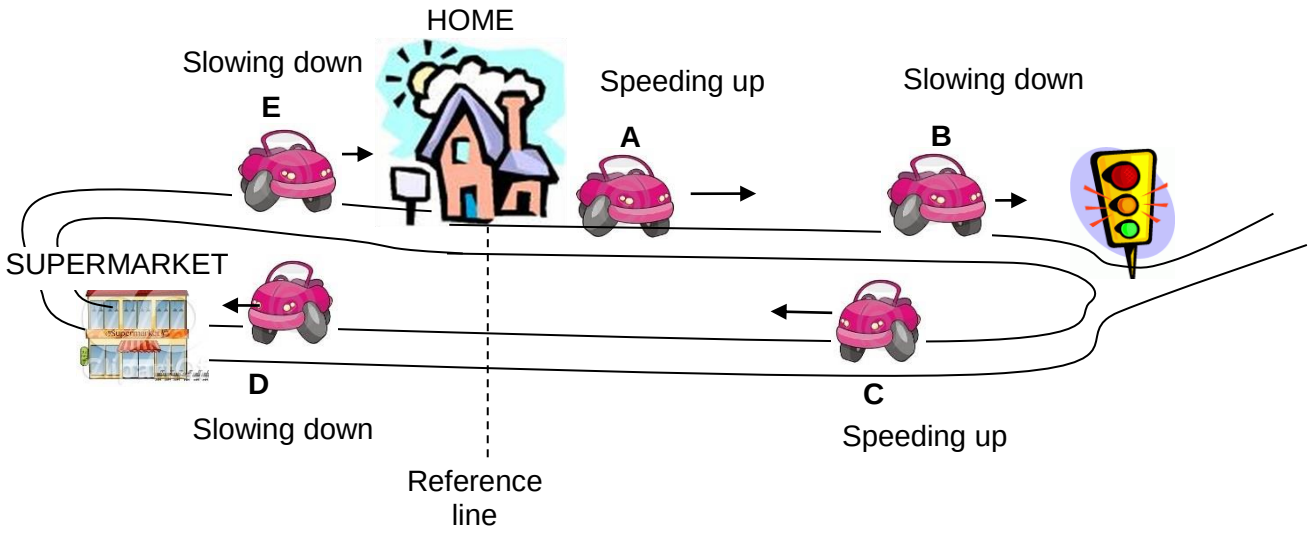
1. For an object moving in one-dimension (rectilinear), there are only two possible directions in which the object can move. One direction is opposite the other. Hence, for convenience, we define one direction as positive and the other negative.
2. For vertical motion, we usually define upward direction as positive*: +ve ↑
3. For horizontal motion, we usually take direction to the right as positive*: → +ve

*But this depends on the problem given, if it is more convenient to define downwards as positive or leftwards as positive, you may do so. In any case, **you should always state clearly the convention you have adopted.**

[3] You will learn in Dynamics that the direction of any object's acceleration is determined by the direction of the resultant force acting on the object (Newton's 2nd Law of Motion).

Example 2

John drove a car from his home towards the traffic junction. As the traffic light turned red, he stopped and waited for his turn. He then made a U-turn and drove to a supermarket to buy some groceries. John drove back to his home after making the stop at the supermarket. The journey is illustrated in the figure below, where points A, B, C, D and E are located along his journey.



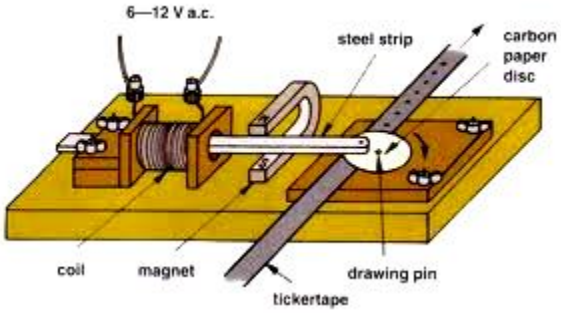
The reference line passes through John's home, and displacement **to the right** of it is defined as positive. Indicate the sign for John's displacement, velocity and acceleration at A, B, C, D and E.

	Displacement	**Velocity**	**Acceleration**
A (speeding up)			
B (slowing down)			
C (speeding up)			
D (slowing down)			
E (slowing down)			

2.2 Describing Rectilinear Motion using Diagrams

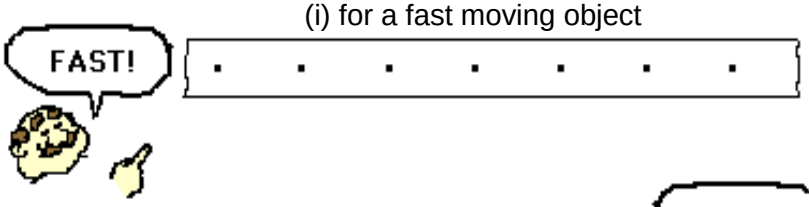
2.2.1 Ticker tape diagrams

A common way of analyzing the motion of objects in physics labs is to perform a ticker tape analysis. A long tape is attached to a moving object and threaded through a device that places a tick upon the tape at regular intervals of time – say every 0.1 second. As the object moves, it drags the tape through the "ticker," thus leaving a trail of dots. The trail of dots provides a history of the object's motion and is therefore a representation of the object's motion.

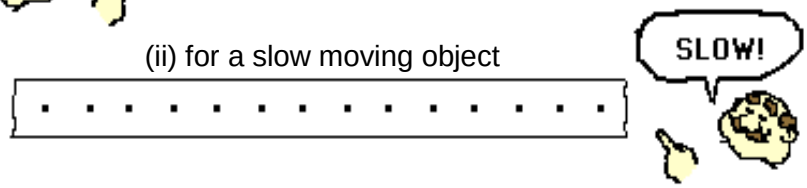


The **distance between dots** on a ticker tape represents the object's position change during that time interval.

(i) for a fast moving object

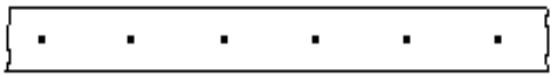


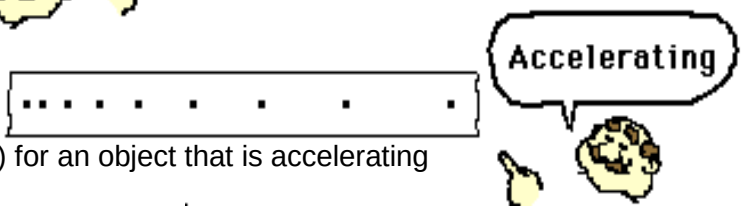
(ii) for a slow moving object



The analysis of a ticker tape diagram will also reveal if the object is moving with a constant velocity or undergoing acceleration.

(iii) for an object moving at constant speed



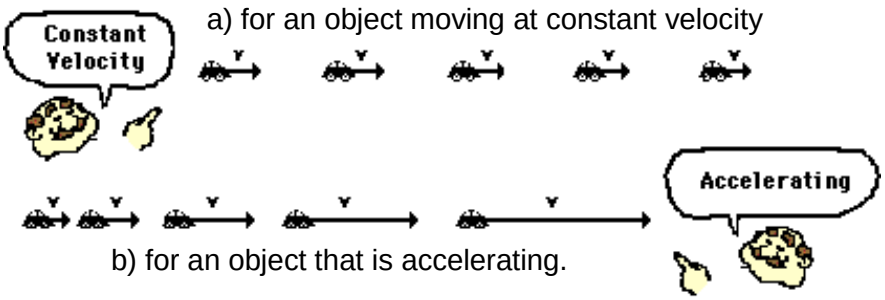


(iv) for an object that is accelerating

Usually, ticker tape diagram reflects the time axis in the usual sense we understand it, meaning, the dots on the right were made at a later time. For example, in diagram (iv), the separation between dots increases to the right, indicating acceleration.

2.2.2 Vector diagrams

Vector diagrams use vector arrows to depict the direction and relative magnitude of a vector quantity. Vector diagrams can be used to describe the velocity of a moving object during its motion. For example, the velocity of a car moving down the road could be represented by a vector diagram. In a vector diagram, the **magnitude** of the vector is represented by the length of the **vector** arrow.



a) for an object moving at constant velocity

b) for an object that is accelerating.

2.2.3 Recording motion using stroboscope

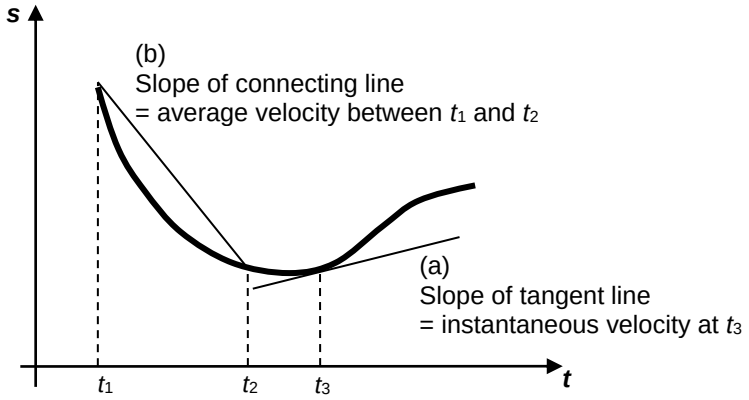
Multi-flash photo of a bouncing soccer ball

Multi-flash photographs or stroboscopic photographs are photographs of a subject illuminated with a flashing light, taken by a camera with an open shutter. By keeping the subject in the dark and then illuminating the subject only at certain times, the resultant image will show a series of different positions of the subject.

Today, by using digital video camera and digital image compositing, a multi-flash photograph can be easily re-constructed from several video frames. The video camera enables us to record motion conveniently, even at high speed. A single video camera captures 2-dimensional motion on a plane parallel to the image plane of the camera. To record complex 3-dimensional motion, many video cameras are used. This technique is useful for sports biomechanical analysis and character analysis in animated film production.

2.3 Describing Rectilinear Motion using Graphs

2.3.1 Displacement vs. Time Graph (*s*-*t* graph)

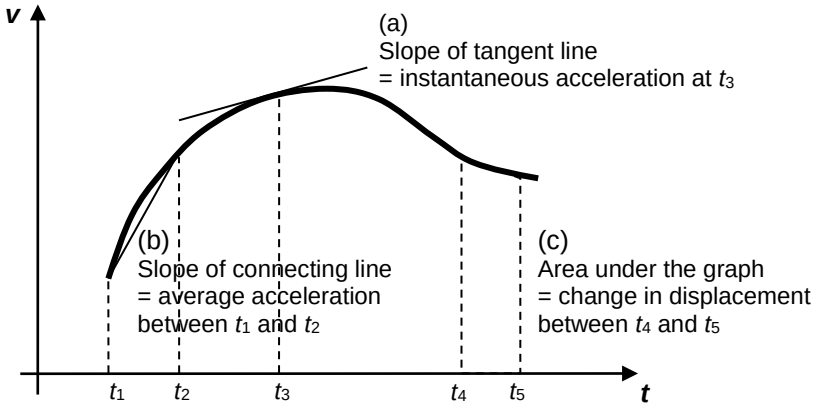


On an ***s*-*t*** graph,

The **instantaneous velocity** is the **slope** (or **gradient**) of the **tangent line** at a given **instant** of time.

- Average velocity is the **slope** of the **straight line connecting two points** corresponding to a given time **interval**.
- If an object moves with **constant** velocity, its ***s*-*t*** graph will be a **straight** line graph, and the instantaneous velocity at any instant is equal to the average velocity over the entire time interval.

2.3.2 Velocity vs. Time Graph (*v*-*t* graph)



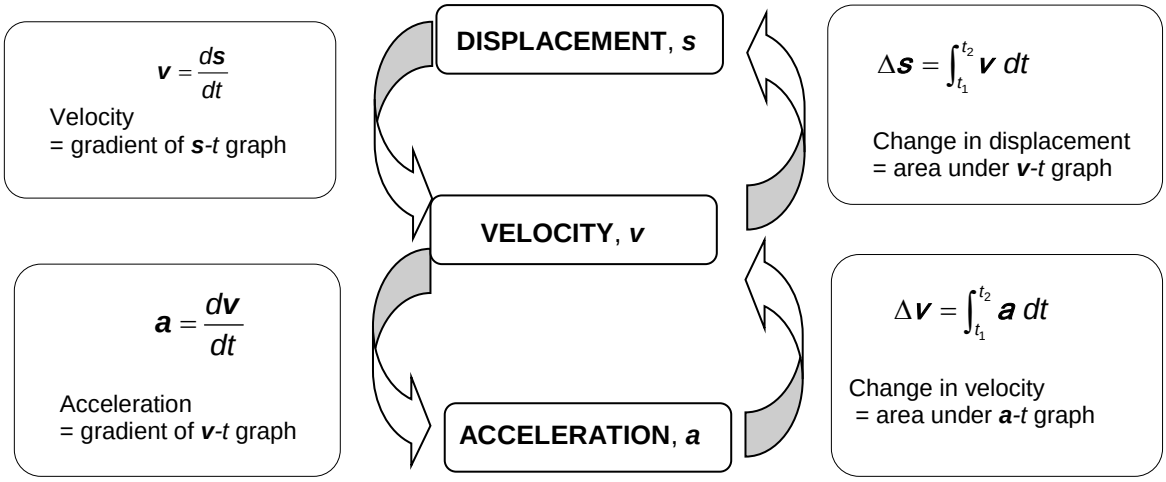
On a ***v*-*t*** graph,

The **instantaneous acceleration** is the **slope** (or **gradient)** of the **tangent line** at a given instant of time.

- Average acceleration is the **slope** of the **straight line connecting two points** corresponding to a given time **interval**.
- If an object moves with **constant** acceleration, its ***v*-*t*** graph will be a **straight** line graph, and the instantaneous acceleration at any instant is equal to the average acceleration over the entire time interval.

The **area under a *v*-*t* graph** for a time interval of Δt gives the **change in displacement** of the object in that time interval.

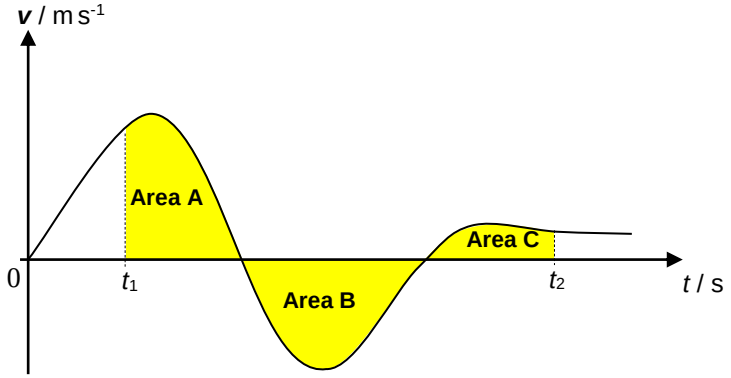
2.3.3 Summary of relations among *s*-*t*, *v*-*t* and *a*-*t* graphs



Example 3

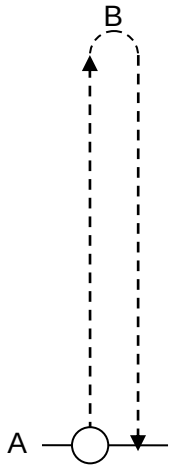
What is the expression of the change in displacement, Δs, of the object from t_1 to t_2?

If you need to find distance travelled by the object in the same time interval, what would be your answer in terms of A, B and C?



Example 4

A student throws a ball vertically upwards at a speed of 20 m s^{-1} from level A. Neglect air resistance.



(a) Sketch the ***s*** - *t*, **v** - *t* and ***a*** - *t* graphs from the moment the ball is thrown to the moment it returns to the student's hand.

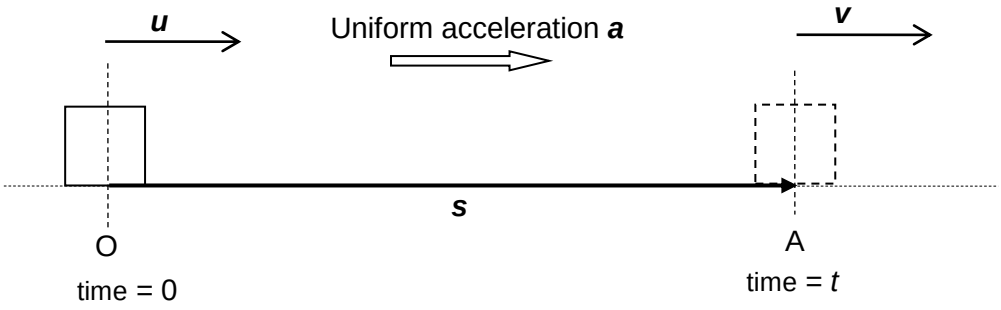
(b) Calculate the time it takes for the ball to return to the student's hand. [Ans: 4.1 s]

2.4 Describing Rectilinear Motion using Equations (for uniform acceleration)

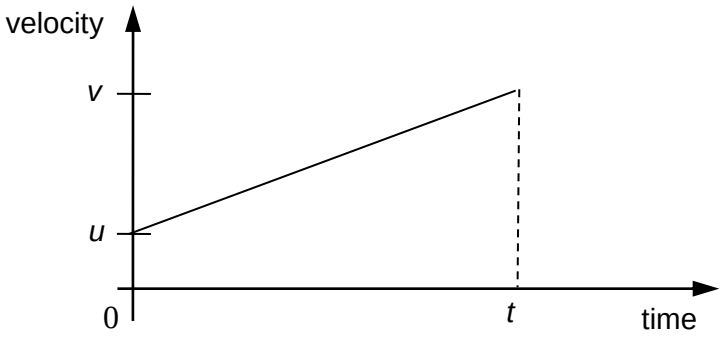
2.4.1 Derivation of the Equations of Motion (for uniform acceleration)

In this section, we will analyse the motion for an object moving in a straight line with **uniform acceleration** and derive equations that could be used for similar motions with uniform acceleration.

Suppose at time $= 0$, the object is at point O with an initial velocity of ***u***. The object moves at uniform acceleration of ***a*** until time $= t$, and has travelled through a displacement of ***s*** to arrive at point A with the final velocity of ***v***.



Let us look at its velocity-time graph:



Note: acceleration of object, *a* is given by the gradient of its ***v*** - *t* graph,

$$a = \frac{v - u}{t}$$

$$\boxed{v = u + a\,t} \qquad (1)$$

Note: The displacement of object from point O, *s*, is given by the area under its ***v*** - *t* graph from time = 0 to time = *t*,

$$s = \frac{1}{2}(u + v)\,t \qquad (2)$$

Substitute Eqn (1) into (2), we have

$$s = \frac{1}{2}[u + (u + a\,t)]\,t$$

$$\Rightarrow \quad \boxed{s = u\,t + \frac{1}{2}a\,t^2} \qquad (3)$$

From Eqn (1), we have $t = \dfrac{v - u}{a}$ and substitute into Eqn (2), we have

$$s = \frac{1}{2}(u + v)\left(\frac{v - u}{a}\right)$$

$$\Rightarrow \quad \boxed{v^2 = u^2 + 2\,a\,s} \qquad (4)$$

Note:

- The equation $s = u\,t + \frac{1}{2}\,a\,t^2$ sets the reference point $\boldsymbol{s}$ = 0 at t = 0.

- The "boxed" equations can only be used if **the acceleration is constant**.

- The equations are **vector equations.** Therefore, variables have to be analysed along the same dimension. For example, if we choose to analyse the motion along the *x*-direction, then we can only use quantities like $\boldsymbol{u_x}$, $\boldsymbol{v_x}$ and $\boldsymbol{a_x}$. Acceleration along the *y*-direction $\boldsymbol{a_y}$ will not affect motion along the *x*-direction.

TIPS Solving Rectilinear Motion Problems:

In solving problems involving rectilinear (one-dimensional) motion, follow the following steps:

1. Identify key words/data from the question such as:
   - 'start' means when time t = 0
   - 'from rest'/'dropped' means initial speed, u = 0
   - 'stop' means final speed, v = 0
   - 'maximum height' means v at maximum height = 0
2. Assign a positive direction and ensure the correct signs for vectors.
3. Select and apply the relevant kinematics equations of motion.

Example 5

A motorist traveling at 13 m s^{-1} approaches traffic lights, which turn red when he is 25 m away from the stop line. His reaction time is 0.70 s. If he brakes fully such that the car slows down at a rate of 4.5 m s^{-2},

On which side of the stop line will he stop, and how far from the stop line would he stop?
[Ans: Beyond the stop line, 2.9m]

2.5 Free-Falling Bodies

The simplest example of motion with constant acceleration is a free falling body.

A free falling object is any object moving only under the influence of gravity, *i.e.* the effects of resistive forces (i.e. drag forces) and upthrust are negligible. For free falling objects, whether they are initially thrown upward, downward or released from rest, they will accelerate downwards at a rate of **9.81 m s^{-2}**, regardless of the objects' weight, size and shape. Let us illustrate this below.

Consider a mass *m* under free falling. Since air resistance is negligible, the only force acting on the mass is its own weight. Applying Newton's 2nd Law,

$$F_{net} = ma$$
$$mg = ma$$
$$a = g$$

In essence, even though a larger mass experiences a larger gravitational force, its larger inertia also makes it harder to accelerate. These two effects cancel out cleanly. This curious coincidence causes a feather to fall at the same rate as a lead ball.

For a video explanation:

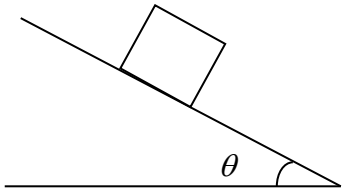
Acceleration of free fall by xmphysics
https://www.youtube.com/watch?v=HTt7C0zgn2c

Example 6

The following are 2 cases of objects in motion.

Case 1: A box of mass *M*, sliding down a smooth incline.

Assume that air resistance is negligible.



Case 2: A feather of mass *m*, falling in vacuum.

(a) Determine the acceleration of the box and the feather.

(b) Explain whether the cases are in free fall.

(c) Explain if the equations of motion can be applied to each case.

2.6 Viscous Drag Force

In fluid dynamics, drag (or fluid resistance) refers to forces that oppose the relative motion of an object through a fluid (a liquid or gas). When a body moves through a fluid, it has to push the surrounding fluid molecules away in order to move. Hence it exerts a force on the fluid molecules and by Newton's Third Law, the molecules exert an equal and opposite force back on the body. Hence there is resistance to motion.

It follows that drag force depends on the **velocity** of the body, *i.e*. the faster the motion of the body through the fluid, the larger the resistance. It also means that if the body is at rest relative to the fluid, there is no resistance.

Other factors affecting the magnitude of drag force include the **shape** and **dimension** of the body (a more streamline shape allows a body to experience less drag as it pushes through the fluid) and the type of fluid (**viscosity** of the fluid).

When there is no turbulent flow (*e.g.* small and streamline objects traveling at low speeds) the drag force is proportional to the speed of the object. The magnitude of drag force can be stated empirically as follows:

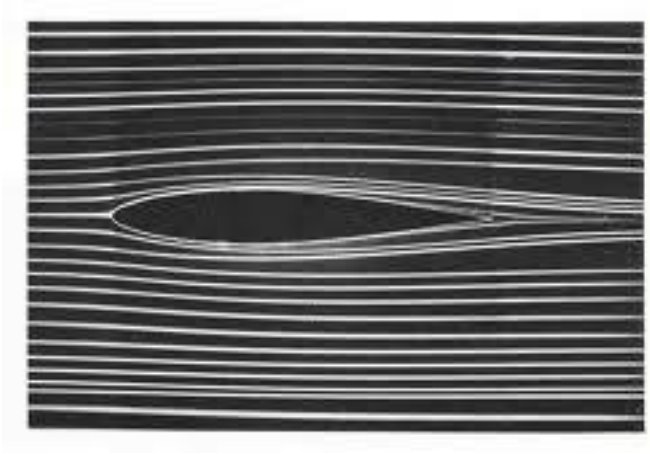
$F_D = kv$

F_D : magnitude of drag force
k : a constant that depends on the dimensions of the body and the type of fluid
v : the velocity of the body relative to the fluid

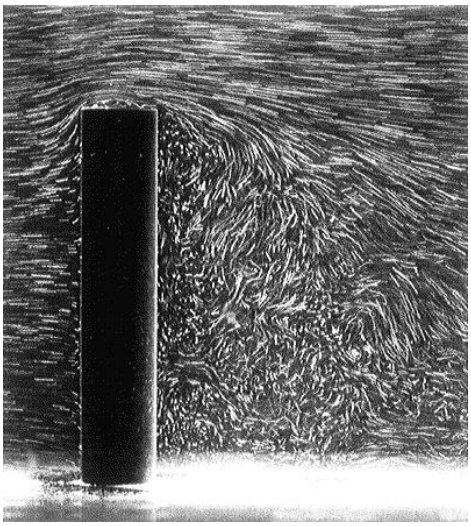
If the fluid flow is turbulent (*e.g.* objects moving at high speeds through air, such as aircraft, skydiver and baseball), the drag force is approximately proportional to the square of the speed. The drag force is then given by:

$F_D = kv^2$

F_D : magnitude of drag force
k : a constant that depends on the dimensions of the body and the type of fluid
v : the velocity of the body relative to the fluid



Laminar flow over a streamline body



Turbulent flow results in higher drag

2.7 Non Free-Falling Bodies

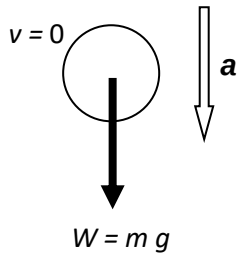
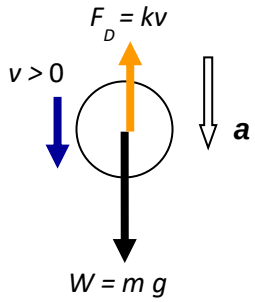
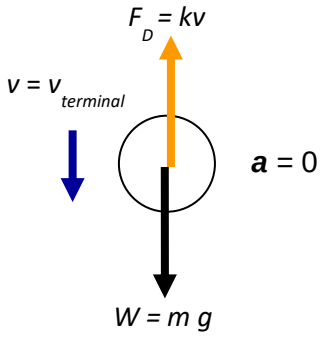
In practice, free fall does not happen because of resistive forces. The effects on a diver or bungee jumper may not be significant, but this is clearly not the case for a parachutist where air resistance is (deliberately) used to limit the maximum downward velocity of the parachutist.

2.7.1 Velocity-time graph of object falling through air

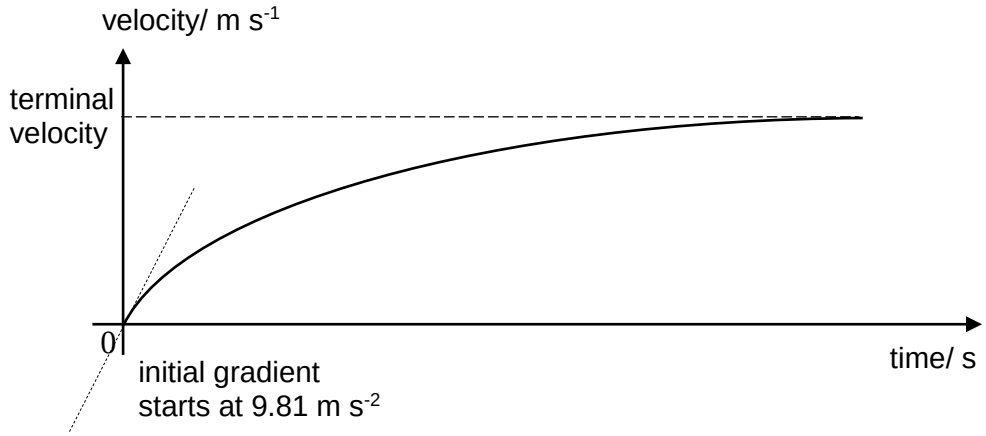
Let us now consider the motion of an object, dropped from rest, falling through air. If air resistance is not negligible, the forces acting on the object are:

W : Weight of object
F_D : Drag force

For ease of analysis, break the motion up into stages as seen below. Let us assume that drag force is proportional to speed.

Time	Forces acting on the body	Motion of the object
$t = 0$		• Object is initially at rest. • The only force acting on the object is the gravitational force (weight, $W = m\,g$). *By Newton's 2nd Law,* $(\downarrow) \quad \Sigma F = ma \Rightarrow \quad mg = ma \Rightarrow \quad a = g$ • The body begins to fall with an acceleration of ***a* = *g***.
$t = t_1$		• As the body gains velocity, air resistance which is dependent on the velocity also increases. *By Newton's 2nd Law* $(\downarrow) \quad \Sigma F = ma \Rightarrow \quad mg - kv = ma \Rightarrow a = g - \frac{kv}{m} < g$ • As F_D increases, the net force decreases and hence the acceleration decreases and falls below ***g***. • The object is nevertheless still accelerating and hence the speed still increases (just at a decreasing rate), so the air resistance continues to increase.
$t = t_2$		• Eventually the air resistance increases to a point where it balances the weight of the body. i.e. $W = F_D$ *By Newton's 2nd Law* $(\downarrow) \quad \Sigma F = ma \Rightarrow \quad mg - kv = ma = 0 \Rightarrow \quad a = 0$ • Resultant force acting on the body is zero. Hence acceleration of the body is **zero**. • The object then travels at a constant velocity called the **terminal velocity**.

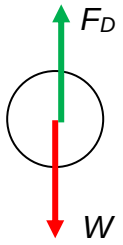
The following graph shows the non-linear velocity change with time for a falling object experiencing air resistance.



2.7.2 Terminal Velocity

We can now explain why heavier objects fall at a faster rate than light ones in the presence of air resistance. For simplicity, let us assume air resistance is proportional to speed *v, i.e.* $F_D = kv$.

Consider the free body diagram of an object falling at terminal velocity,



At terminal velocity, the acceleration is zero and so the net force is zero. Hence,

$$F_D = W$$

Assuming air resistance is proportional to speed, therefore $F_D = kv_t$

$$kv_t = mg$$
$$v_t = \frac{mg}{k}$$

Based on this equation, we can now conclude:

1. Between two objects of the same dimensions but different densities, the denser one reaches a higher terminal velocity because it has a larger *m* but the same *k*.

2. Between two objects of the same mass but different shapes, the more streamlined one reaches a higher terminal velocity because it has a smaller *k* but the same *m*.

For a video explanation:

Terminal velocity by xmphysics
https://www.youtube.com/watch?v=tuS98lyEzzg

Example 7

A steel ball bearing is released from a height above a deep pool of oil and it reaches a terminal velocity in the oil. Sketch the graph to show the variation of the velocity ***v*** of the ball bearing against time *t* when the ball enters the oil the drag force acting on the ball is

(a) smaller than the weight of ball (Case A)

(b) larger than the weight of ball (Case B)

Indicate clearly on your diagram, t_1, the time when the ball bearing hits the oil. Assume that upthrust is negligible.

2.8 Two-Dimensional Kinematics

In this part of the chapter, we deal with the kinematics of a particle moving in a plane, or two-dimensional motion. Some common examples of motion in a plane are the motion of projectiles, motion of satellites and the motion of charged particles in an electric field. For the beginning we shall deal with the special case of motion of **an object with a uniform velocity in one direction and a uniform acceleration in a perpendicular direction.**

2.8.1 Projectile Motion

Anyone who has observed a soccer ball in motion (or, for that matter, any object thrown into air) has observed *projectile motion*.

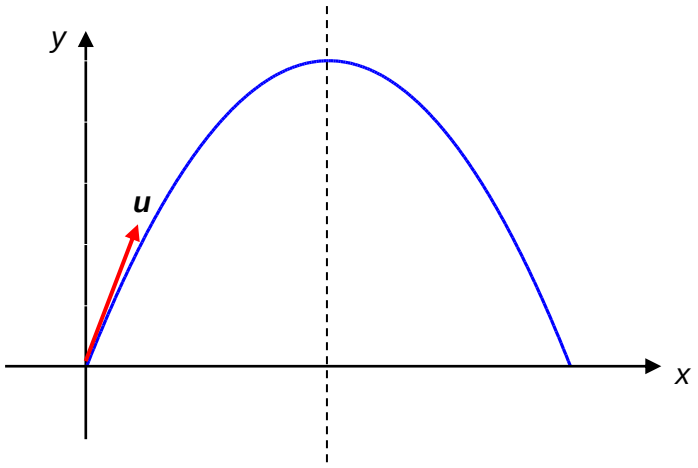
Projectile motion refers to the motion of an object in the presence of a gravitational field. In this section, we are interested in the motion of an object projected with an initial velocity at an angle to the horizontal. To analyze this type of motion, we will start with an idealized model. The assumptions made in our idealized model are:

(1) the effect of air resistance is negligible.

(2) the acceleration due to gravity ***g*** *is constant*[4], *acts downwards* and has a *magnitude equal to* $g = 9.81\ m\ s^{-2}$.

With the above assumptions, we find that the path of a projectile, which we call its ***trajectory*****, is always a parabola[5].**

Key Characteristics of a Parabolic Path:



- The path of the projectile is symmetrical about the highest point.
- The speed of the projectile is symmetrical about the highest point.
- The time it takes for the object to go from $y = 0$ to the highest point

  is equal to

  the time it takes for the object to go from the highest point to $y = 0$.

It was Galileo, who first accurately described projectile motion. He showed that it could be understood by analysing the horizontal and vertical components of the motion separately. He found that:

The horizontal and vertical motions of the projectile are independent of each other.

[4] We shall restrict ourselves to objects whose distance travelled and maximum height above the Earth are small compared to the Earth's radius (6400 km) so that ***g*** can be considered constant.

[5] The proof that trajectory of a projectile is parabolic will be shown in Example 11.

To examine this idea further, let us examine the motion of 2 balls:

(1) Ball 1 is a ball released from rest and allowed to fall vertically through a hole on the table, and

(2) Ball 2 is allowed to roll off the end of a table with an initial velocity $\boldsymbol{u_x}$ in the horizontal (***x***) direction.

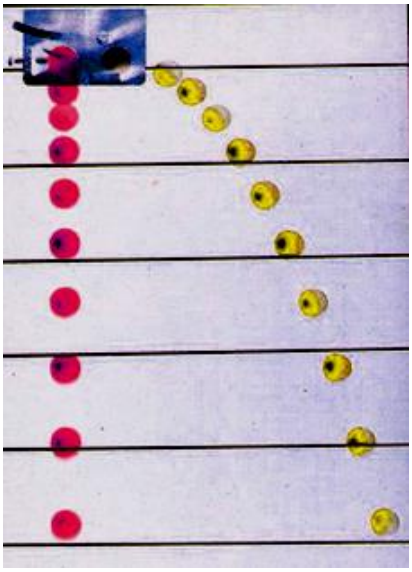
Both the balls leave the table at the same time. The stroboscopic photograph below depicts the motion of both balls. Let us examine the vertical and horizontal motion separately.

- **Vertical Motion:**

  Consider the motion of both balls in the vertical direction.

  Note,

  - The **vertical displacement of Ball 2 matches the vertical displacement of Ball 1**.
  - The **vertical displacement and vertical velocity** of Ball 2 **are only affected by the vertical acceleration *g*** (as in Ball 1).



- **Horizontal Motion:**

  Consider the motion of Ball 2. Note that

  - The **horizontal displacements** made by the ball **at equal intervals of time are approximately the same**.
  - The **horizontal velocity remains constant** with time.
  - The **horizontal acceleration is zero**.

    Consistent with our assumption that only the weight acts on Ball 2 and no other forces act in the horizontal direction. Hence, horizontal acceleration is zero.
  - The **vertical acceleration hence has no effect on the horizontal velocity and horizontal displacements.**

Hence, we see that horizontal and vertical motions are **independent** of each other and projectile motion can be simplified by analysing the vertical and horizontal motions separately. Another result of this analysis, which Galileo himself predicted, is that *an object projected horizontally will reach the ground in the same time as an object dropped vertically.* This is because the vertical motions are the same in both cases.

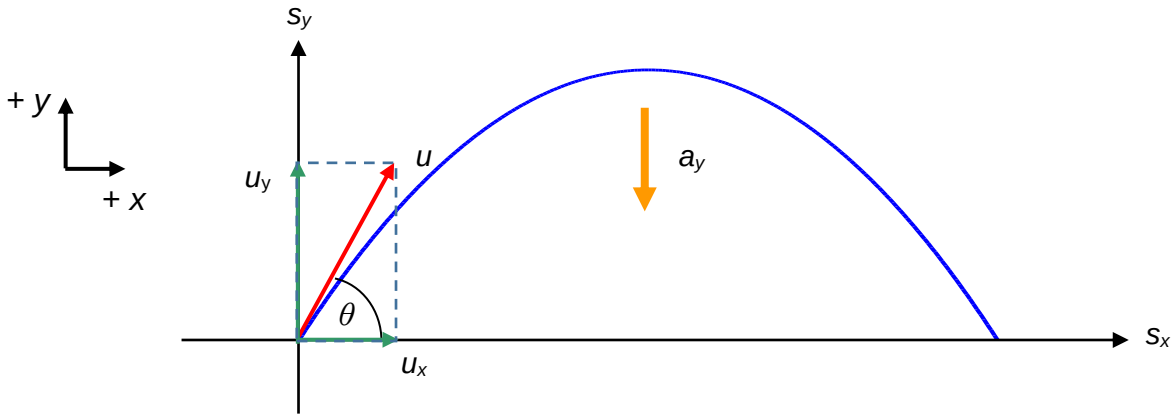
2.8.2 Projectile Motion: General Launch Angle

Let us now use our equations of motion for one-dimensional uniform acceleration motion and extend this to two-dimensional motion.

Recall our three equations of motion for one-dimensional motion:

$$v = u + a\,t$$
$$v^2 = u^2 + 2\,a\,s$$
$$s = u\,t + \tfrac{1}{2}\,a\,t^2$$

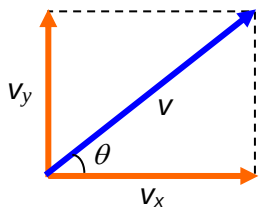
Consider an object projected above the horizontal at an angle of θ with an initial speed of u, as shown in figure below.



Analysing our projectile horizontally and vertically, and choosing our coordinate system as in the diagram, we have:

Horizontally				Vertically			
	$\boldsymbol{a_x}$	$= 0$			$\boldsymbol{a_y}$	$= -g$	
	$\boldsymbol{v_x}$	$= \boldsymbol{u_x}$			$\boldsymbol{v_y}$	$= \boldsymbol{u_y} + \boldsymbol{a_y}\,t$	
$\Rightarrow$	$\boldsymbol{v_x}$	$= \boldsymbol{u}\cos\theta$	(1)	$\Rightarrow$	$\boldsymbol{v_y}$	$= \boldsymbol{u}\sin\theta - gt$	(3)
	$\boldsymbol{s_x}$	$= \boldsymbol{u_x}\,t$			$\boldsymbol{s_y}$	$= \boldsymbol{u_y}\,t + \tfrac{1}{2}\,\boldsymbol{a_y}\,t^2$	
$\Rightarrow$	$\boldsymbol{s_x}$	$= (\boldsymbol{u}\cos\theta)\,t$	(2)	$\Rightarrow$	$\boldsymbol{s_y}$	$= (\boldsymbol{u}\sin\theta)\,t - \tfrac{1}{2}\,gt^2$	(4)
					v_y^2	$= u_y^{\,2} + 2\,\boldsymbol{a_y}\,\boldsymbol{s_y}$	
				$\Rightarrow$	v_y^2	$= u^2\sin^2\theta - 2\,g\,\boldsymbol{s_y}$	

In addition, the **resultant velocity** $\boldsymbol{v}$ at any instant of time can be found from the horizontal and vertical component of the velocities:



Magnitude of the velocity, $v = \sqrt{v_x^2 + v_y^2}$

Direction of the velocity θ can be determined from:

$$\tan\theta = \frac{v_y}{v_x}$$

IMPORTANT TO KNOW

In exam questions, it is essential to give both the magnitude and direction for any vector quantity.

From the above equations, we can easily use them to obtain other information of the projectile, like:

- **Maximum Height:** Maximum vertical displacement with respect to the initial starting point (vertical velocity, $\boldsymbol{v_y}$ = 0 m s^{-1}).
- **Horizontal Range:** Maximum horizontal displacement with respect to the initial starting point.
- **Angle of Projection for Maximum Horizontal Range.**
- **Equation of the trajectory of the projectile (see Example 13(v)).**

TIPS Solving Projectile Motion Problems:

In solving problems involving projectile motion, follow the following procedures:

1. *Read* the question carefully.
2. *Draw* a careful diagram summarising the given information.
3. Choose an *x*-*y* coordinate system.
4. Identify the accelerations in the *x* and *y* directions.
5. If you are given the initial velocity, you may want to resolve it into its *x* and *y* components.
6. Analyse the horizontal and vertical motion separately.
7. Apply the relevant equations. Remember to take sign conventions into consideration.

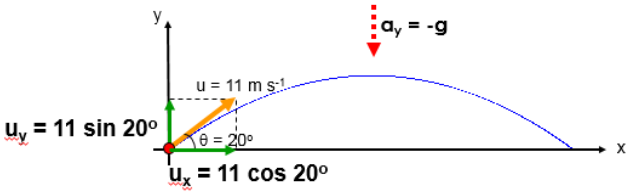
We shall now work through several examples of projectile motion to familiarize ourselves with the strategy and the application of the equations.

Example 8

A long-jumper leaves the ground at an angle of 20^o to the horizontal and at a speed of 11 m s^{-1}.

(a) How far does he jump? [Ans: 7.93 m]

(b) What is his maximum height reached? [Ans: 0.72 m]

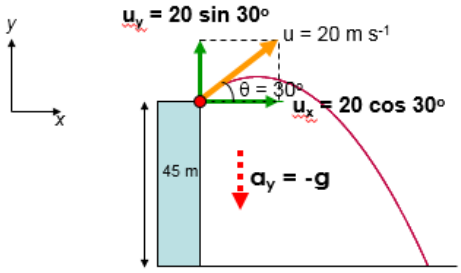


Example 9

A stone is thrown from the top of a building upward at an angle of 30° to the horizontal with an initial speed of 20 m s^{-1}. If the height of the building is 45 m,

(a) How long is the stone in flight? [Ans: 4.22 s]

(b) What is the speed of the stone just before it strikes the ground? [Ans: 35.9 m s^{-1}]

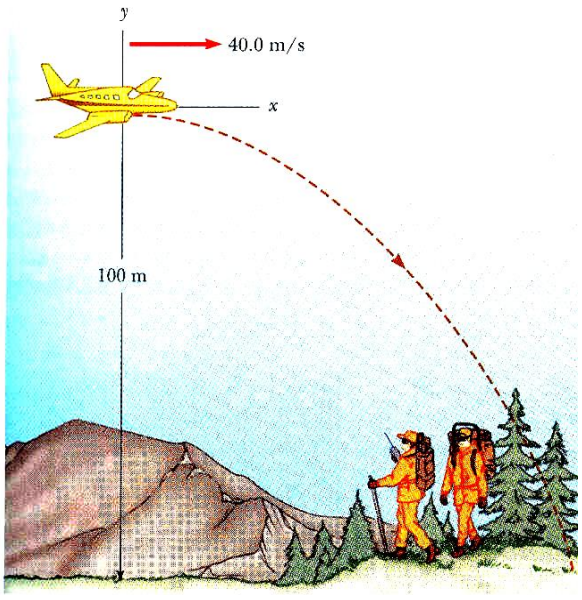


Example 10

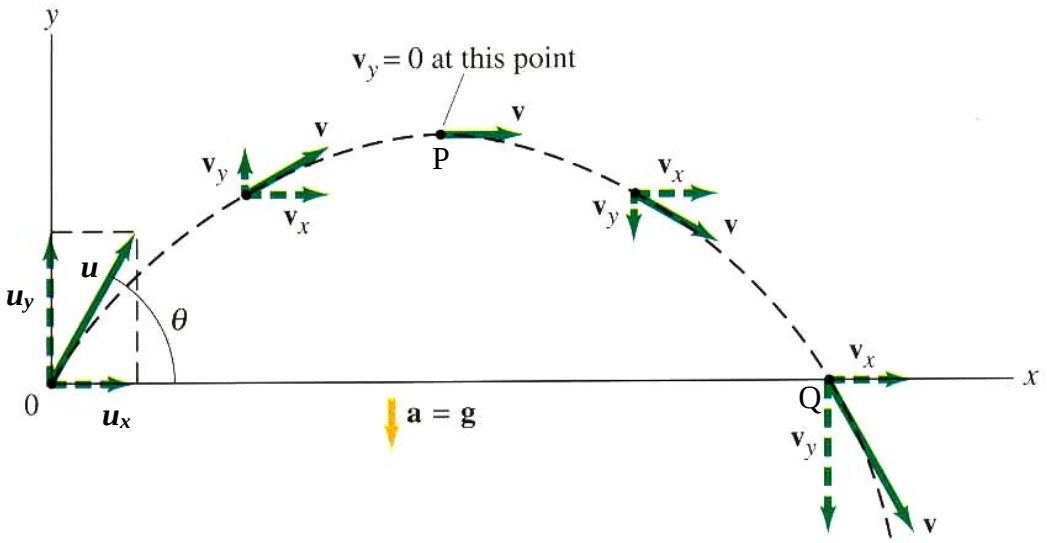
An Alaskan rescue plane drops a package of emergency rations to a stranded party of explorers, as shown in the figure below. The plane is travelling horizontally at 40.0 m s^{-1} and is 100 m above a point O on the ground.

(a) Determine how far from the point O does the package strike the ground. [Ans: 181 m]

(b) Determine the position of the package relative to the airplane when it strikes the ground.



Example 11: More about Projectile Motion



Consider a projectile projected with an initial velocity ***u*** at an angle above the horizontal, diagram above. Assuming that air resistance is negligible and the acceleration of free fall ***g*** remains constant, derive using the equations of motion expressions of:

(a) maximum height *H* reached by the projectile,

(b) duration of time of flight t_{flight},

(c) horizontal range *R,*

(d) optimal angle of projection θ_{max} (for maximum range),

(e) equation of the trajectory of the projectile.

Adopt the sign conventions as given in the diagram above.

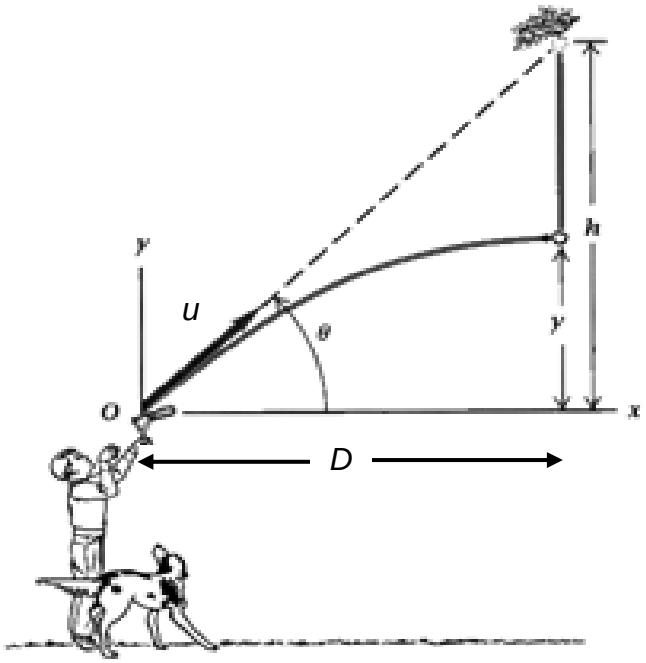
Example 12

A boy aims his slingshot directly at an apple hanging in a tree. At the moment he **shoots** a pebble, the apple **drops**.

The initial speed of the pebble is 25 m s^{-1}. The position of the apple is 5.0 m above the horizontal level, (h = 5.0 m). The pebble hits the apple 0.50 s later.

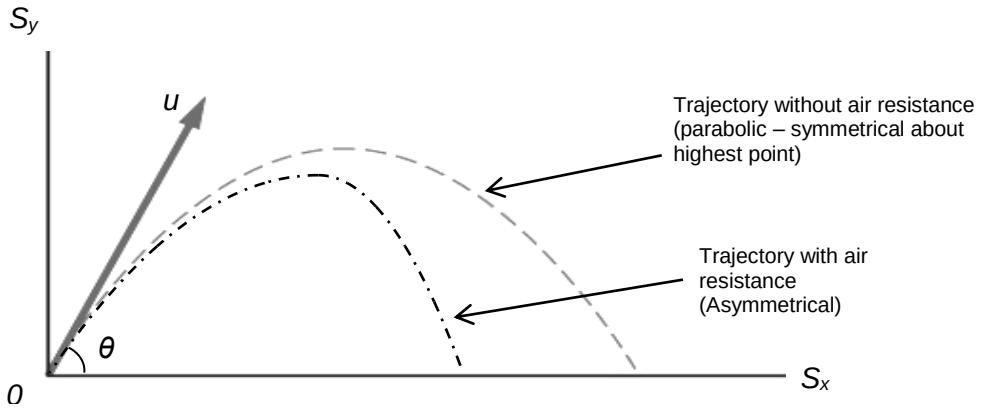
(a) Determine the angle at which the boy must aim his slingshot towards the apple. [Ans: 24°]

(b) Find the horizontal distance D. [Ans: 11.4 m]



2.8.3 2D Motion with Air Resistance

Consider an object projected above the horizontal with an initial speed of u and at angle of θ with respect to the horizontal.



Suppose that the air resistance experienced by the projectile is not negligible. Then **throughout the motion, the projectile will experience a drag force in opposite direction to its motion** (velocity).

Let us further analyse the horizontal motion and vertical motion independently,

Horizontally,

draw force(s) acting on the projectile ◯	The drag force acts in (the same/opposite) direction to the horizontal velocity. ⇒ The horizontal velocity (increases/decreases/remains the same) with time. ⇒ The **horizontal range is reduced** (as compared with the case if no air resistance present).

Vertically,

On its flight up,	On its flight down,
draw force(s) acting on the projectile 	draw force(s) acting on the projectile
The drag force acts in (the same/opposite) direction as the weight *mg*. The magnitude of the net acceleration is (less/more) than *g*. ⇒ **Maximum height reached** is lower than if there were no air resistance present.	The drag force acts in (the same/opposite) direction to the weight *mg*. The magnitude of the net acceleration is (less/more) than *g*. ⇒ The final (vertical) speed v_y will be (smaller/larger) than the initial (vertical) speed u_y (at the point of projection). ⇒ The average (vertical) speed upwards is (smaller/larger) than the average (vertical) speed downwards. Since, distance travelled up = distance to come down and Average speed = Total distance / Time taken ⇒ **The time taken to travel up is shorter than the time taken to travel down.**

Furthermore, comparing the vertical motion of the projectile on its flight up and on its flight down, we see that:

- The (vertical) acceleration experienced by the projectile when it is going up is different from the (vertical) acceleration experienced by the projectile when it is coming down. ($a_{y,up} \neq a_{y,down}$)
- For a particular height, the speed of the projectile when it is going up is also different for the speed of the projectile when it is coming down.

Hence, we can see that the **path of the projectile will not be symmetrical about the highest point**, *i.e.* the path is asymmetrical about the highest point.

In summary,

The key characteristics of a projectile motion with air resistance (compared to if air resistance is negligible) are:

1. Horizontal range is shorter.
2. Maximum height is lower.
3. The time taken to reach maximum height must be shorter than the time taken to return to the starting point.
4. Path is asymmetrical about a vertical line through its highest point.

Further Readings and References

Appendix I : Understanding the Calculus Notations

Although rigorous treatment with calculus is not required in the A-level syllabus, it is however useful for us to understand the physical significance of some of the calculus notations as they are commonly used to define physical quantities.

The Derivative : $\frac{dy}{dx}$

Consider a graph of y vs. x as indicated by the curve in Figure 1. $\frac{\Delta y}{\Delta x}$ indicates the gradient of the line joining the 2 points A and B on the line.

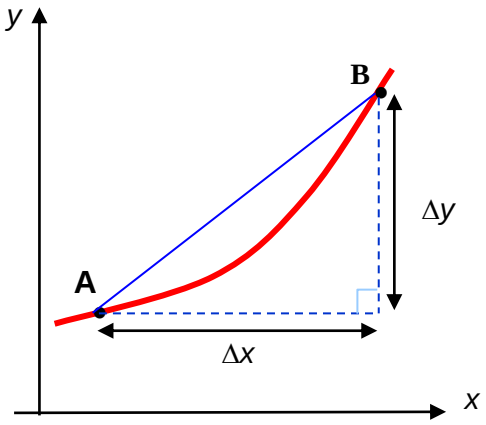


Figure 1. Graph of y vs. x

Figure 2 shows that if we move point B closer towards point A, eventually the 2 points will meet. At this point gradient of the line AB will be exactly equal to the gradient of the tangent of the curve at A. (as represented by the bold straight line). Mathematically, we say:

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \frac{dy}{dx} = \text{gradient of the tangent of the } y - x \text{ graph}$$

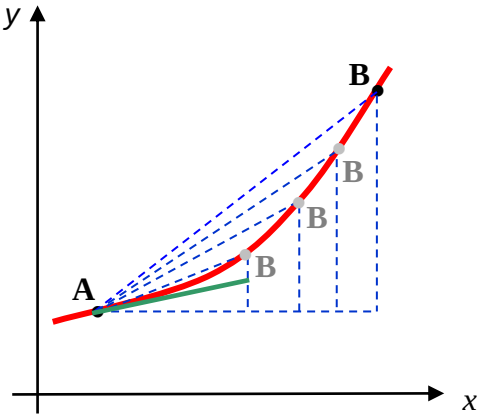


Figure 2. Tangent of the curve at **A**.

This physical interpretation is very important in physics as many physical quantities are expressed in this derivative form.

For example, in our study of kinematics, we have :

Velocity, v = Rate of change of displacement with respect to time

$$= \frac{ds}{dt}$$

= Gradient of tangent of the s - t graph

and

Acceleration, a = Rate of change of velocity with respect to time

$$= \frac{dv}{dt}$$

= Gradient of the tangent of the v - t graph

The Integral: $\int y\ dx$

The reverse process to taking the derivative is referred to either as *integration* or finding the *anti-derivative*. Graphically, it is equivalent to finding the area under the curve. Let us see why.

Consider a *y*-*x* graph as shown in Figure 3. Suppose we wish to find the area under the graph from some initial *x* value, $x_{initial}$, to some final *x* value, x_{final}. One way to do so is break the area up into *n* number of small rectangles each of some area, A_i.

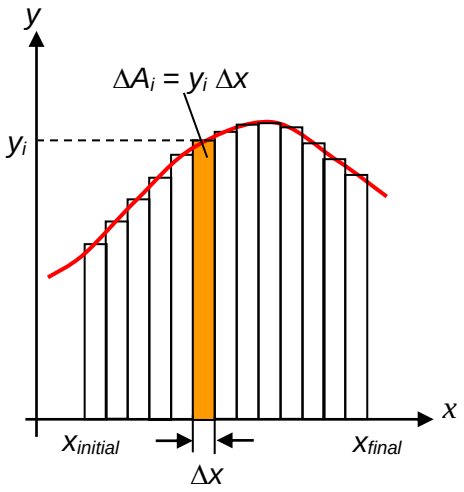


Figure 3. Finding the area under a curve. The area of the shaded rectangle is equal to the product of the width of the rectangle Δx and the *y* value of the curve at the particular *x*, y_i.

Hence, the area of each small rectangle can be written as: $A_i = y_i\,\Delta x$

and the total area under the curve can be estimated by summing the area of all these little rectangles, i.e.

$$\text{Total area, } A = \sum_{i=1}^{n} y_i \Delta x$$

where the symbol $\sum$ (upper case Greek sigma) signifies a sum over all terms, i.e. over all values of *i*.

For the approximation of the area to be even better, we can try to divide the area into smaller and smaller rectangles, by decrease the interval Δx, i.e. we say that we let $\Delta x \to 0$. Therefore,

$$A = \lim_{\Delta x \to 0} \sum_{i=1}^{n} y_i \Delta x$$

This is also written as $\int_{x_{initial}}^{x_{final}} y(x)\ dx$. So, $\int_{x_{initial}}^{x_{final}} y(x)\ dx$ is the area under the *y*-*x* graph.

We now use the defining equations for acceleration and velocity to illustrate the use of the above concept.

The defining equation for acceleration of an object is $a = \frac{dv}{dt}$

Therefore, the change in velocity of an object is

$$v_{final} - v_{initial} = \int_{t_{initial}}^{t_{final}} a(t)\, dt$$

which is the area under *a*-*t* graph.

Similarly, velocity of an object is $v = \frac{ds}{dt}$

Rearranging the equation, we also have the change in displacement of an object given as

$$s_{final} - s_{initial} = \int_{t_{initial}}^{t_{final}} v(t)\, dt$$

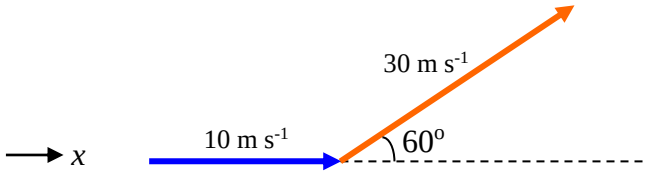
which is the area under *v*-*t* graph.

Tutorial 2A Kinematics - 1D Motion

Self-Review Questions

Use these questions to test your familiarity with the concepts. These questions should be sufficiently easy such that you can solve them on your own, with a little bit of thinking, without help from the tutors. The solutions are made available on the cohort Google Classroom for self-check.

S1 A particle moves with an initial velocity of 10 m s^{-1} in the x-direction for 5.0 s. It then changes its velocity to 30 m s^{-1} at an angle of 60° to the horizontal and moves for a further 2.0 s.



Calculate

(a) the change in speed

(b) the change in velocity of the particle

(c) the average speed

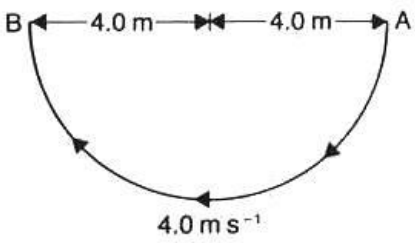
(d) the average velocity of the particle

S2 A motorist drives north for 35 minutes at 85 km h^{-1} and then stops for 15 minutes. He then continues northeast travelling 130 km in 2 h. (Hint: Draw a vector diagram to illustrate the motion)

(a) What is his total displacement?

(b) What is his average velocity?

S3 An object moves along a semi-circular path AB of radius 4.0 m as shown in the figure on the right, at a constant speed of 4.0 m s^{-1}.



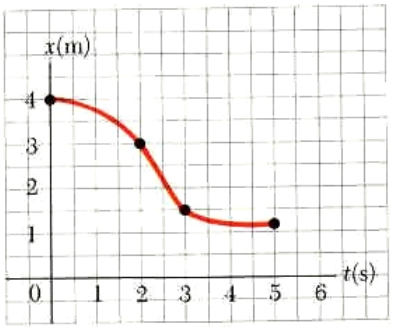
Calculate

(a) the time taken,

(b) the average velocity,

(c) the change in velocity

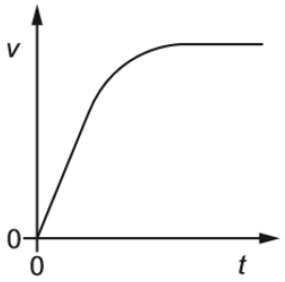
S4 Using the figure on the right, determine



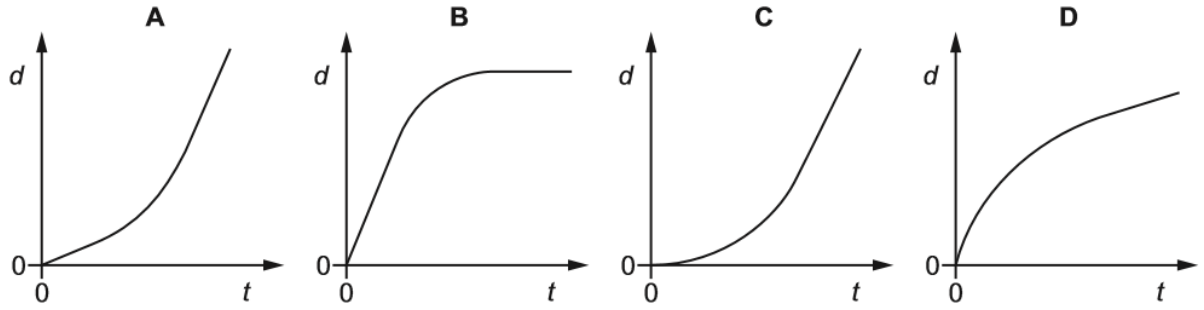
(a) the average velocity between t = 0.0 s and t = 3.0 s,

(b) the instantaneous velocity at t = 3.0 s.

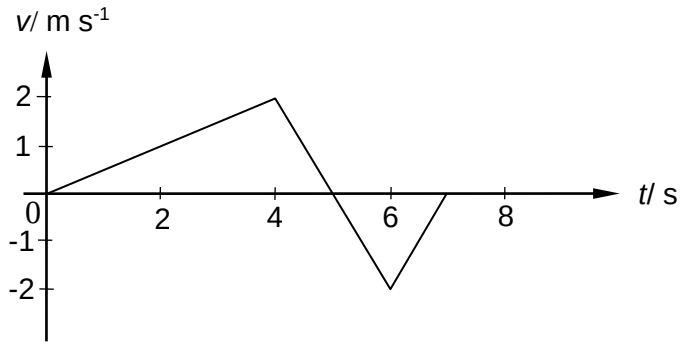
S5 A sky-diver falls vertically from a helicopter and reaches constant (terminal) velocity. The graph shows the variation with time t of the speed v of the sky-diver.



Which graph shows the variation with time t of the distance d fallen by the sky-diver?



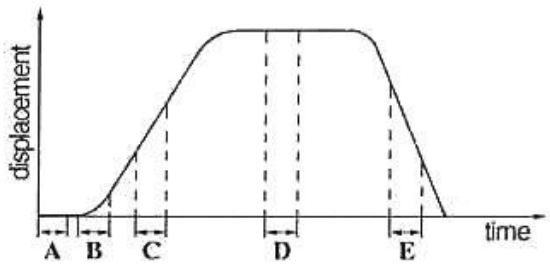
S6 The velocity-time graph of a particle along a straight line is shown.



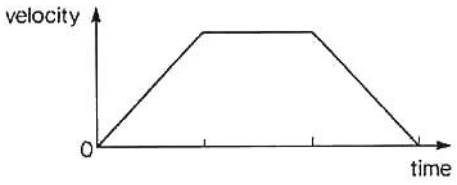
The displacement of the particle from the origin after 7.0 s is

A 1.0 m **B** 3.0 m **C** 5.0 m **D** 7.0 m

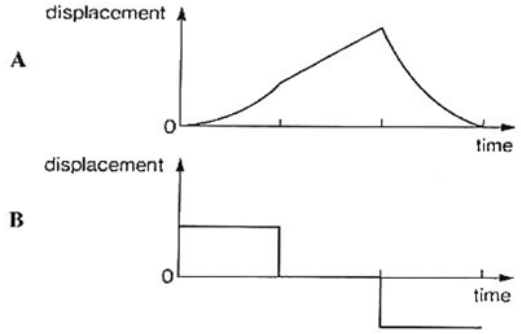
S7 **[J92/I/3]** The graph represents how displacement varies with time for a vehicle moving along a straight line. During which time interval does the acceleration of the vehicle have its greatest numerical value?

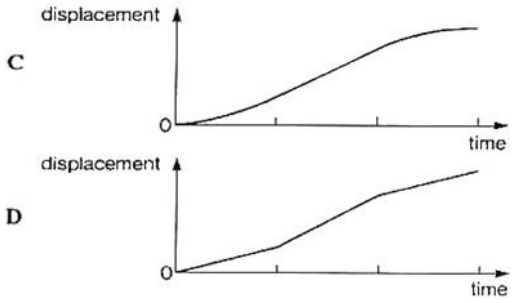


S8 **[N95/I/5]** The graph of velocity against time for a moving object is shown.



Which of the following is the corresponding graph of displacement against time?





S9 A hot air balloon, initially at ground, is travelling vertically upward at a constant speed of 5.0 m s^{-1}. When it is 21.0 m above the ground, a package is released from the balloon.

(a) How long after being released, is the package in the air?

(b) What is the velocity of the package just before it hits the ground?

Discussion Questions

These questions require more thinking and they are worth further discussions during tutorial sessions. You should attempt these questions with proper working on foolscap, to your best ability, and prepare to share your work as well as to learn from others their different approaches. Very often, in learning physics, the process of getting the answer is more important than the answer itself.

D1 Give examples to answer the following questions.

(a) Is it possible for a body to be accelerating while travelling at constant speed?

(b) Can a body have a constant velocity and a varying speed?

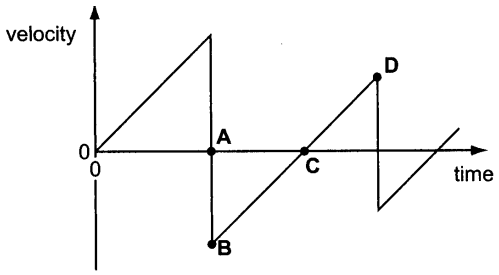
(c) Can the direction of a body's velocity change when it has a constant acceleration?

(d) If a body has zero velocity at an instant, can it be accelerating?

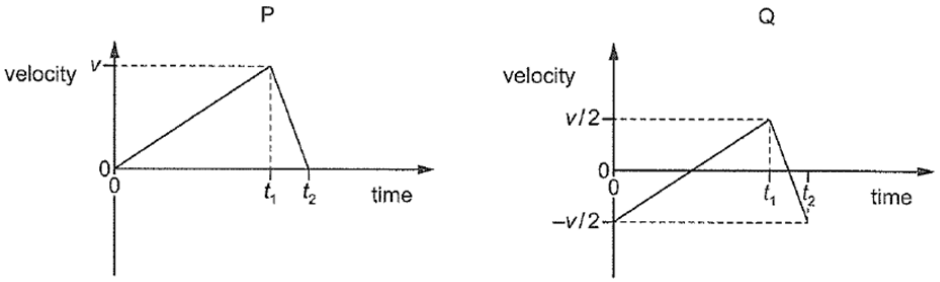
(e) Can a body be accelerating in a direction opposite to that of its velocity?

(f) If the average velocity of an object is non-zero for some interval, does this mean that the instantaneous velocity is never zero during this interval?

D2 A ball is released from rest above a hard, horizontal surface. The graph shows how the velocity of the bouncing ball varies with time. At which point on the graph does the ball reach its maximum height after the first bounce?



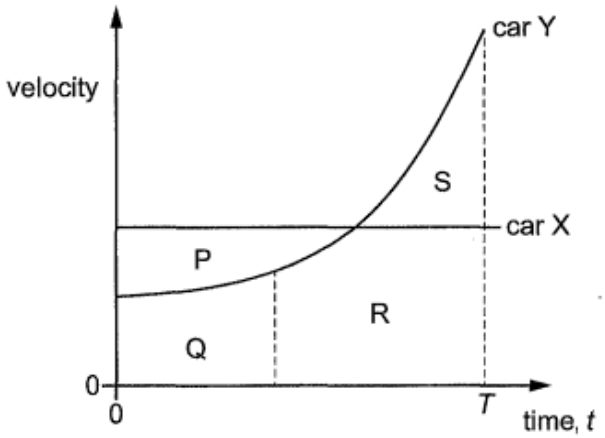
D3 [N18/I/2] The motions of two objects P and Q along a straight line are shown in the following velocity-time graphs.



How do the magnitudes of the displacement and the total distance travelled compare for the two objects from time 0 to time t_2?

	displacement	**distance**
A	different	different
B	different	same
C	same	different
D	same	same

D4 [N20/I/4] The velocity-time graph shows car X travelling with a constant velocity and car Y travelling with increasing acceleration. Car X overtakes car Y at time $t = 0$.



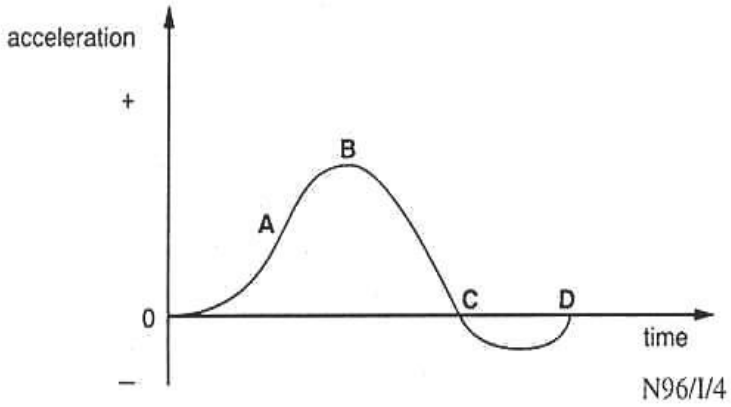
Four areas P, Q, Rand S are labelled on the graph.

Car Y overtakes car X at time $t = T$.

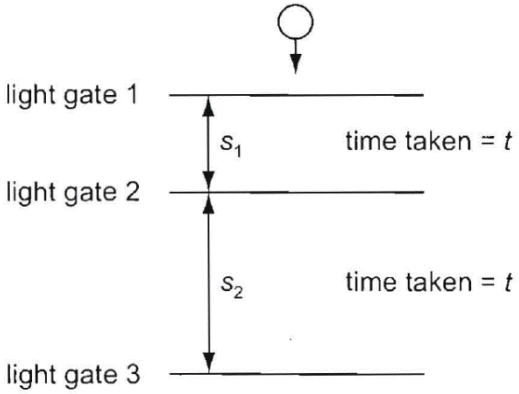
Which relationship must be correct?

A P = S **B** R = S **C** P + Q = R **D** Q + R = P + S

D5 [N96/I/4] A car is travelling along a straight road. The graph shows the variation with time of its acceleration during part of the journey. At which point on the graph does the car have its greatest velocity?



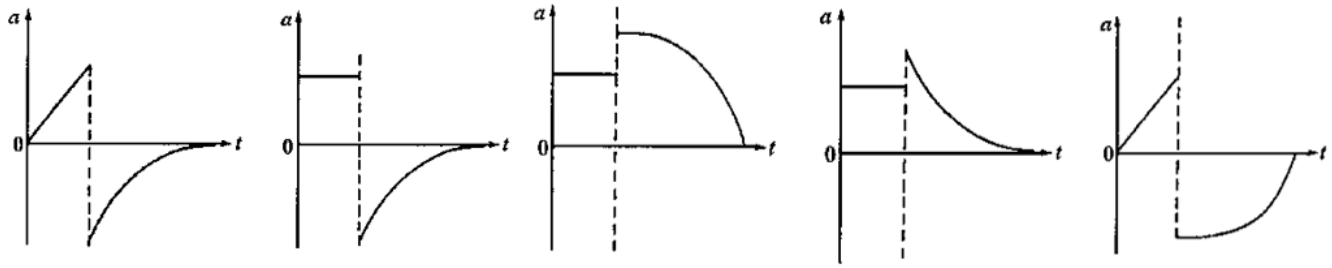
D6 [N10/I/5] An object falls freely with a constant acceleration *a* from above three light gates. It is found that it takes a time *t* to fall between the first two light gates a distance of s_1 apart. It then takes an additional time, also *t,* to fall between the second and third light gates a distance s_2 apart.



What is the acceleration in terms of s_1, s_2 and t ?

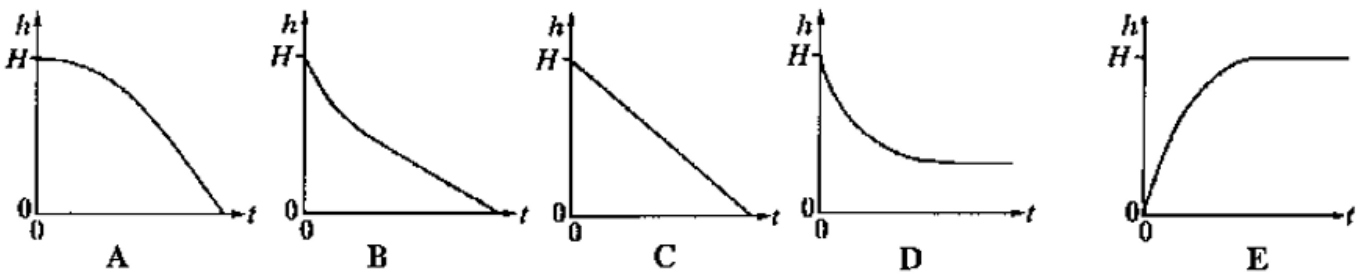
A $(s_2 - s_1)/t^2$ **B** $(s_2 - s_1)/2t^2$ **C** $2(s_2 - s_1)/3t^2$ **D** $2(s_2 - s_1)/t^2$

D7 [J83/II/5] A small metal sphere is released from rest at a height of a few centimetres above the surface of a viscous liquid. On entering the liquid, the sphere experiences a viscous drag proportional to its velocity. Which one of the following graphs most closely represents the variation of acceleration a with time t of the sphere?



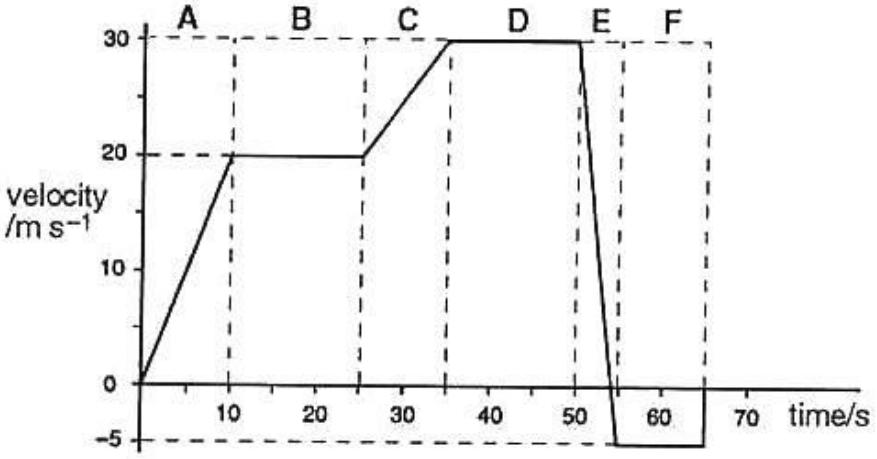
A **B** **C** **D** **E**

D8 [N83/II/4] A steel ball bearing is released at the surface of a viscous liquid contained in a tall, wide jar. In falling through the liquid, the ball bearing experiences a retarding force proportional to its velocity. If the depth of the liquid in the jar is H, which one of the following graphs best represents the variation of the height h of the ball bearing above the base with t?



A **B** **C** **D** **E**

D9 [N97/II/1] Figure below shows a velocity-time graph for a journey lasting 65 s. It has been divided up into six sections for ease of reference.



(a) Using the information from the graph, obtain:

(i) the velocity 10 s after the start,

(ii) the acceleration in section A,

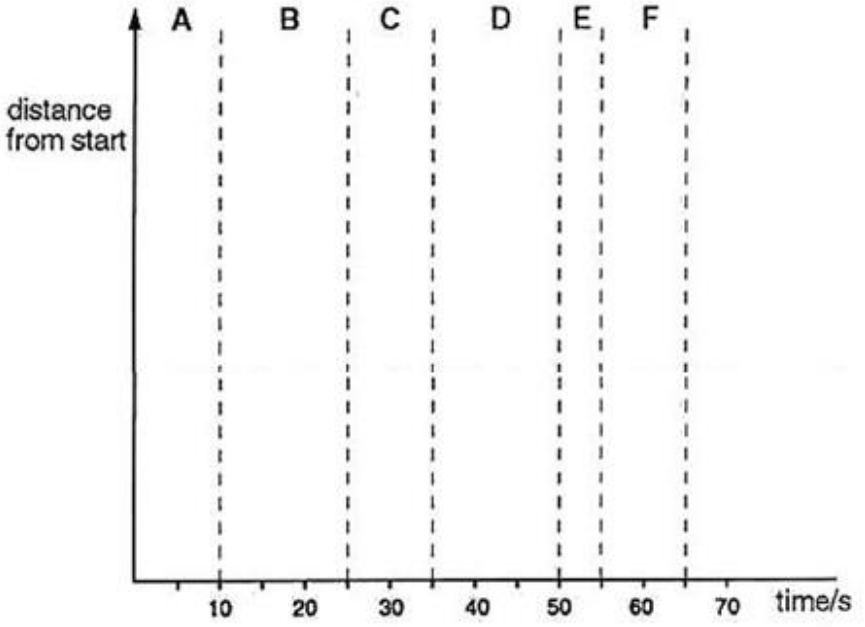
(iii) the acceleration in section E,

(iv) the distance travelled in section B,

(v) the distance travelled in section C. [5]

(b) Describe qualitatively in words what happens in sections E and F of the journey. [4]

(c) Copy the figure below into your answer script and sketch the shape of the corresponding distance-time graph. You are not expected to make detailed calculations of the distance travelled. [3]



D10 [N17/III/1] An object is dropped from a hot-air balloon. The variation with time *t* of the vertical velocity *v*

of the object for the first 8.0 s of its descent is shown in Fig. 1.1.

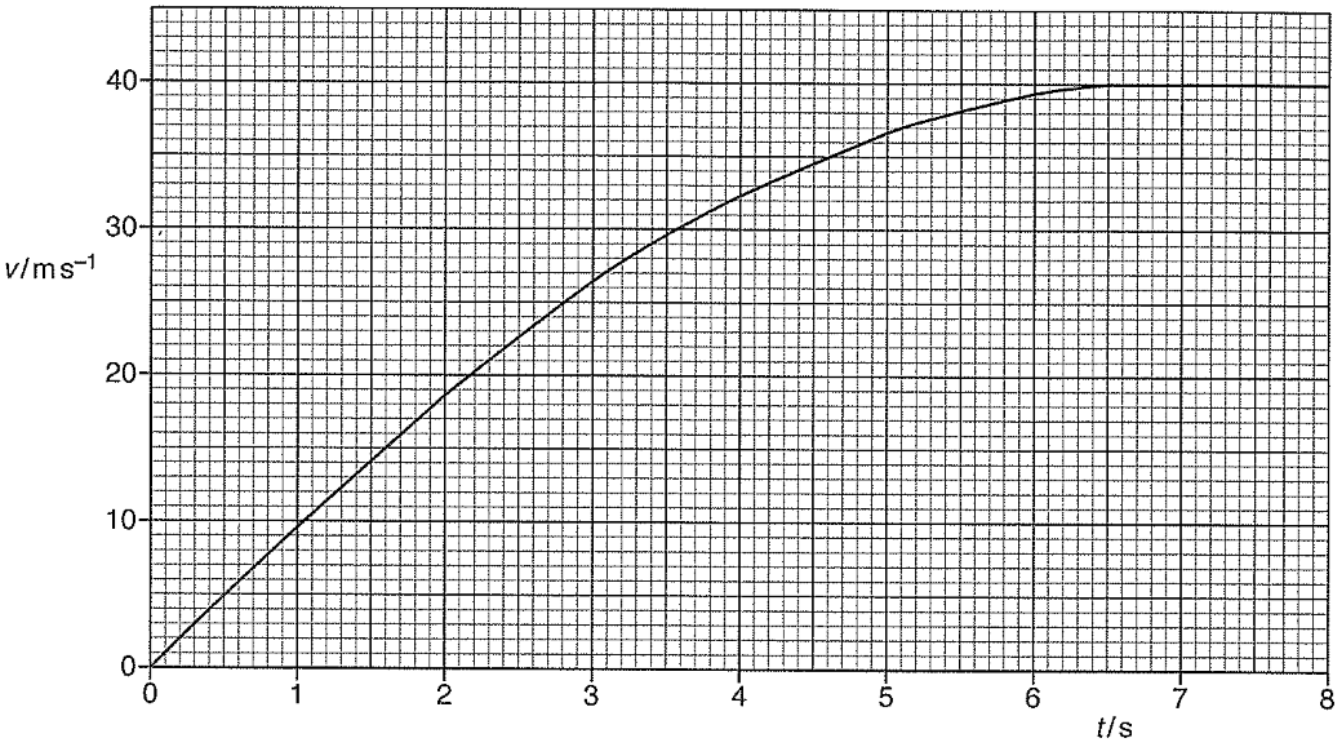


Fig. 1.1

(a) Explain why the object reaches a constant velocity. [3]

(b) A student suggests that the resistive force *F* acting on the object of mass *m* is proportional to its velocity. That is

$$F = kv$$

where k is a constant.

At speed *v*, the acceleration of the object is *a*.

(i) Show that, based on the suggestion of the student, and without using data from the graph in Fig. 1.1,

$$(g - a) = kv / m$$

where *g* is the acceleration of free fall. [2]

(ii) Use information from Fig. 1.1 or otherwise to complete Fig. 1.2. [4]

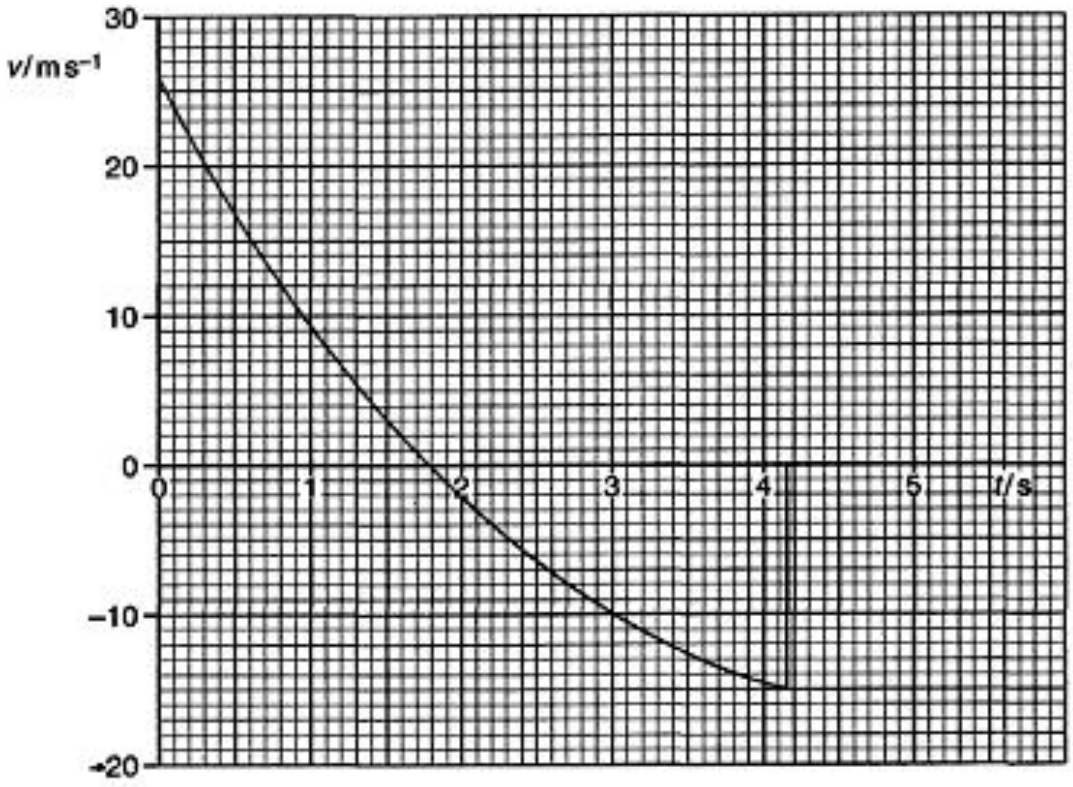
velocity v/m s^{-1}	acceleration a/m s^{-2}	$(g - a)$/m s^{-2}
0	.....	.....
20	8.2	1.6
30	.....	.....
40	.....	.....

Fig. 1.2

(iii) Use completed Fig. 1.2 to deduce whether the student's suggestion for velocities of 20 m s^{-1} and 30 m s^{-1} is correct. [2]

D11 [N03/III/1(c), (d) and (e)]

(c) The graph below shows the variation with time *t* of the velocity *v* of a ball from the moment it is thrown with a velocity of 26 m s^{-1} vertically upwards.



(i) State the time at which the ball reaches its maximum height. [1]

(ii) State the feature of a velocity-time graph that enables the acceleration to be determined. [1]

(iii) Just after the ball leaves the thrower's hand, it has a downward acceleration of approximately 20 m s^{-2}. Explain how this is possible. [2]

(iv) State the time at which acceleration is *g*. Explain why the acceleration has this value only at this particular time. [2]

(v) Sketch an acceleration-time graph for the motion (in the air). Show the value of *g* on the acceleration axis. [3]

(d) Explain why, for all real vertical throws, the time taken to reach maximum height must be shorter than the time taken to return to the starting point. [2]

(e) The ball in **(c)** starts with kinetic energy of 54 J.

(i) Calculate the mass of the ball. [2]

(ii) Describe qualitatively how the amount of kinetic energy changes during the motion. [2]

Answers to Tutorial 2A: Kinematics - 1D Motion

S1 (a) 20 m s^{-1}

(b) The change in velocity is 26.5 m s^{-1} at an angle of 79.1° anticlockwise from the horizontal

(c) 15.7 m s^{-1}

(d) The average velocity is 13.6 m s^{-1} at 33.0° to the initial velocity

S2 (a) 169 km, bearing of 33.0° (b) 60 km h^{-1}, bearing of 33.0°

S3 (a) 3.14 s (b) 2.55 m s^{-1}, A to B (c) 8.0 m s^{-1}, vertically upwards

S4 (a) -0.83 m s^{-1}, -0.67 m s^{-1}, (allow -0.61 m s^{-1} to -0.88 m s^{-1})

S5 C

S6 B

S7 B

S8 C

S9 (a) 2.64 s (b) -20.9 m s^{-1}

D9 (a)(i) 20 m s^{-1} (ii) 2 m s^{-2} (iii) -7 m s^{-2} (iv) 300 m (v) 250 m

D10 b(ii) acceleration: 9.8, 8.2, 5.8, 0 ; (g – a): 0, 1.6, 4.0, 9.8

D11 (c)(i) 1.80 s (iv) 1.80 s (e) (i) 0.160 kg

Tutorial 2B Kinematics - 2D Motion

Self-Review Questions

S1 As a projectile moves through its parabolic trajectory, which of these quantities, if any, remain constant: (a) speed, (b) acceleration, (c) horizontal component velocity, (d) vertical component velocity?

S2 A transport plane travelling horizontally at a steady speed of 50 m s^{-1} at an altitude of 300 m releases a parcel when directly above a point X on level ground. Calculate

(a) the time of flight of the parcel,

(b) the velocity of impact of the parcel,

(c) the horizontal distance from X to the point of impact.

S3 The angle of elevation of an anti-aircraft gun is 70° and the muzzle velocity is 900 m s^{-1}. How long after firing will the shell first be at an altitude of 1700 m?

A 2.0 s **B** 4.0 s **C** 6.0 s **D** 8.0 s

S4 A ball is projected horizontally from the top of a cliff on the surface of the Earth with a speed of 40 m s^{-1}. Assuming that there is no air resistance, what will its speed be after 3 s?

A 30 m s^{-1} **B** 40 m s^{-1} **C** 50 m s^{-1} **D** 60 m s^{-1}

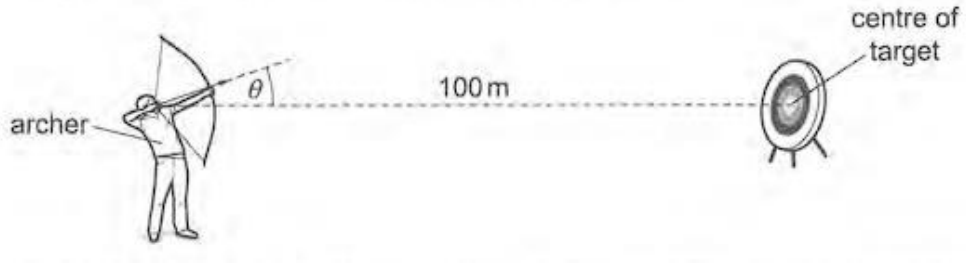
S5 A stone thrown horizontally at a speed of 24 m s^{-1} from the top of a cliff takes 4.0 s to hit the sea. What is the height *h* of the cliff-top above the sea and the distance *d* from the base of the cliff to the point of impact?

	Height *h*/m	Distance *d*/m
A	78	96
B	78	116
C	90	96
D	90	116

S6 Water shoots out from a horizontal pipe which is at a height of 52 cm from the floor. If the horizontal distance travelled by the water before it hits the floor is 100 cm, what is the velocity of the water when it leaves the pipe?

S7 A particle can be projected at 40 m s^{-1}. Find the angle(s) of projection required for it to achieve a horizontal range of 150 m.

S8 [N22/I/2] An archer is standing level with the centre of a target, as shown.



The archer is at a distance of 100 m from the target. The length of the arrow is negligible.

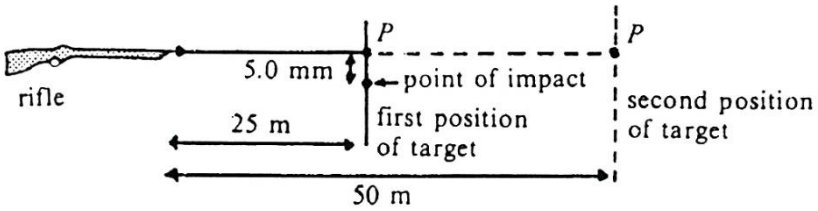
The archer fires an arrow at an angle θ above the horizontal. The arrow hits the centre of the target 2.0s later. Air resistance is negligible.

What is θ?

A 2.8° **B** 5.5° **C** 11° **D** 79°

Discussion Questions

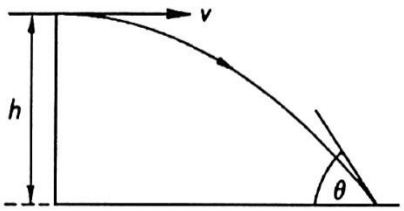
D1 [N82/II/4] When a rifle is fired horizontally at a target *P* on a screen at a range of 25 m, the bullet strikes the screen at a point 5.0 mm below *P*. The screen is now moved to a distance of 50 m and the rifle again fire horizontally at *P* in its new position. See figure below.



Assuming that the air resistance may be neglected, what is the new distance below *P* at which the screen would now be struck?

A $5\sqrt{2}$ mm **B** 10 mm **C** 15 mm **D** 20 mm **E** 25 mm

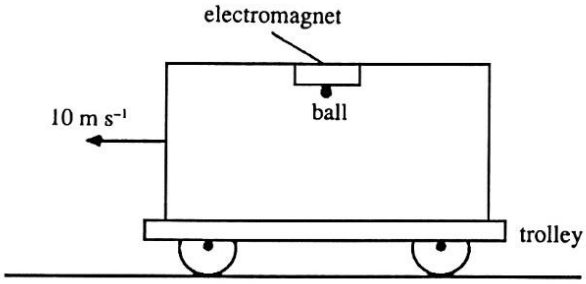
D2 [J88/I/3] The diagram shows the path of a projectile fired with a horizontal velocity *v* from the top of a cliff of height *h*.



Which of the following values for *v* and *h* will give the greatest value of the angle θ?

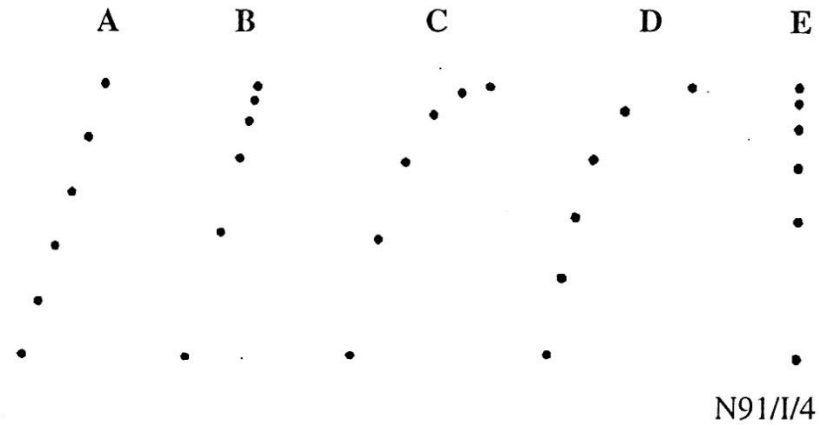
	v/m s^{-1}	*h*/m
A	10	30
B	10	50
C	30	30
D	30	50
E	50	10

D3 [N91/I/14] A ball is suspended from an electromagnet attached to a trolley which is travelling at a constant speed of 10 m s^{-1} to the left. The trolley is illuminated by a stroboscope flashing at a constant rate. The diagram represents the viewpoint of a stationary camera.



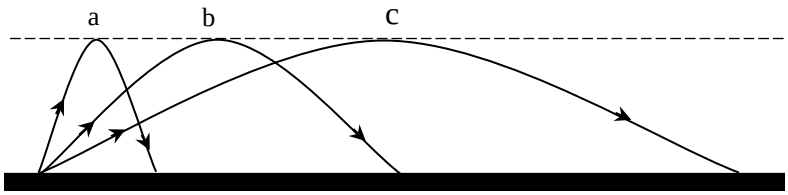
The ball is released and a series of stroboscopic images of the ball are recorded on a single photographic plate.

Which diagram best represents what is seen on the photographic plate?



N91/I/4

D4 Trajectories are shown in the figure below for three kicked footballs.



Ignoring the effects for air resistance on the footballs, order the trajectories from smallest to largest, indicating any ties, according to:

(a) initial vertical velocity component,

(b) time of flight,

(c) initial horizontal velocity component and

(d) initial speed.

D5 A Hollywood daredevil plans to jump the canyon shown in the figure. If he desires a 3.0 s flight time,

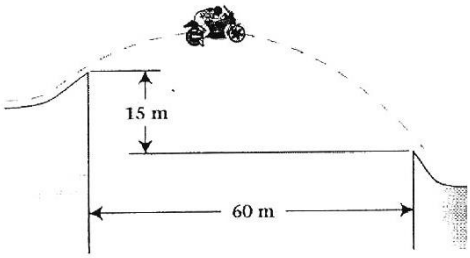
Calculate

(a) the correct angle for his launch ramp,

(b) his correct launch speed,

(c) the correct angle for his landing ramp, and

(d) his predicted landing speed.



D6 A cannon ball is projected at an angle of 30^{o} above horizontal and at a speed of 15.0 m s^{-1} from the top of a cliff. The top of the cliff is at a height of 100.0 m above the sea level. A hot-air balloon is at a horizontal distance of 50.0 m away from the bottom of the cliff and rises vertically at a constant speed. When the hot-air balloon is at a height of 30.0 m from sea level, the cannon ball is launched. What is the speed of the balloon if the cannon ball hit it?

D7 A ball is projected horizontally from the top of a building. One second later, another ball is projected horizontally from the same point with the same velocity.

(a) At what point in the motion will the balls be closest to each other?

(b) Will the first ball always be travelling faster than the second ball?

(c) How much time passes between the moments the first ball and the second one hit the ground?

(d) Can the horizontal projection velocity of the second ball be changed so that the balls arrive at the ground at the same time?

D8 [2010 C1 BT/II/2] A person stands on top of a hill at point **O** and throws a stone at an initial angle of 60° with respect to the horizontal with a speed of 10 m s^{-1}. The stone landed at point A on the slope *t* seconds later. Point **A** is *d* metres from point **O**, in the direction 45° down slope as shown in Figure 9. You may neglect air resistance.

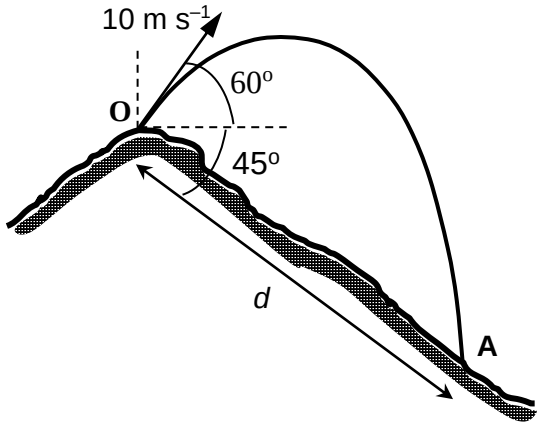


Figure 9

(a) Write down the expressions, in terms of *t*, for
(i) the horizontal displacement and [1]
(ii) the vertical displacement of the stone at point **A** with respect to point **O**. [1]

(b) Calculate *t*. [3]

(c) Find *d*. [1]

D9 From a point O at a height *h* above the ground, two particles P and Q are simultaneously projected with initial speed *u* at an angle of θ with the horizontal as shown below. The acceleration due to gravity is represented by *g*.

(a) Copy the diagram above into your foolscap and sketch the trajectories of P and Q in your diagram

(b) Obtain an expression in terms of *u, g* and θ, the distance between the points of impact of P and Q.

Answers to Tutorial 2B: Kinematics - 2D Motion

S1 (a) Not constant (b) Constant (c) Constant (d) Not constant

S2 (a) 7.82 s (b) 91.6 m s^{-1}, 57.0° below the horizontal (c) 391 m

S3 A

S4 C

S5 A

S6 3.07 m s^{-1}

S7 33.4° and 56.6°

S8 C

D5 (a) 26° (b) 22 m s^{-1} (c) 45° (d) 28 m s^{-1}

D6 6.81 m s^{-1}

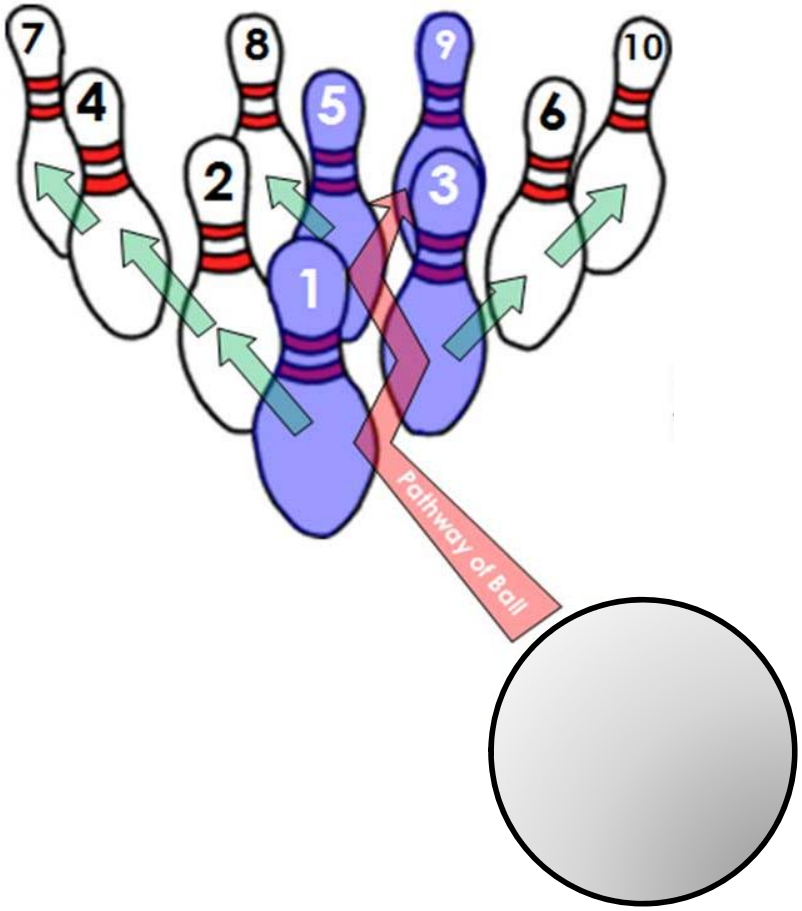
D7 (c) 1 s

D8 (b) 2.78 s (c) 19.7 m

D9 (b) $\frac{u^2 \sin 2\theta}{g}$

Chapter 3

Dynamics



"Although you can get a strike by hitting the head pin square-on, it is not a very high percentage shot. Hitting the pins in "the pocket", which is the gap between the head pin and the next pin on the right (the 3-pin) for right handed bowlers, and the next pin to the left (the 2-pin) for lefties, will strike with the most consistency."

(Neil Lazarski, "How to Raise Your Bowling Average")

H2 Physics Syllabus 9749

Topic 3: Dynamics

Content

- Newton's laws of motion
- Linear momentum and its conservation

Learning Outcomes

Students should be able to:

(a) state and apply each of Newton's laws of motion.

(b) show an understanding that mass is the property of a body which resists change in motion (inertia).

(c) describe and use the concept of weight as the force experienced by a mass in a gravitational field.

(d) define and use linear momentum as the product of mass and velocity.

(e) define and use impulse as the product of force and time of impact.

(f) relate resultant force to the rate of change of momentum.

(g) recall and solve problems using the relationship $F = ma$, appreciating that resultant force and acceleration are always in the same direction.

(h) state the principle of conservation of momentum.

(i) apply the principle of conservation of momentum to solve simple problems including inelastic and (perfectly) elastic interactions between two bodies in one dimension. (Knowledge of the concept of coefficient of restitution is not required.)

(j) show an understanding that, for a (perfectly) elastic collision between two bodies, the relative speed of approach is equal to the relative speed of separation.

(k) show an understanding that, whilst the momentum of a closed system is always conserved in interactions between bodies, some change in kinetic energy usually takes place.

Topic 4: Forces (Part)

Learning Outcomes

Students should be able to:

(b) describe the forces on a mass, charge and current-carrying conductor in gravitational, electric and magnetic fields, as appropriate.

(c) show a qualitative understanding of normal contact forces, frictional forces and viscous forces including air resistance. (No treatment of the coefficients of friction and viscosity is required.)

(d) show an understanding that the weight of a body may be taken as acting at a single point known as its centre of gravity.

Contents

3.1 Introduction 4
3.1.1 Types of Forces 4
3.1.2 Weight 5
3.1.3 Contact Forces 6
3.1.4 Tension 7
3.2 Newton's Laws of Motion 8
3.2.1 Newton's 1st Law 8
3.2.2 Linear Momentum 8
3.2.3 Newton's 2nd Law 10
3.2.4 Impulse 11
3.2.5 Newton's 2nd Law for Constant Mass 13
3.2.6 Newton's 3rd Law 16
3.3 Principle of Conservation of Linear Momentum 18
3.3.1 Two-body Systems 20
3.3.2 Types of Collisions 21

Appendix A: Static and Kinetic Friction 27
Appendix B: Newton's Law in Car Safety 28

Tutorial 3: Dynamics 29
Self-Review Questions 29
Discussion Questions 31

Videos of Lecture Examples can be found at
https://youtube.com/playlist?list=PL_b5cjrUKDlYDkaMJ2sKNvZNiASvpiSfT

3.1 Introduction

Dynamics is the branch of classical mechanics concerned with the study of forces and their effects on motion. Isaac Newton defined the fundamental physical laws which govern dynamics in physics, especially his second law of motion.

3.1.1 Types of Forces

Broadly speaking, forces may be categorized into two kinds: **contact** forces and **non-contact** forces.

3.1.1.1 Contact Forces

Forces acting between two bodies that are in physical contact.

Contact forces	Description
Friction	Force that resists the **relative** motion or the tendency for **relative** motion between two surfaces that are in contact.
Normal contact force	Net electrostatic repulsion between two surfaces that are in contact.
Contact force	Vector sum of friction and normal contact force, meaning the total contact force.
Spring force	Force exerted by one end of a spring when the spring is extended or compressed from its natural length. The spring forces acts throughout a spring, in both directions.
Tension	Force exerted by one end of a string/rope which is being pulled taut. The tension is constant throughout the string/rope and acts in both directions.
Upthrust	**Net** upward force exerted by the fluid on a body fully or partially submerged in the fluid.
Viscous drag force	Force that resists a body moving **relative** to a fluid.
Air resistance	Drag force exerted by air on a body moving in air. Air resistance is a specific form of drag.

3.1.1.2 Non-contact Forces

Forces acting between two bodies which are NOT in physical contact (separated by a distance). This characteristic of the force is described as **action-at-a-distance**.

Non-contact forces	Description	Physical property of a body on which the force acts on.
Gravitational force	Attractive force exerted between masses.	Mass
Electric force	Attractive/repulsive force exerted between electrically charged bodies.	Electric charge
Magnetic force	Force exerted between currents	Moving charges

3.1.2 Weight

The weight of a body is due to the body being in a gravitational field. In most cases, the object of interest is close to the Earth and its weight refers to the gravitational force exerted on its body by the Earth. However, an object may have a different weight elsewhere, for instance, when it is in the gravitational field of the Moon.

Notes:

- Weight is a **force** (hence a **vector**).
- Direction: For everyday-life problems, the weight of the object is taken to act **vertically downwards** at the centre of gravity of the object.
- Magnitude: $W = mg$

  W: weight of an object (SI unit: N)
  m: mass of the body (SI unit: kg)
  g: acceleration of free fall/ gravitational field strength at the point (SI unit: m s^{-2})

- Weight is a non-contact force. It does not only act on a person sitting in a chair, but also on a flying bird, regardless of whether we take into account air resistance.

Distinction between mass and weight

Mass	Weight
A measure of the inertia of a body	A measure of the gravitational pull on a body
Scalar quantity	Vector quantity
SI unit: kg	SI unit: N
Same regardless of location	Varies with gravitational field strength at each location
Same regardless of the method of measurement.	Some methods of measurement do not measure the true weight but the apparent weight (see example 7)

In everyday language, the terms ‘mass’ and ‘weight’ are used interchangeably, e.g. I “weigh” 70 kg. In physics, and especially during examinations, such casual usage is <u>not tolerated</u>.

3.1.2.1 Centre of Gravity

An object of a significant size has its mass distributed (not necessarily uniformly) throughout its volume.

The **centre of gravity** of an object is the point at which the weight of the object *appears to* act.

Notes:

- In free-body diagrams, the weight should be drawn from the centre of gravity of an object.
- The centre of gravity of an object is not necessarily on the object. It may lie outside the object.
- In a uniform gravitational field, the centre of gravity always coincides with the centre of mass of the object.

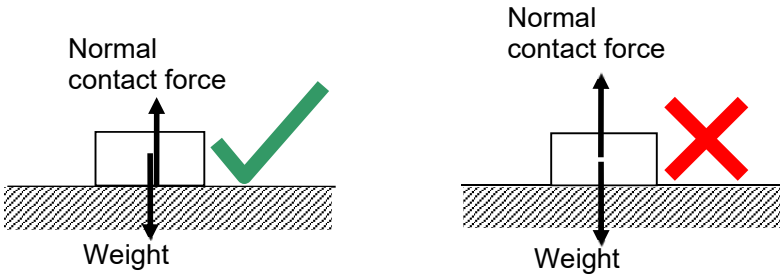
3.1.3 Contact Forces

When two surfaces come into contact, they both experience contact forces arising from electrostatic forces between atoms of the surfaces. For ease of analysis, this contact force is often resolved into two components: the component perpendicular to the surfaces is called **normal contact force**, and the component parallel to the surfaces is given the name **friction**.

3.1.3.1 Normal Contact Force

A block can rest on a table because its weight is balanced by an upward force the table exerts on it. This force arises from the electrostatic repulsion between the molecules of the table and the molecules of the block. It is called the normal contact force since it acts perpendicularly (normal) to the surfaces.

In free-body diagrams (see Section 3.2.5.1 for details), contact forces must be drawn from a contact point (e.g. somewhere along the base of the block). A common mistake is contact forces that are drawn from the centre of gravity (see diagram below).



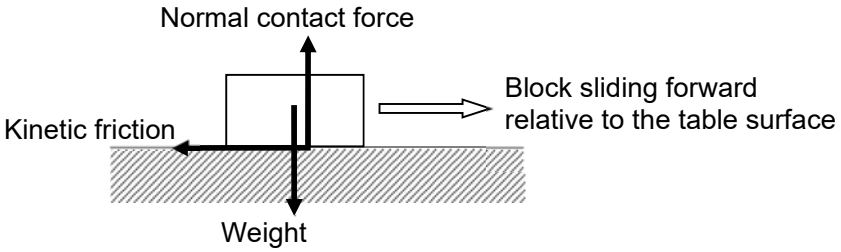
- By definition, the normal contact force is always directed **perpendicular to the surfaces**.
- The magnitude depends on how hard the two surfaces are pressed into each other.

3.1.3.2 Friction

Friction is the force that resists relative motion between two sliding surfaces (called kinetic friction), or the tendency for relative motion between two stationary surfaces (called static friction).

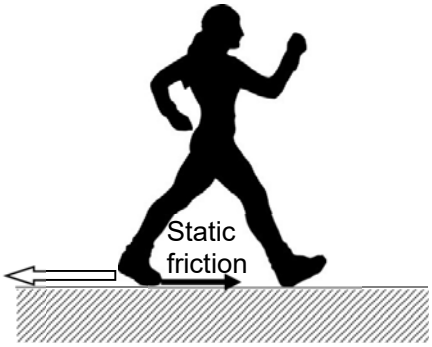
Direction
Friction is always directed in the direction opposite to the direction of **relative motion**.

For example, a block sliding along a rough table is gradually decelerated to rest. It experiences a (kinetic) friction directed backward because the block's surface is sliding forward relative to the table's surface.



On the other hand, a person is able to walk forward only if he experiences a (static) friction that is directed forward. The sole of his shoe tends to rub backward relative to the floor's surface. When an exam question refers to friction, you have to infer from the question whether the relevant force is a *retarding* frictional force or a *propelling* frictional force.

Sole rubbing backward relative to the surface

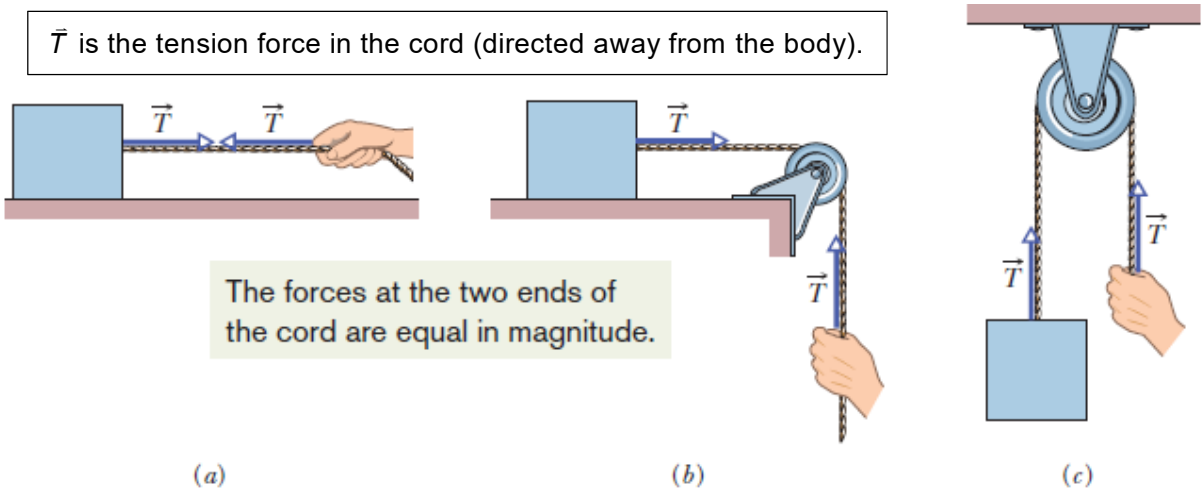


3.1.4 Tension

In physics, tension refers to the magnitude of the pulling force exerted by a string, rope, cable or chain on the object tied to its end when it is pulled **taut**.

- *At the end of a string*, the direction of the tension force exerted on the object is **along the string**, directed **away from the object**.
- *Within the string*, the magnitude of the tension is constant, and the tension acts in both directions equally. The magnitude of the tension in the string is equal to the magnitude exerted at the ends.
- The magnitude of the tension force depends on how much the string is stretched. Once the string is **slack**, the tension falls to zero.
- In many situations, we idealise the string, rope, cable, etc. to be **massless** and **inextensible** to simplify analysis.
- Strings or ropes are often used together with pulleys. Again, in many situations, we idealise a pulley to be **massless** and **frictionless** (at the axle) to simplify analysis.
- The tension is equal to the magnitude of the forces applied by the ends of the string, even if the bodies and the cord are accelerating and even if the cord runs around a massless, frictionless pulley (figs. b and c).

$\vec{T}$ is the tension force in the cord (directed away from the body).



(a) (b) (c)

3.2 Newton's Laws of Motion

3.2.1 Newton's 1st Law

Statement of Newton's 1st law of motion

A body stays at rest or continues to move with a constant speed in a straight line unless acted upon by a ***net*** *external force*.

Notes

- Changes in a body's state of motion refer to changes in its velocity. It can be a change in either its **speed** or **direction** (or both).
- An object's resistance to change its state of motion is called **inertia**. As such, Newton's 1st law is also referred to as the Law of Inertia.
- An object's inertia is dependent on its mass: The larger its mass, the larger its inertia.

3.2.2 Linear Momentum

Everyday experience tells us that two objects moving at the same velocity can result in very different consequences. Consider a 3-gram ping pong ball and a 3-ton lorry, both travelling at 5 m s^{-1}. Which would you rather collide with?

In physics, we define a "quantity of motion" called **momentum**, ***p***.

Definition of Linear Momentum

The linear momentum of a body is the product of its mass and its velocity.
Mathematically, linear momentum, $\vec{p} = m\vec{v}$

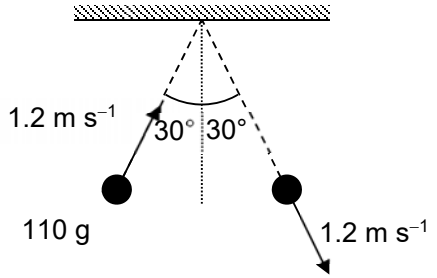
Notes

- Momentum is a **vector quantity**. It has the same direction as the object's velocity.
- Momentum has the SI unit: **kg m s^{-1}**. Sometimes the alternative unit **N s** is used.

In H2 Physics, when we refer to momentum, we mean linear momentum. It is not to be confused with angular momentum, which is not in the syllabus.

Example 1a: Change in Momentum

A 110 g billiard ball rebounds off a wall, with velocities as shown in the (top view) diagram. The ball moves on a horizontal plane. Evaluate the change in momentum of the ball.



3.2.3 Newton's 2nd Law

Statement of Newton's 2nd law of motion

The ***rate of change of linear momentum*** of a body is ***directly proportional*** to the ***resultant force***[1] acting on it and is in the direction of the resultant force.

Strictly speaking, Newton's 2nd law only states that $\vec{F}_{net} \propto \frac{d\vec{p}}{dt}$.

Hence, $F_{net} = k\frac{dp}{dt}$, where k is some constant of proportionality. When SI units are used for all quantities in the equation, the constant k has a value of 1.

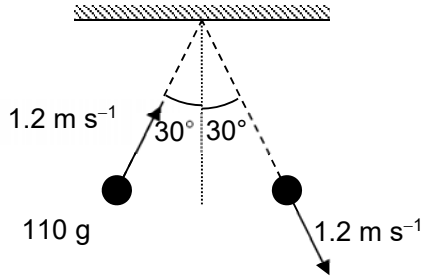
Mathematically, net force, $\vec{F}_{net} = \frac{d\vec{p}}{dt} = \frac{d(m\vec{v})}{dt}$

Average net force, $\langle F_{net} \rangle = \frac{\Delta\vec{p}}{\Delta t} = \frac{\Delta(m\vec{v})}{\Delta t}$

The SI unit for force is the newton (N).

Example 1b: Rate of Change in Momentum

The time interval of contact between the billiard ball and the wall is 0.100 s. Find the average net force acting on the wall during the collision.



[1] Resultant force is also known as net force, or (vector) sum of forces.

3.2.4 Impulse

Definition of Impulse

The impulse of a force is the product of the **average force** and the time interval over which it is applied.

Mathematically, $\vec{J} = \langle \vec{F} \rangle \Delta t$

Notes

- Impulse is a **vector** quantity; its direction is the same as that of the average force *F*.
- The SI units for impulse are **N s**. They can be shown to be equivalent to **kg m s^{-1}**, the same as the units of momentum.

3.2.4.1 Relation between Impulse and Momentum

From Newton's second law,

$$\vec{F} = \frac{d\vec{p}}{dt}$$

Rearranging and integrating, we obtain

$$\int \vec{F} dt = \int d\vec{p}$$

The left-hand side of the equation, $\int \vec{F}\, dt$, is the **impulse** of the force.

The right-hand side, $\int d\vec{p} = \vec{p}_f - \vec{p}_i = \Delta\vec{p}$, is the **total change in momentum**.

This result, which links impulse to the change in momentum, is known as the **impulse-momentum theorem**.

Impulse-Momentum Theorem

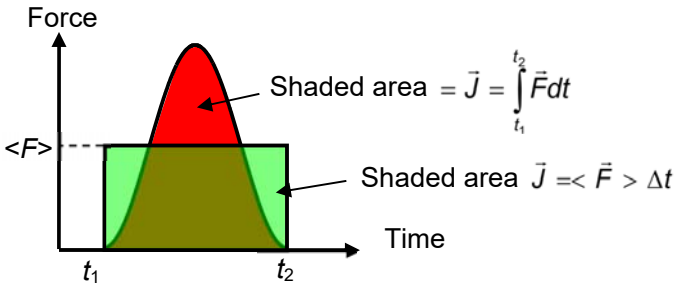
The impulse-momentum theorem states that the ***total change in*** **momentum** of a particle during a time interval **equals the impulse of the net force** that acts on the particle during that interval.

Mathematically, $\vec{J} = \int \vec{F} dt = \Delta\vec{p}$

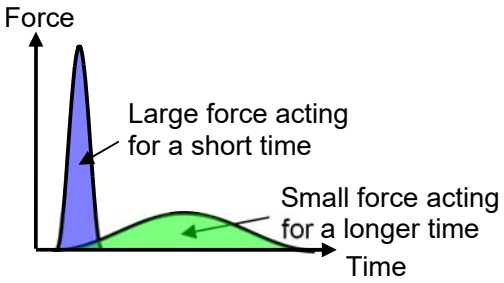
For **average** or **constant** forces, this simplifies to $\vec{J} = < \vec{F} > \Delta t = \Delta\vec{p}$.

3.2.4.2 Graphical Representation of Impulse

Since $\vec{J} = \int \vec{F} dt = \Delta \vec{p}$, impulse is represented by the **area under the *force-time* graph**. The same area also represents the **(total) change in the momentum** of the body.



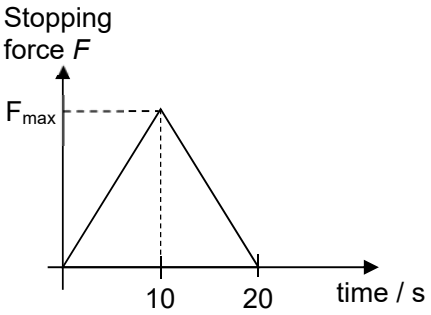
To illustrate further, study the *F-t* graphs below. Notice that, even though the two forces have different peak magnitudes, they both deliver the same impulse (i.e. same area under each graph). Hence, the two forces will cause the same change in momentum to the object they are applied on.



Example 2 (N89/II/8 modified)

In order to stop a car of mass 1500 kg travelling at 30 m s^{-1}, the driver applies his brakes so that *F*, the total stopping force, increases steadily to a maximum and then decreases to zero as shown in the figure.

Calculate
(a) the momentum of the car when it is travelling at 30 m s^{-1},
(b) the impulse due to the braking force,
(c) the magnitude of the average stopping force, <*F*>,
(d) the value of F_{max}.



Answer:

3.2.5 Newton's 2nd Law for Constant Mass

From Newton's 2nd Law, $F_{net} = \frac{dp}{dt} = \frac{d(mv)}{dt}$

For constant mass, $\vec{F}_{net} = m\frac{d\vec{v}}{dt}$

$\vec{F}_{net} = m\vec{a}$

Notes

- $\vec{F}_{net} = m\vec{a}$ is **not** a complete representation of Newton's second law. It only represents a special case of Newton's second law for constant mass.
- The product F_{net} represents the **resultant** of all the actual physical forces acting on a body. It is **not** a new or additional physical force. As such, we do not draw F_{net} as a separate force in free-body diagrams.
- Be clear about which is the cause (F_{net}), and which is the effect (acceleration).
  - A non-zero F_{net} *always results in* an acceleration (change in speed or direction).
  - If acceleration is detected, it *implies* F_{net} is non-zero. It does not result in a new force.

TIPS **Solving $F_{net} = ma$ Problems**

Step 1 Draw a picture of the problem if no diagram is provided.
Step 2 Consider the expected direction of acceleration and resultant force.
(This is to facilitate setting up the $F_{net} = ma$ equation.)
Step 3 Draw the free-body diagram (see below) for the body (or bodies) of interest. Only include forces that act on the object you are drawing the diagram for.
Step 4 Establish the coordinate system, with **one axis parallel to the expected acceleration**, and a perpendicular axis if necessary.
Step 5 Resolve the forces into components parallel to your chosen axis/axes.
Step 6 Note additional constraints. For example, two objects moving in physical contact or with a taut rope connected in between must have the same acceleration.
Step 7 Apply $F_{net} = ma$ and solve for all unknowns.
(If there are n unknowns, you have to solve a system of *n* equations.)

3.2.5.1 Free-Body Diagram

A free-body diagram (FBD) is a simple representation of an object and a vector representation of all the forces (including all contact and non-contact forces) acting on the object.

TIPS **Drawing Free-Body Diagrams**

Step 1 Draw a simplified diagram if one is not provided.
Step 2 Identify and isolate the body/bodies (or system) of interest.
Step 3 Draw **all** the forces acting **on** the body/bodies (or system) using labelled arrows.
- The forces must be drawn from the point of application of the forces.
- Draw the forces to appropriate lengths, since the relative lengths of the arrows represent the relative magnitude of the forces.
- Label the arrows clearly with words or properly defined symbols.

Step 4 For two-dimensional problems, resolve the forces into two mutually perpendicular directions to simplify analysis.
- E.g. horizontal and vertical direction or parallel to a slope and perpendicular to the slope.
- Note: the resolved components should be drawn in **dotted** lines so as to distinguish them from the actual physical forces.

Applying $F_{net} = ma$ to a system with forces in 1D - the importance of identifying the system

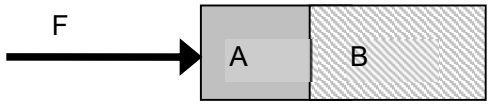
Example 3

A helicopter of mass *M* and weight *W* rises with vertical acceleration, *a*, due to the upward thrust *U* generated by its rotor. The crew and passengers of total mass *m* and total weight *w*, exerts a combined force *R* on the floor of the helicopter.

Draw an appropriate free-body diagram, and write down an equation for the motion of
(a) the helicopter, (b) the crew and passengers, (c) helicopter, crew and passengers

Example 4

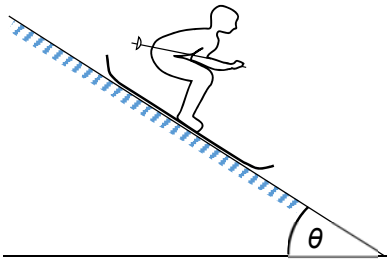
Two blocks A and B, of masses 2*M* and 4*M*, respectively, are pushed along a smooth horizontal surface by a force of *F* as shown in the diagram.
What is the magnitude of the force exerted by block A on block B during the acceleration?



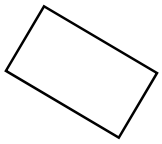
Applying $F_{net} = ma$ to a system with forces in 2D

Example 5

A 75 kg skier is accelerating at 2.6 m s^{-2} down a slope at an angle $\theta = 30°$. Find the friction f and the normal contact force N acting on the skier.



Step 1	Choose/Identify the system, in this case the skier, and draw a simple box to represent it.
Step 2	Draw the free-body diagram.
Step 3	For two-dimensional problems, resolve forces into two convenient, mutually perpendicular directions to simplify analysis. - We choose the components *along* and *perpendicular to* the slope. - The frictional force is parallel to the slope. - The normal contact force is perpendicular to the slope. - The weight has components along and perpendicular to the slope.



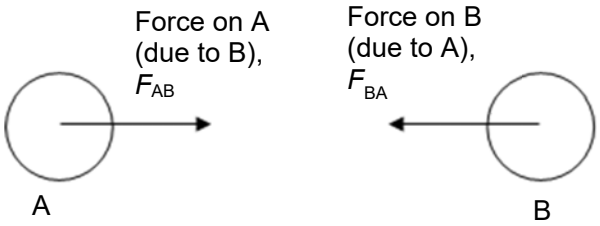
Step 4 Apply Newton's second law to the 2 perpendicular directions:

Don't know how to resolve the weight into the 2 components? Watch this:

3.2.6 Newton's 3rd Law

Statement of Newton's 3rd law

If body A exerts a force on body B, then body B will exert an equal and opposite force on body A.



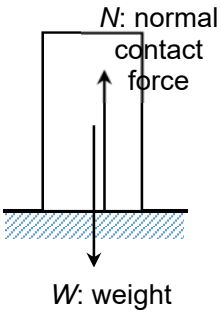
$$\vec{F}_{AB} = -\vec{F}_{BA}$$

It is easy to confuse action-reaction pairs with force pairs that happen to be pairs of "equal-but-opposite" forces. Performing the following "mental checklist" may help to detect "impostors".

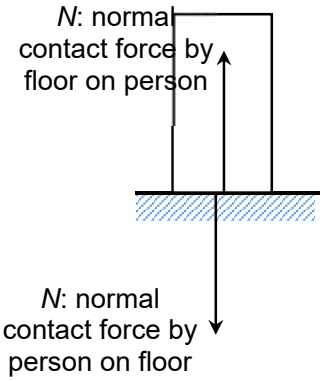
For two forces to be Action-Reaction Pair, the two forces must be

- ☐ Equal in magnitude
- ☐ Opposite in direction
- ☐ Acting on different bodies
- ☐ Of the same nature (e.g. gravity-gravity, friction-friction, tension-tension)

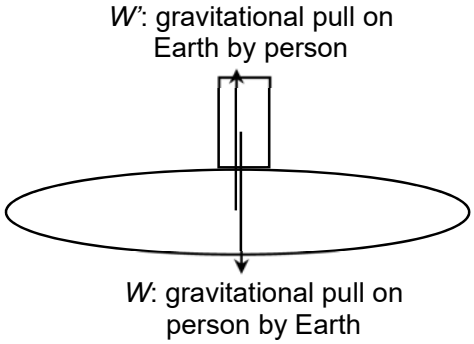
To illustrate, consider the forces when a person is standing on the floor.



Forces acting on the person



1st third-law pair



2nd third-law pair

Example 6 (Man in the lift)

An 80 kg man weighs himself by standing on a weighing scale inside a lift. What does the scale read if the lift
(a) is at rest,
(b) moves with an upward acceleration of 1.8 m s^{-2},
(c) the lift is moving upwards with a constant velocity of 2.2 m s^{-1},
(d) the lift slows from its velocity in (c) to rest at rate of 1.9 m s^{-2}.

Answer:

This YouTube video sheds more light on apparent weight and true weight:
https://www.youtube.com/watch?v=AbNJv1VNWu8

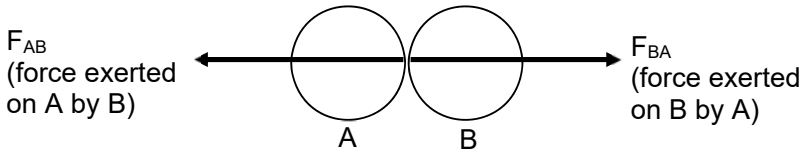
Example 7:

A hovering jetpack with nozzles of total cross-sectional area A expels water at constant velocity v relative to the jetpack. Show that the force exerted by the expelled water on the jetpack is $F = \rho A v^2$, where ρ is the density of the water jet.

Answer:

3.3 Principle of Conservation of Linear Momentum

Momentum is an important concept because, momentum is a **conserved** quantity. Consider, for example, the head-on collision of two billiard balls, A and B.



Assuming **no net *external* force act on the two bodies**, the only forces experienced by the two bodies A and B during the collision will be the mutual normal contact forces, F_{AB} and F_{BA}, respectively.

By Newton's 3rd law, the two forces must be equal but opposite:

$$\vec{F}_{AB} = -\vec{F}_{BA}$$

Since the entire interaction is simultaneous and over the same duration, the impulses of the two forces must also be equal but opposite:

$$\int \vec{F}_{AB}\ dt = \int -\vec{F}_{BA}\ dt$$
$$\vec{J}_{AB} = -\vec{J}_{BA}$$

It follows that the momentums exchanged between A and B must also be equal but opposite:

$$(\Delta\vec{p})_A = -(\Delta\vec{p})_B$$

If we see the two billiard balls as making up one system, then the **total momentum of the system does not change** in the collision:

$$\begin{aligned}\Sigma\Delta\vec{p} &= (\Delta\vec{p})_A + (\Delta\vec{p})_B \\ &= (\Delta\vec{p})_A + -(\Delta\vec{p})_A \\ \Sigma\Delta\vec{p} &= 0\end{aligned}$$

The above derivation can be extended to include any number of interacting objects in three-dimensional space, involving forces of any nature (magnetic, electric, gravitational, etc.), as long as the condition of **no resultant external force** holds. The logic is surprisingly straight-forward:

$$\Sigma\vec{F} = 0 \quad \Rightarrow \quad \Sigma\vec{J} = 0 \quad \Rightarrow \quad \Sigma\Delta\vec{p} = 0$$

Thus the general statement of the principle of conservation of momentum is

Statement of the Principle of Conservation of Linear Momentum

The **total linear momentum** of a system **remains unchanged** if ***no net external force*** acts on the system. Mathematically, $\Sigma m_n\vec{u}_n = \Sigma m_n\vec{v}_n$

The principle is **NOT restricted to collisions!**
Sadly, every year, many students get zero marks during examinations when asked to state the principle of conservation of linear momentum, because they write about momentum before and after collisions and stuff.

Notes:

- By a system, we mean a set of objects that we choose.
- Some textbooks refer to the system as being *isolated*. An isolated system is one in which the only (significant) forces are those between the objects in the system. The sum of all these "internal" forces within the system will be zero because of Newton's 3rd Law.
- If there is a net external force – then the total momentum of the system will *not* be conserved. E.g., if we take as our system a rock falling under gravity, the momentum of this system is *not* conserved because an *external* force, the gravitational pull by the Earth is acting on it. However, if we include the Earth in the system, the total momentum of rock plus Earth *is* conserved.

Example 8: (Freedman)

An open-topped freight car with mass 24 000 kg is coasting along without friction along a level track. It is raining very hard and the rain is falling vertically downward. Originally the car is empty and moving with a speed of 4.00 m/s. What is the speed of the car after it has collected 3000kg of rainwater?

Answer:

3.3.1 Two-body Systems

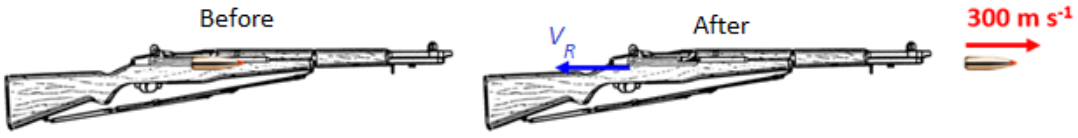
For a **two-body system**, the principle of conservation of momentum can be mathematically simplified as follows:

$$(\Delta\vec{p})_A = -(\Delta\vec{p})_B$$
$$(\vec{p}_f - \vec{p}_i)_A = -(\vec{p}_f - \vec{p}_i)_B$$
$$m_A(\vec{v}_A - \vec{u}_A) = m_B(\vec{u}_B - \vec{v}_B)$$
$$m_A\vec{u}_A + m_B\vec{u}_B = m_A\vec{v}_A + m_B\vec{v}_B$$

Example 9

A marksman holds a rifle of mass M = 3.00 kg loosely in his hands, so as to let it recoil freely when fired. He fires a bullet of mass m = 5.00 g horizontally with a velocity of v_B = 300 m s^{-1}.
What is the recoil velocity v_R of the rifle?

Answer:

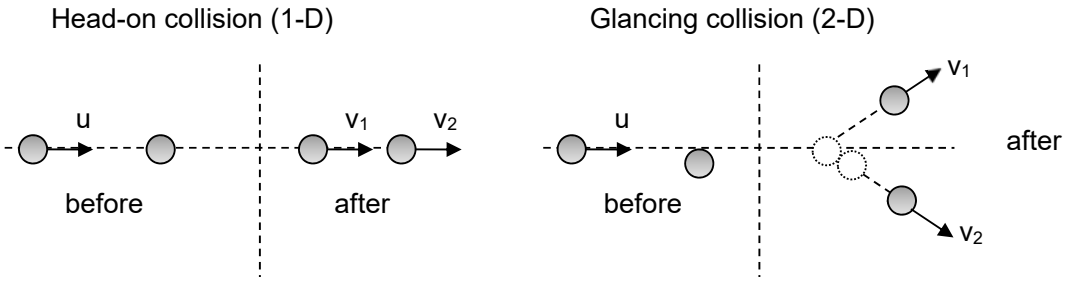


Note that since both the rifle and bullet are **stationary** before firing, the total initial momentum of the system is **zero**. Since momentum is a vector quantity, in order for the final total momentum to be zero, the **rifle must move in the direction opposite to the bullet** (recoil).

3.3.2 Types of Collisions

A **collision** is an isolated event in which two or more interacting bodies exert **relatively strong** forces on each other for a relatively short time. Note that the two bodies do not necessarily have to be in any physical contact, as interaction may involve a force at a distance.

3.3.2.1 Head-on vs Glancing Collisions

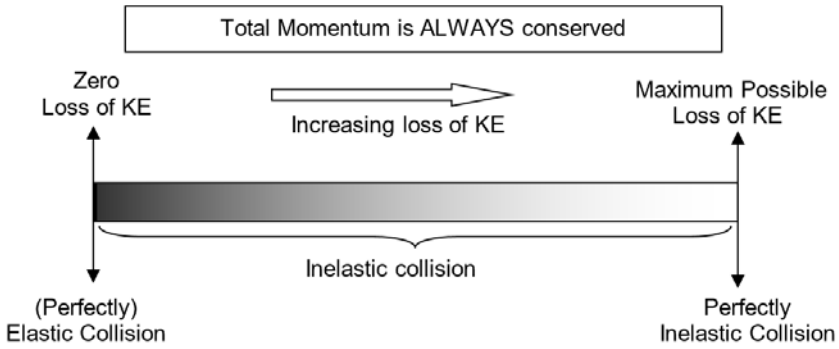


A **head-on** collision is one in which the directions of motion before and after the collision are along the same line of motion.

3.3.2.2 Elastic vs Inelastic Collisions

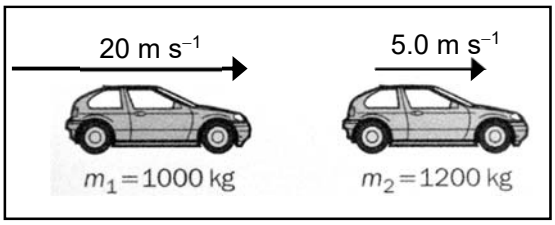
The total *kinetic* energy of the colliding bodies is *not* necessarily conserved during a collision. (This is unlike the total momentum, which is *always* conserved during a collision.) Some kinetic energy is usually converted into vibrational energy of the atoms, causing a heating effect, sound, and deformation energy. How much of the original kinetic energy is conserved after the collision depends on the **elasticity** of the collision.

- An **elastic** collision is one where 100% of the kinetic energy is conserved, e.g. two diamonds bouncing off each other. Note that *during* an elastic collision, some of the kinetic energy can temporarily be converted to potential energy.
- An **inelastic** collision is one that is *not* elastic, i.e. some or all of the kinetic energy is converted into vibrational energy, sound, deformation energy, etc.
- A **perfectly inelastic** collision is one where the 2 objects coalesce (stick together) and move with **common velocity** after the collision, e.g. two lumps of clay that don't bounce at all, but stick together after the collision. This represents the **maximum possible loss** of kinetic energy without violating the principle of conservation of energy; it does *not* necessarily mean all the kinetic energy is lost.

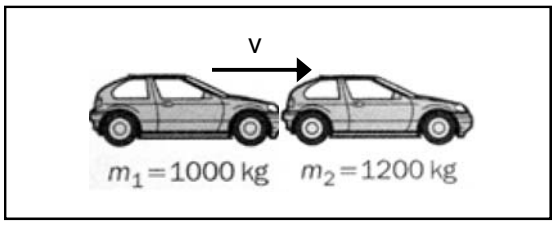


Note that **momentum** is **always** conserved, regardless of whether a collision is elastic or inelastic. In most exam questions, you will have to apply the principle of conservation of linear momentum.

Example 10 (Perfectly Inelastic collision)
A car of mass 1000 kg moving at 20.0 m s^{-1} collides with a car of mass 1200 kg, moving at 5.0 m s^{-1} in the same direction.
(a) If the collision is perfectly inelastic, determine their final velocity after the collision.
(b) Calculate also the ratio of (total final kinetic energy) to (total initial kinetic energy).



before

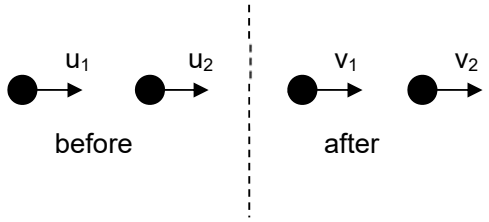


after

Answer:

3.3.2.3 Relative Speed of Approach/Separation during Elastic Collisions

Consider a two-body **elastic** head-on collision.



By the principle of conservation of momentum,

$$m_1u_1 + m_2u_2 = m_1v_1 + m_2v_2$$

Rearranging, we get

$$m_1(u_1 - v_1) = m_2(v_2 - u_2) \text{ ----------(1)}$$

Since the kinetic energy is conserved during an elastic collision,

$$\frac{1}{2}m_1u_1^{\,2} + \frac{1}{2}m_2u_2^{\,2} = \frac{1}{2}m_1v_1^{\,2} + \frac{1}{2}m_2v_2^{\,2}$$

Rearranging, we get

$$m_1(u_1^{\,2} - v_1^{\,2}) = m_2(v_2^{\,2} - u_2^{\,2})$$

$$\Rightarrow \quad m_1(u_1 - v_1)(u_1 + v_1) = m_2(v_2 - u_2)(v_2 + u_2) \text{ ----------(2)}$$

Dividing question (2) by equation (1),

$$(u_1 + v_1) = (v_2 + u_2)$$

Rearranging, we get

$$(u_1 - u_2) = (v_2 - v_1)$$

The $(u_1 - u_2)$ term represents the relative speed of approach before the collision.

The $(v_2 - v_1)$ term represents the relative speed of separation after the collision.

We thus come to the following conclusion:

Statement of Relative Speed of Approach/Separation

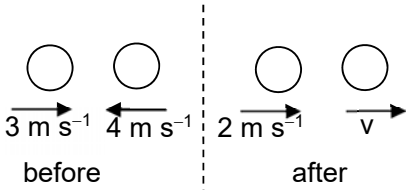
For **elastic** collisions, the relative speed of approach is equal to the relative speed of separation.

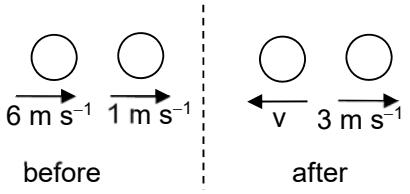
Mathematically, $u_1 - u_2 = v_2 - v_1$

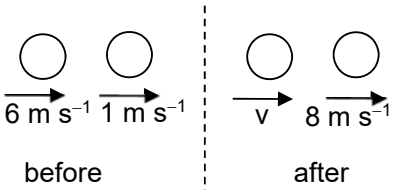
This result can only be applied for a 2-body **elastic** collision. During examinations, the above statement must be **made** before it can be applied mathematically in solving problems.

Example 11

Assuming the following are **head-on elastic collisions**, solve for the unknown velocity *v* in each case. Note that the masses of the objects are not (necessarily) equal.







TIPS **Solving 2-body Collision Problems**

Step 1 Indicate velocity vectors for each of the bodies before and after the collision. Choose a convenient positive direction, usually to the right. You can assume the unknown velocities are all positive. If the solutions turn out to be negative later, it means these velocities are in the negative direction.

Step 2 Set up the equation for the principle of conservation of linear momentum. The positive direction chosen in step 1 helps us to define the signs of the (known) velocities in the equations: velocities in the positive direction get a "+" and those in the negative direction get a "–".

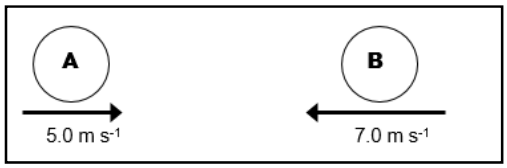
Step 3 **IF** the collision is perfectly inelastic, use one variable to represent the final velocities of both bodies (i.e. make $v_1 = v_2 = v$).

IF the collision is elastic, you will probably have two unknown velocities and need another equation. (You need as many independent equations as there are unknowns to solve them uniquely.) Use the equation for relative speed of approach/separation (preferred) of that for the conservation of kinetic energy (which is more tedious, as it is a quadratic equation).

Step 4 Solve for the unknowns, e.g. mass, initial or final velocities.

Example 12 (Elastic collision)
Two spheres, A and B, of mass 4.0 kg and 6.0 kg, respectively, collide head-on. Their initial velocities are 5.0 m s^{-1} and 7.0 m s^{-1} in opposite directions, as shown in the diagram.
Assuming the collision is **elastic**, determine their final velocities.

Answer:
The directions of motion of the spheres after the collision are not given in the question. Hence, we will need to draw the "after collision" diagram to indicate the directions used in the analysis. It is ok if the directions chosen are incorrect as the final answer will tell us if the sphere are really moving in the predicted direction in the "after collision" diagram.

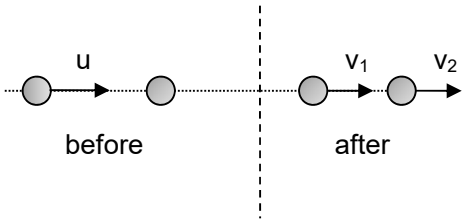


Before

After

3.3.2.4 Elastic Collision, One Body Initially at Rest *(Good to know)*

Let's study a head-on elastic collision between two bodies, one of which is initially at rest.



By the principle of conservation of linear motion:

$$m_1u = m_1v_1 + m_2v_2 \quad \text{---------- (1)}$$

For an elastic collision, relative speed of approach equals relative speed of separation:

$$u - 0 = v_2 - v_1$$

Substitute $v_2 = u + v_1$ into (1):

$$m_1u = m_1v_1 + m_2u + m_2v_1$$
$$v_1 = \frac{m_1 - m_2}{m_1 + m_2}u$$

Substitute $v_1 = v_2 - u$ into (1):

$$m_1u = m_1v_2 - m_1u + m_2v_2$$
$$v_2 = \frac{2m_1}{m_1 + m_2}u$$

Case 1: $m_1 = m_2$

$v_1 = 0$, $v_2 = u$

E.g. during a billiard shot, the cue ball comes to a complete rest after hitting the target ball (of the same mass), passing all the momentum to the target ball.

Case 2: $m_1 << m_2$

$v_1 \approx -u$, $v_2 \approx 0$

E.g. a ping-pong ball lands on the table and bounces back with the same speed as before.

Case 3: $m_1 >> m_2$

$v_1 \approx u$, $v_2 \approx 2u$

E.g. After hitting a bowling pin, the bowling ball continues with no noticeable drop in speed, while the pin is sent flying forward at about twice the ball's speed.

Obviously, these results are only valid for head-on elastic collisions between two bodies, one initially at rest. During examinations, you are *not* allowed to quote and use these results without derivation. However, they may come handy as "sanity checks" and in multiple-choice questions.

Appendix A: Static and Kinetic Friction

Static Friction

The magnitude of *static* friction between two surfaces can be expressed as

$$f_s \leq \mu_s N$$

where μ_s is the coefficient of static friction between the two surfaces in question and
N is the magnitude of the normal contact force between the two surfaces.

Static friction thus has a maximum magnitude, called the limiting friction, $f_{max} = \mu_s N$.

Kinetic Friction (or sliding friction or dynamic friction)

The magnitude of *kinetic* friction between two surfaces can be expressed as

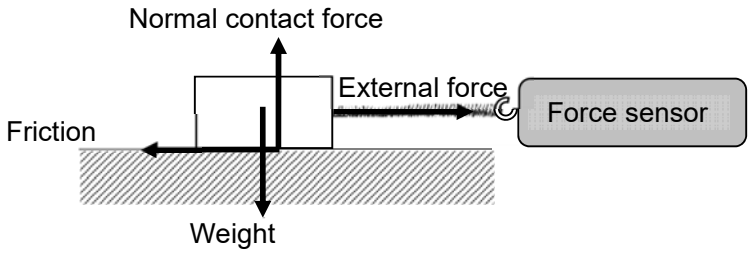
$$f_k = \mu_k N$$

where μ_k is the coefficient of kinetic friction between the two surfaces in question and
N is the magnitude of the normal contact force between the two surfaces.

Unlike the static friction, kinetic friction is a constant value. It is also independent of the relative speed of motion.

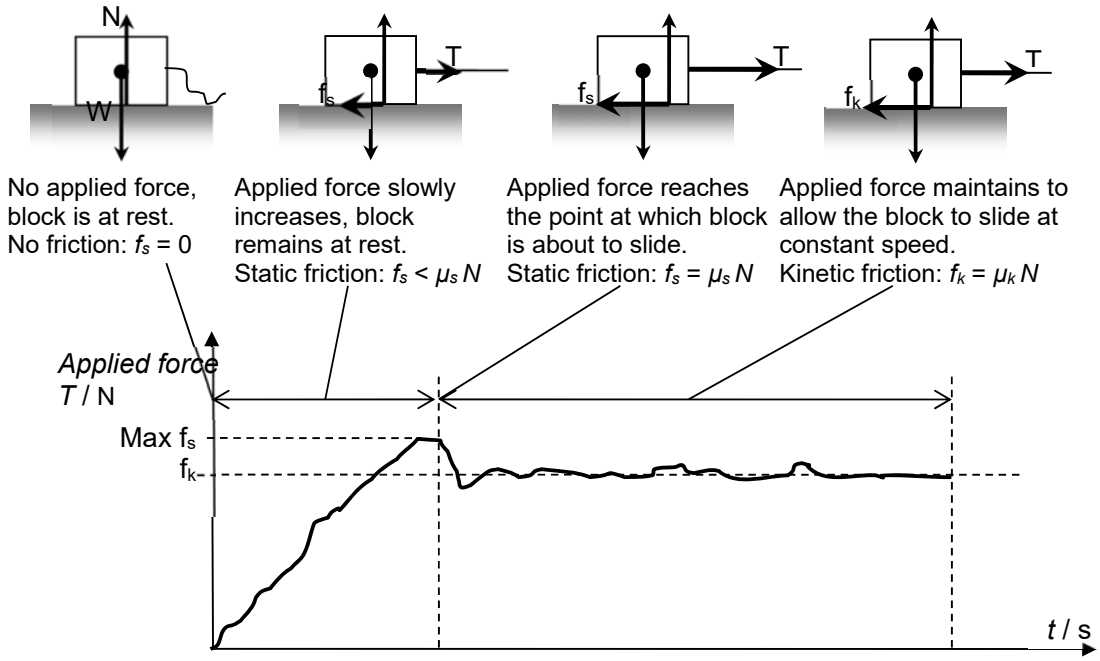
Both μ_s and μ_k are empirically found to be always less than 1 (friction is always less than the normal contact force). μ_k is found to be always less than μ_s (kinetic friction is always smaller than the limiting static friction).

To better understand the two, consider the results of the following experiment.



An external force is applied on the block and is increased very gradually until the block starts moving and continues to move at a constant speed. Since the block is always either stationary or moving at constant speed, the magnitude of the frictional force can be inferred from the applied force as measured by the force sensor.

The graph of the applied force against time is shown below in relation to the magnitudes of friction, both static and kinetic:



Appendix B: Newton's Law in Car Safety

Using Newton's laws, discuss how each of the following features in vehicles increases the safety of its passengers:

a) seat belts,
b) head restraints,
c) air bags.

(a) Seat belt exerts a retardation force on the passenger, without which he would have continued in his state of uniform motion in a straight line into the windscreen. (Newton's 1st Law).

If a moving car crashes into another (stationary) car, it comes to an abrupt stop. If the passenger continues to move forward, he will crash into the windscreen and likely suffer from severe injuries, or even death. Without the seat belt, the only retardation force on the passenger is the frictional force between the seat and the passenger's bum. But this force is limited. If the vehicle decelerates too fast, the frictional force is not large enough to decelerate the passenger at the same rate as the vehicle.

When passengers crash into the wind screen, it is due to their inertia: they are already moving forward and continue to do so. It is not due to some mysterious "forward force" that pushes them into the wind screen

(b) If a car is hit from the rear, the head restraint provides a forward force on the passenger's head so that the head is accelerated at the same rate as the rest of his body. If the passenger's body accelerates forward much faster than his head, the passenger may suffer from whiplash (a kind of neck injury). The muscles in the passenger's neck do not provide enough force to accelerate his head as fast as the rest of his body. (Newton's 2nd Law)

(c) A collision with the car's wind screen would bring the passenger to rest in a very short period of time. Airbags exert a relatively small force over a relatively long duration of time, safely bringing the passenger to rest by reducing the passenger's momentum at a relatively slower rate. (Newton's 2nd Law)

Tutorial 3: Dynamics

Self-Review Questions

Part 1: Newton's Laws, Inertia, Force, Momentum, Impulse

S1 When a force of 4 N acts on a mass of 2 kg for a time of 2 s, what is the rate of change of momentum?

A 1 kg m s^{-2} **B** 2 kg m s^{-2} **C** 4 kg m s^{-2} **D** 8 kg m s^{-2} **E** 16 kg m s^{-2}

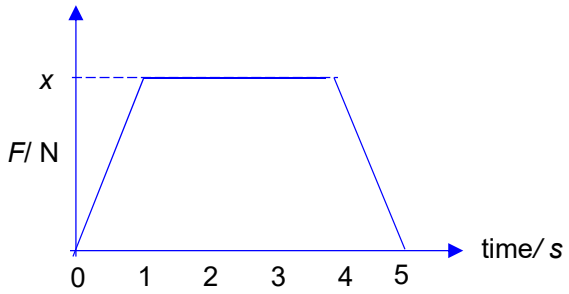
S2 On an aircraft carrier, an F-18 Super Hornet fighter aircraft that is being launched can be catapulted from 0 to 270 km h^{-1} in 2.0 s.

For a jet having a mass of 20 000 kg, what is the average force experienced by it?

S3 **[N83/P2/2]**

When a force, *F*, varying as shown below is applied to a mass of 10 kg, the gain in momentum in 5 s is 40 kg m s^{-1}.

What is the value of *x*?



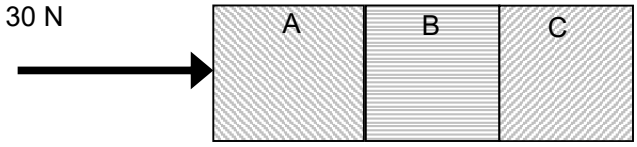
A 4 **B** 8 **C** 10 **D** 15

S4 **[N84/P3/Q3]**

A man is parachuting at constant speed towards the surface of the Earth. The force which, according to Newton's 3rd law, makes an action-reaction pair with the gravitational force on the man is

A The tension in the harness of the parachute.
B The viscous force of the man and his parachute on the air.
C The gravitational force on the Earth due to the man.
D The viscous force of the air on the man and his parachute.
E The tension in the fabric of the parachute.

S5 Three blocks A, B and C, of masses 2.0 kg, 4.0 kg and 3.0 kg, respectively, are pushed along a smooth horizontal surface by a force of 30 N as shown in the diagram.

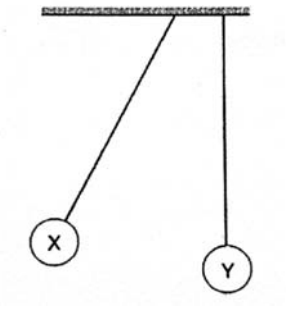


a) Calculate the net force experienced by each of the blocks, A, B and C.
b) What is the force exerted by block B on (i) block C? (ii) block A?

Part 2: Conservation of Linear Momentum/Collisions

S6 **[N02/P1/Q5]**

Two steel balls X and Y are suspended on a string. Ball X is pulled to one side as shown. After ball X is released, the balls collide.

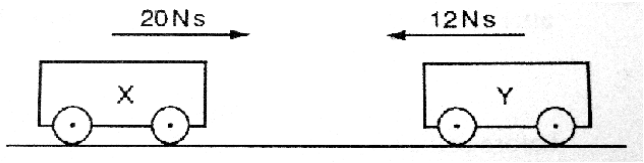


Which quantities must be conserved in the collision?

A Kinetic energy, total energy and momentum
B Kinetic energy and momentum only
C Kinetic energy and total energy only
D Total energy and momentum only

S7 **[N98/P1/Q4]**

The diagram shows two trolleys, X and Y, about to collide and gives the momentum of each trolley before the collision. After the collision, the directions of the motion of both trolleys are reversed and the magnitude of the momentum of X is then 2 N s.

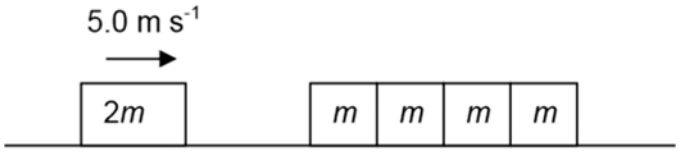


What is the magnitude of the corresponding momentum of Y?

A 6 N s **B** 8 N s **C** 10 N s **D** 30 N s

S8 **[Promo2007/I/6]**

Four identical railway trucks, each of mass m, are coupled together and are at rest on a smooth horizontal track. Initially, a fifth truck of mass $2m$ is moving at 5.0 m s^{-1}. It collides and moves off together with all the stationary trucks.



Calculate the speed of the trucks after impact.

A 0.83 m s^{-1} **C** 1.3 m s^{-1}

B 1.0 m s^{-1} **D** 1.7 m s^{-1}

S9 During a rescue operation, a 5500 kg helicopter hovers above a fixed point. The rotating blades of the helicopter's rotor send air downwards with a speed of 60.0 m s^{-1} relative to the helicopter, producing enough upward force for helicopter to hover.

a) What is the net force acting on the helicopter?
b) What is the upward force produced by the helicopter's rotor?
c) What mass of air must pass through the blades every second to produce enough thrust for the helicopter to hover?

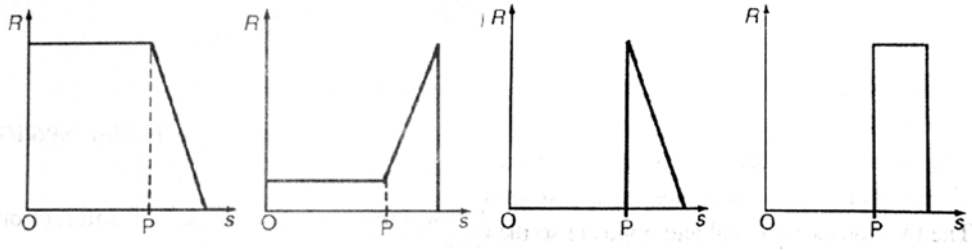
Discussion Questions

Part 1: Newton's Laws, Inertia, Force, Momentum, Impulse

D1 **J88/P1/Q5**

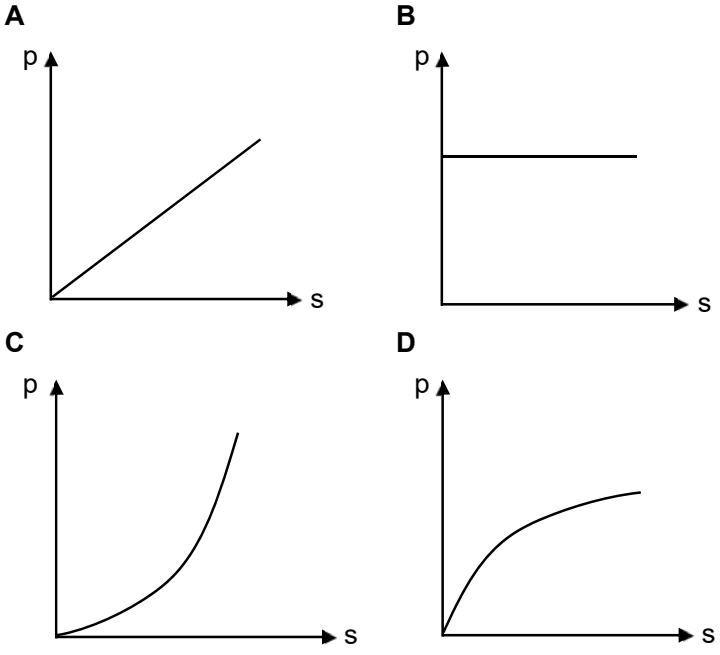
An object falls vertically through air at a constant velocity and then strikes soft ground in which it becomes embedded. Its deceleration during impact is constant. If P represents the point of impact, which of the following graph best represents the variation of the total force R on the object with distance s?

A **B** **C** **D**



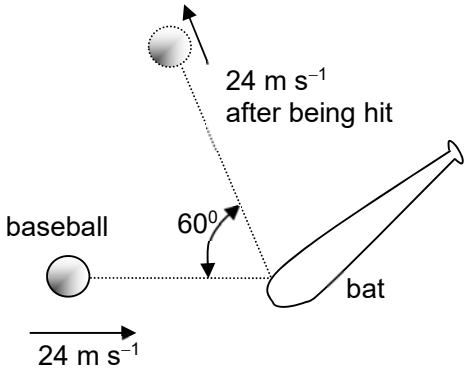
D2 **TJC/2008/P1/Q7**

A railway carriage starts from rest and is driven along a straight horizontal track by a motor which exerts a constant force. The effects of friction and air resistance can be neglected.
Which of the graphs below best represents the variation of the momentum p with distance travelled s?

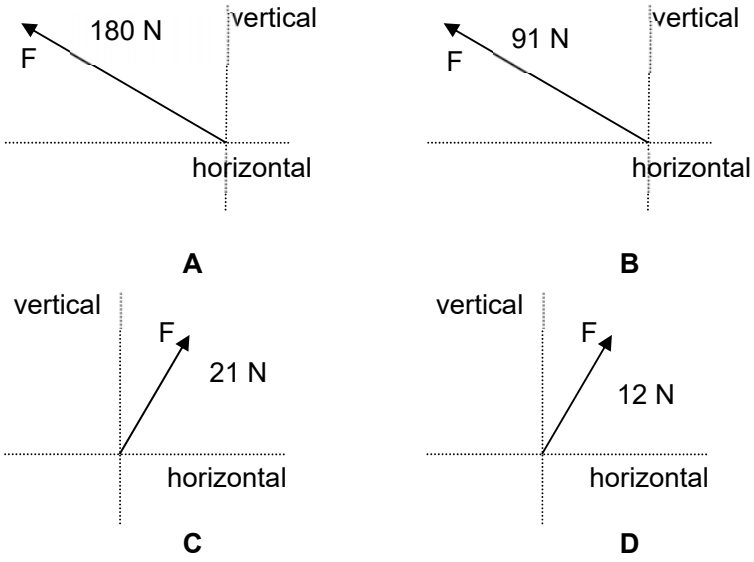


D3 **HCI/2007/P1/Q8**

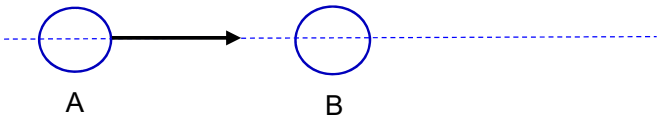
A baseball of mass 0.11 kg is thrown with a horizontal velocity of 24 m s^{-1} towards a batter. After the ball is struck by the bat, it has a speed of 24 m s^{-1} in the direction 60° above the horizontal as shown on the right.



If the ball and bat are in contact for 0.025 s, the magnitude and direction of the average force exerted on the ball by the bat is



D4 Two identical billiard balls A and B collides, with B being at rest before the collision.



(top view of the two billiard balls)

Explain why in order for the initial speed of ball B after the collision to be in the same direction as the initial direction of travel of A, it must be a head-on collision.

D5 **[2017 P1 Q4]**

An object of mass 3.0 kg is falling vertically from rest. Air resistance *R*, in newtons, is given by the empirical equation

$$R = 0.60v$$

where *v* is the velocity in metres per second.

What is the maximum velocity of the object and what is the acceleration of the object when its velocity is 12 m s^{-1}.

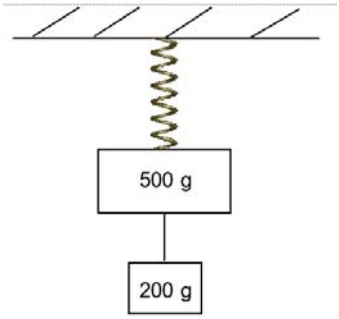
D6 **[J80/II/2&3, adapted]**

A car of mass 1000 kg pulls a caravan of mass 1000 kg. The total resistance to motion has a constant value of 4000 N. One quarter of this resistance acts on the caravan. At first, the acceleration of the car and the caravan is 2.0 m s^{-2}, but eventually they move at a constant speed of 6.0 m s^{-1}.

a) Calculate the force exerted on the car by the tow bar when the acceleration is 2.0 m s^{-2}.
b) Calculate the force exerted on the car by the tow bar when the car and caravan are moving at a constant speed of 6.0 m s^{-1}.

D7 **HCI/2007C2-BT2/P1/Q4**

The diagram below shows a system in static equilibrium in which two masses, joined by a light thread, are suspended from a light vertical spring. If the thread is cut, what is the value of the acceleration of the 500 g mass immediately after the thread is cut?

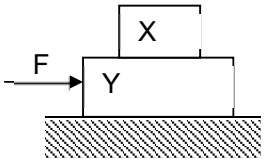


A *g* **B** 0.4*g* **C** 0.2*g* **D** zero

D8 **HCI/2008Promo/P1/Q5**

Two blocks, X and Y, of masses *m* and 3*m*, respectively, are accelerated along a smooth horizontal surface by a force *F* applied to block Y as shown.

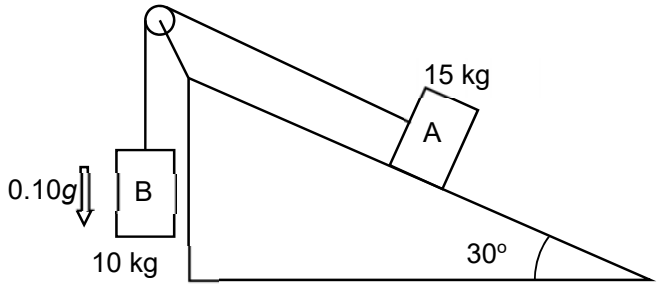
The interface between block X and Y is rough and friction is present. If block X does not slide on block Y, what is the frictional force acting on block X?



D9 **HCI/2007BT2/P1/Q4**

Two blocks, A and B, of mass 15 kg and 10 kg respectively, are connected by a light cord passing over a light, free-running pulley as shown in the diagram below. When released, block B accelerates downwards at 0.10 *g*.

What is the frictional force experienced by block A?

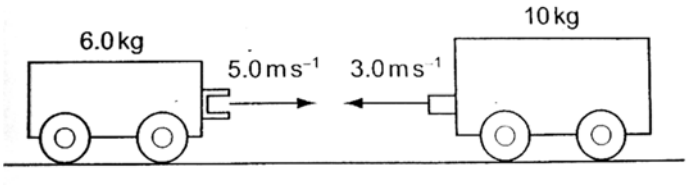


A 0 N **B** 15 N **C** 50 N **D** 90 N

Part 2: Conservation of Linear Momentum/Collisions

D10 **H2 2008/P1/Q6**

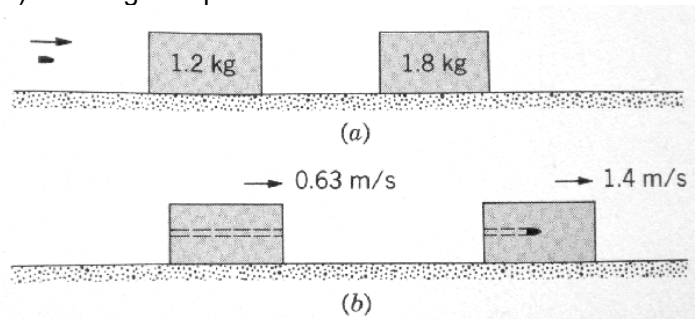
A trolley of mass 6.0 kg travelling at a speed of 5.0 m s^{-1} collides head-on and locks together with another trolley of mass 10 kg which is travelling in the opposite direction at a speed of 3.0 m s^{-1}. The collision lasts for 0.20 s.



What is the total momentum of the two trolleys before the collision and the average force acting on each trolley during this collision?

	Total momentum before the collision/ kg m s^{-1}	Average force on each trolley / N
A	0	300
B	60	150
C	0	150
D	60	300

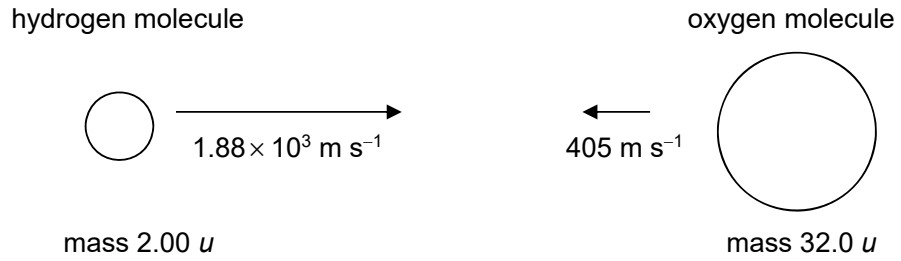
D11 A 3.5 g bullet is fired horizontally at two blocks resting on a smooth table top, as shown in the figure. The bullet passes through the first block, with mass 1.2 kg, and embeds itself in the second, with mass 1.8 kg.
The final speeds of the blocks are 0.63 m s^{-1} and 1.4 m s^{-1} respectively.
Neglecting the mass removed from the first block by the bullet, find
a) the speed of the bullet immediately after emerging from the first block and
b) the original speed of the bullet.



D12 N2000/III/1(part)

a) (i) Define linear momentum. [1]
(ii) State the principle of conservation of momentum. [2]

b) In a gas a hydrogen molecule, mass 2.00 *u* and velocity 1.88×10^3 m s^{-1}, collides elastically and head-on with an oxygen molecule, mass 32.0 *u* and velocity 405 m s^{-1}, as illustrated in the figure below.



In qualitative terms, what can be stated about the subsequent motion as a result of knowing that
(i) the collision is elastic,
(ii) the collision is head-on.
[3]

c) Using your answers to b),
(i) determine the velocity of separation of the two molecules after the collision.

Since collision is elastic,
(ii) apply the law of conservation of momentum to the collision.
(iii) determine the velocity of both molecules after the collision. [6]

D13 N2007/P3/Q1 (modified)

A tritium nucleus moves towards a deuterium nucleus as illustrated in Fig. 1.1.

Fig. 1.1

The nuclei initially have the same speed *v*. The tritium nucleus consists of two neutrons and a proton. The deuterium nucleus consists of a neutron and a proton. The proton and the neutron have the same mass *m*.

The two nuclei having like charges, repel one another.

a) Explain why it is **not** possible for the nuclei to stop at the same instant. [2]

b) At one instant during the interaction between the nuclei, they are both travelling **in the same direction** with the same speed. Calculate this speed, in terms of *v*. [2]

c) Fig. 1.2 is a velocity-time sketch graph showing how the velocity of each nucleus varies. The interaction between the nuclei is elastic.

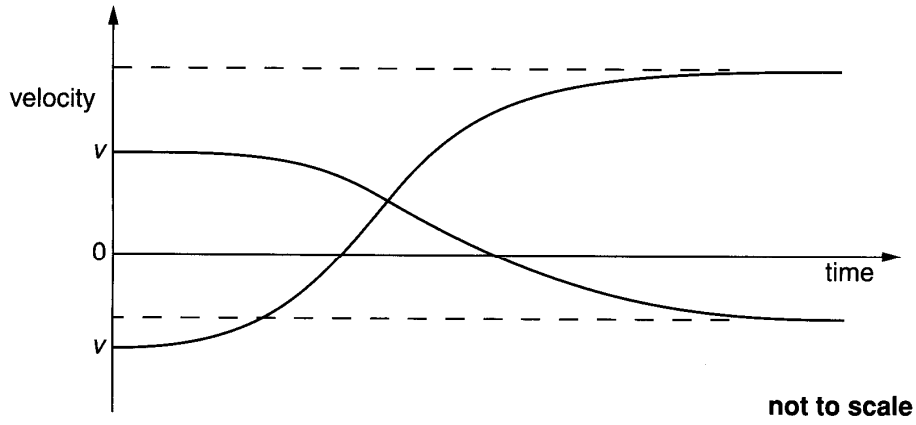


Fig. 1.2

(i) Label the graph to show
1. which curve is for the tritium nucleus,
2. the times at which each nucleus stops,
3. the time at which they are at their distance of closest approach. [3]

(ii) Determine the final speed of each nucleus in terms of *v*. [4]

D14 HCI BT2010/II/3(b)

The rocket engines of the space shuttle can exert a force of 28.6 MN at lift-off when the total mass is 2.00 x 10^6 kg.

(i) Determine the acceleration of the rocket on vertical lift off. [2]

(ii) Suppose a 65.0 kg astronaut in the space shuttle were to stand on a bathroom weighing scale that indicates "weight" in kilograms during lift-off.
Determine the reading on the scale. [2]

(iii) The rocket engine ejects gas at a rate of 2.00 x 10^4 kg s^{-1} on lift-off. Determine the speed of the ejected gas relative to the rocket engine. [2]

D15 H2 2011/P3/Q1

A toy car with a rocket engine moves along a horizontal track, as shown in Fig. 1.1.

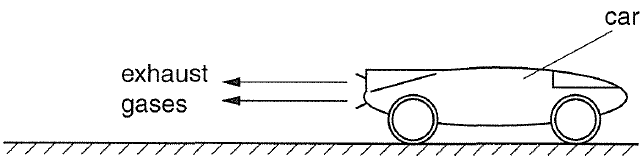


Fig 1.1

The rocket engine produces a constant forward force of 4.6 N. The car loses mass continuously as exhaust gases are produced by the rocket.

a) Use momentum consideration to explain why the rocket produces a forward force on the car.[3]

b) The variation with time *t* of the speed *v* of the car is shown in Fig. 1.2.

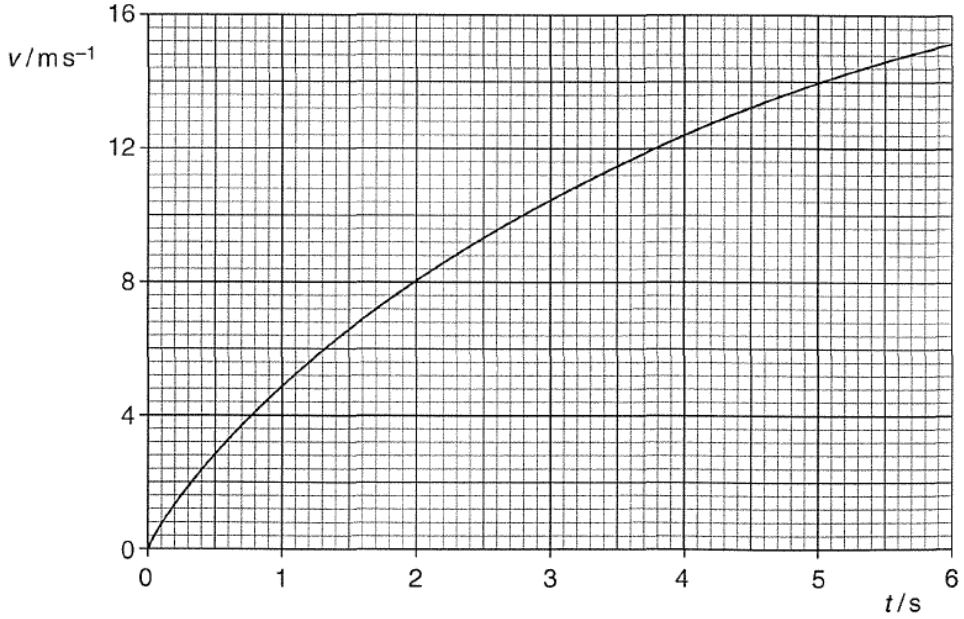


Fig. 1.2

At time *t* = 2.0 s, the mass of the car is 440 g.

(i) For the time *t* = 2.0 s,

1. use Fig. 1.2 to determine the acceleration of the car, [2]
2. use your answer in (i) part 1 to determine the magnitude of the resistive force acting on the car. [2]

(ii) Explain how it can be deduced that the resistive forces acting on the car increase with increase of speed. [2]

c) The toy car is now re-fuelled and then rotated so that it is pointing upwards. It is suggested that the rocket engine produces sufficient force to propel the car vertically. By considering the acceleration of the car at time *t* = 0 in Fig. 1.2, comment on this suggestion. [3]

Answers to Tutorial 3: Dynamics

Self-Review
S1 C **S2** 750 kN
S3 C **S4** C
S5 a) 6.7 N, 13 N, 10 N b) (i) 10 N (ii) 23.3 N
S6 D **S7** C **S8** D
S9 0N, 54N, 900 kg/s

Discussion
D5 (i) 4.49 m s^{-2} (ii) 94.8 kg (iii) 1.43 x 10^3 m s^{-1}
D6 49 m s^{-1}; 7.4 m s^{-2}
D7 a) 3000 N b) 1000 N
D8 0.4g
D9 *F*/4
D10 0
D11 b) (i) 2.7 m s^{-2}, 3.4 N
D13 a) 721 m s^{-1} b) 937 m s^{-1}
D14 c) (i) 2285 m s^{-1} (iii) 136 m s^{-1}, 2420 m s^{-1} both towards the left.
D15 b) 0.20*v* c) (ii) $v_{deuterium} = 1.4v$, $v_{tritium} = 0.60v$

Chapter 4

Forces

"The same principle that allows an airplane to rise off the ground by creating lift under its wings is used in reverse in F1 cars to generate an additional 'downward force' to press the race car against the surface of the track. This increases the contact force between the tires and the road surface, allowing the car to turn corners at amazing speeds. F1 cars achieve downward force-to-weight ratio of 1:1 at about 125 km/h. At 190 km/h (118 mph) the ratio is roughly 2:1."

- Anonymous Formula One Expert

TOPIC 4: Forces

H2 Physics Syllabus 9749

Forces	**Learning Outcomes** Students should be able to:
Types of force	(a) recall and apply Hooke's law ($F = kx$, where k is the force constant) to new situations or to solve related problems (b) describe the forces on a mass, charge and current-carrying conductor in gravitational, electric and magnetic fields, as appropriate *(covered later in relevant chapters)* (c) show a qualitative understanding of normal contact forces, frictional forces and viscous forces including air resistance (no treatment of the coefficients of friction and viscosity is required) *(partly covered in Kinematics)*
Centre of gravity	(d) show an understanding that the weight of a body may be taken as acting at a single point known as its centre of gravity *(covered in Dynamics)*
Turning effects of forces	(e) define and apply the moment of a force and the torque of a couple (f) show an understanding that a couple is a pair of forces which tends to produce rotation only (g) apply the principle of moments to new situations or to solve related problems
Equilibrium of forces	(h) show an understanding that, when there is no resultant force and no resultant torque, a system is in equilibrium (i) use a vector triangle to represent forces in equilibrium
Upthrust	(j) derive, from the definitions of pressure and density, the equation $p = \rho gh$ (k) solve problems using the equation $p = \rho gh$ (l) show an understanding of the origin of the force of upthrust acting on a body in a fluid (m) state that upthrust is equal in magnitude and opposite in direction to the weight of the fluid displaced by a submerged or floating object (n) calculate the upthrust in terms of the weight of the displaced fluid (o) recall and apply the principle that, for an object floating in equilibrium, the upthrust is equal in magnitude and opposite in direction to the weight of the object to new situations or to solve related problems.

Contents

Topic 4: Forces

4.1 Hooke's Law 4
4.2 Upthrust or Buoyant Force 4
4.2.1 Pressure due to Fluid 4
4.2.2 Derivation of the equation $p = \rho gh$ 5
4.2.3 Upthrust 6
4.3 Translational Equilibrium 8
4.4 Moment of a Force 9
4.5 Rotational Equilibrium and Principle of Moments 10
4.6 Static Equilibrium for Rigid Extended Bodies 11
4.7 Three-Force Systems 12
Appendix I Contact force and frictional force 15
Appendix II viscous force 16
Appendix III Buoyancy of a submarine 17
Tutorial 4 Forces 18
Self-Review Questions 18
Discussion Questions 20

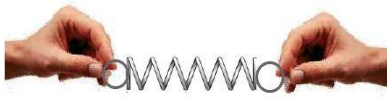
Videos of Lecture Examples can be found at

https://youtube.com/playlist?list=PL_b5cjrUKDlaffEZq6U4qTf1LEZbyIgF-

Let us begin by introducing some forces that we will encounter in this chapter.

4.1 Hooke's Law

When we try to extend a spring by pulling it apart or compressing it with both hands, each of our hands will be subjected to an opposing force by the spring. We refer to this force as the tension or compression in the spring. For our purposes, we usually consider light springs of negligible mass. In 1676, Robert Hooke stated an empirical law that allows us to calculate the magnitude of this force.



Hooke's Law states that the **magnitude of force *F* exerted by a spring[#] on a body** attached to the spring is proportional to the **extension** *x* of the spring from its natural length provided the proportional limit of the spring is not exceeded.

$$F = kx$$

k = constant of proportionality (also referred to as force constant or spring constant or stiffness of spring)

Example 1 (N94/I/23)
A spring, obeying Hooke's Law, has an unstretched length of 50 mm and a spring constant of 400 N m^{-1}. What is the tension in the spring when its overall length is 70 mm?

[Answer: 8.0 N]

[#] Note that, in general, all elastic materials obey Hooke's Law within their elastic limits. The law is not restricted to springs.

4.2 Upthrust or Buoyant Force

4.2.1 Pressure due to Fluid

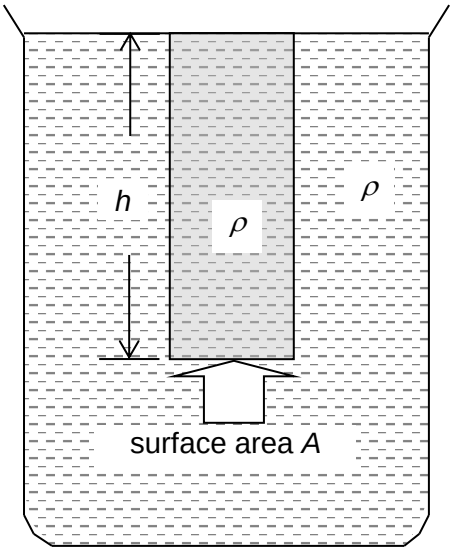
If an object is immersed in a fluid, the fluid will press on the surface of the object. The normal force per unit area of the surface is referred to as the pressure due to the fluid. For a cylinder that is filled with water, the water pressure on the base of the cylinder will increase if more water is added and the water level rises. Similarly, if a small sheet of metal is dropped into the cylinder, the pressure on the surface of the metal sheet will increase as it sinks deeper.

Apart from water, we also experience atmospheric pressure due to the air above us. The atmospheric pressure we experience as we climb up a mountain will decrease as we ascend higher.

4.2.2 Derivation of the equation $p = \rho gh$

Consider a column of fluid as shown in the diagram. This column of fluid is pressed inwards on all sides by the rest of the fluid in the beaker. Let

ρ (Greek letter *rho*) = density of the fluid,
A = surface area of the bottom of the column,
h = depth from the surface of the fluid,
V = volume of the fluid column,
m = mass of the fluid column.



The weight of the fluid is pressing down on the bottom surface A. Since *pressure is force over area*, the pressure due to the weight of the fluid column is

$$p = \frac{mg}{A}.$$

Recall that *density is mass divided by volume*, $\rho = m / V$. Rearranging, we find that $m = \rho V$, from which we get

$$p = \frac{\rho V g}{A}.$$

Note further that the volume of the fluid column is equal to the height of the column multiplied by the surface area, $V = h A$. Thus, we find that $V / A = h$, from which

$$p = \rho g h.$$

Hence, the average pressure due to the weight of a fluid column is $p = \rho gh$.

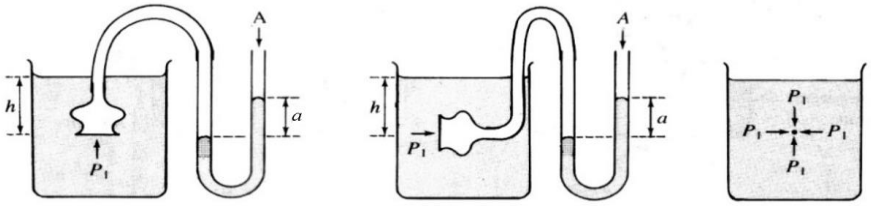
Note that we also have atmospheric pressure on top of the fluid column in the derivation above, pressing the fluid column down as it were. The total pressure at the surface area A is therefore equal to the atmospheric pressure added to the pressure due to the fluid:

$$p_{total} = p + p_{atm},$$

where p_{atm} is the atmospheric pressure at the surface of the fluid and p is the pressure at the bottom surface of the column due to the weight of the fluid alone.

Note that this expression for the pressure on the surface of the column is independent of the cross-sectional area of the column. Hence, if we consider a column where the cross-sectional area of the column approaches zero, we then obtain the same formula for the *pressure at a point* in the fluid.

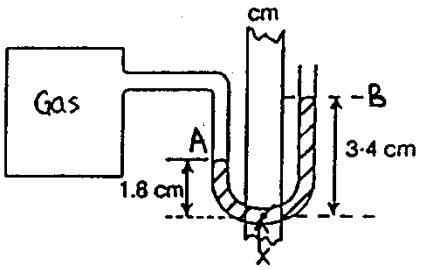
By considering an object that is very small, of negligible volume, we can say that the *pressure at a point in a fluid is the same in all directions*. A simple experiment can be conducted to illustrate this using a manometer. Since the height difference a remains the same as we change the direction that the opening of the tube is facing in the fluid, we can see that the pressure at the opening of the tube in the fluid is the same in all directions.



Example 2

The diagram on the right shows a mercury manometer connected to a gas container. The atmospheric pressure is 760 mm Hg and the point X is at the bottom of the U-tube.

a) What is the pressure of the point X in mm Hg?
b) What is the pressure of the gas in mm Hg?
c) If the manometer is disconnected from the gas container, what will be the vertical heights of the mercury levels from the point X in the left and right limbs of the manometer?



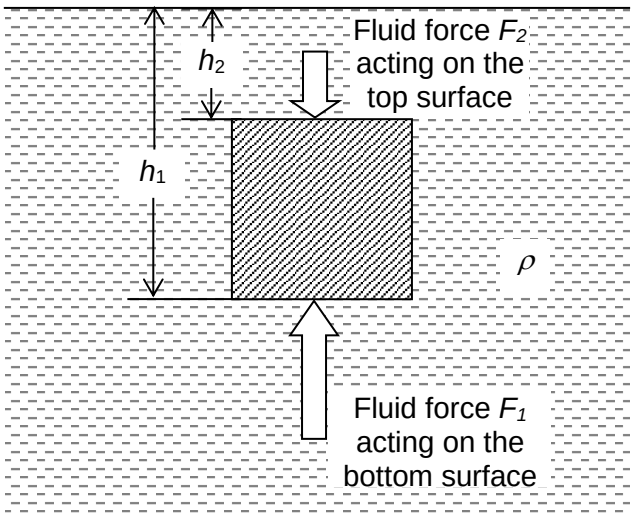
[Answers: 794 mm Hg; 776 mm Hg; 26 mm]

4.2.3 Upthrust

For an object that is immersed in a fluid, the force that the fluid exerts on different parts of its surface differs according to the depth of immersion of that part of the surface. We will now consider the net force that acts on the object due to the fluid.

Upthrust U refers to the **net upward** force exerted **by a fluid** on a body **fully or partially** submerged in the fluid.

By considering the fluid forces acting on a cube fully immersed in a fluid as shown in the diagram, we can generalise the result to find the upthrust experienced by any body immersed in a fluid.



Let ρ = density of the fluid,
A = surface area of a face of the cube,
$\Delta h = h_1 - h_2$ = length of a side of the cube.

The fluid forces acting on the side faces of the cube are equal in magnitude and in opposite directions, thus resulting in **zero net horizontal force**.

The fluid force F_1 acting on the **bottom** surface is $F_1 = \rho g h_1 A$.
The fluid force F_2 acting on the **top** surface is $F_2 = \rho g h_2 A$.
Thus, the **net vertical force** is $F_1 - F_2 = \rho g h_1 A - \rho g h_2 A = (h_1 - h_2)\rho g A = \Delta h \rho g A$
$= \rho V g$ (since the volume of the cube is $V = A\,\Delta h$)
= (mass of fluid displaced) g
= weight of fluid displaced

Reference: Why is upthrust always upwards (xmphysics),
https://www.youtube.com/watch?v=yih8xQMa5zw

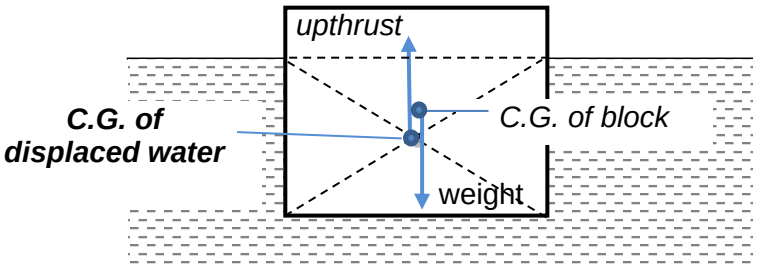
Archimedes' Principle

For a body submerged or floating in a fluid, the upthrust acting on the body **is equal in magnitude and opposite in direction** to the **weight of fluid displaced by the body**.

$$U = \rho V g,$$

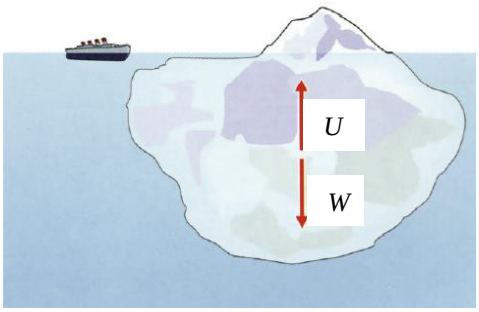
where U = upthrust,
ρ = density of fluid,
V = volume of fluid displaced by the body,
g = gravitational acceleration.

In free-body diagrams, upthrust must be drawn from the centre of gravity of the fluid displaced.



Example 3

Icebergs are often a danger to ships since most of an ice-berg's volume lies beneath the surface of the water. One can only see the "tip of the iceberg." Find the percentage of the iceberg that is submerged. (density of ice, ρ_i = 0.917 g cm^{-3}, density of sea-water, ρ_w = 1.025 g cm^{-3})



[Answer: 89.5%]

4.3 Translational Equilibrium

A large object can exhibit different kinds of motion. When we describe the object as rigid, different parts of the object cannot move relative to each other. However, the rigid object as a whole can still rotate about its centre of mass (rotational motion) while the centre of mass can have translational motion. Let us first consider the translational motion of the centre of mass of the object.

The object is said to be in translational equilibrium when it is stationary or moving at constant velocity. According to Newton's First Law, this happens when the net external force acting on the object is zero.

Condition for Translational Equilibrium:

The net external force acting on a body is zero, i.e. $\Sigma F_{ext} = 0$.

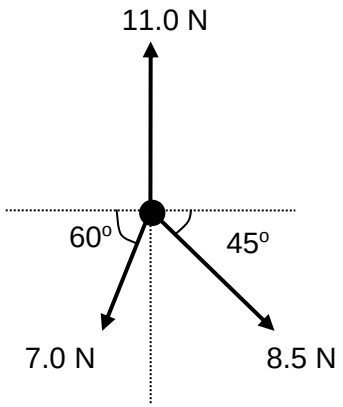
There are two common methods to determine whether an object is in equilibrium. The first method is resolving vectors and the second one is the use of a vector triangle.

In the first method, all the vectors acting on an object are resolved into horizontal and vertical components. If the sum of all the horizontal components and the sum of all the vertical components are zero, then the object is in translational equilibrium. For the horizontal and vertical component of a force F, acting at an angle θ, we typically use the following equations to resolve the force vectors: $F_h = F\cos\theta$, $F_v = F\sin\theta$.

For the special case where three coplanar forces (i.e. forces acting on the same plane) are acting on an object and the object is in translational equilibrium, the forces can be represented in magnitude and direction by the adjacent sides of a triangle. In other words, three force vectors that form a triangle for which the directions of the forces are all either clockwise or anticlockwise around the centre of the triangle have **zero resultant**.

Example 4 (Resolution of forces into component method)

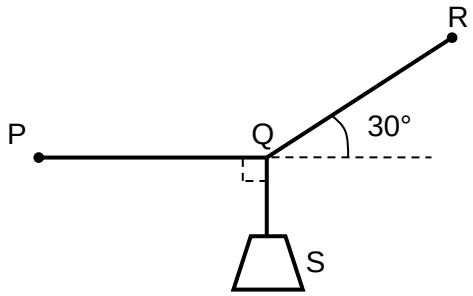
Four forces act on a particle such that it is in equilibrium. Three of the forces are shown in the diagram. Find the magnitude and direction of the unknown force F that maintains the equilibrium.



[Answer: 2.73 N, at an angle of 67° anticlockwise from the force of 11.0 N]

Example 5 (J85/I/2) (vector triangle method)

In the diagram below, a body S of weight *W* hangs vertically by a thread tied at Q to the string PQR. If the system is in equilibrium, what is the tension in the section PQ?



[Answer: 1.73*W*]

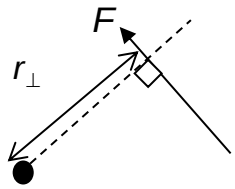
4.4 Moment of a Force

Besides translational motion, objects also exhibit rotational motion. Forces applied on a rigid body can also cause or change its rotational motion. The quantitative measure of the extent to which a force can cause or change the rotational motion of a rigid body is called the **moment of the force**, τ (Greek letter *tau*). The **moment of a force** is also called the **torque**.

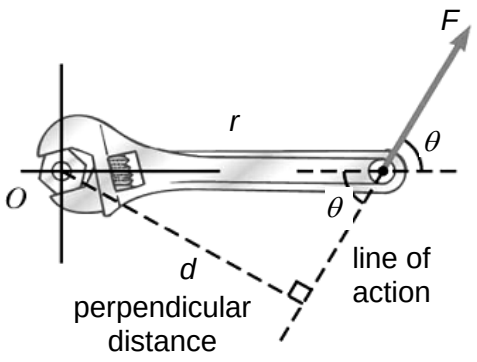
Definition of Moment of a Force

The **moment of a force** τ about a point is the product of the (magnitude of the) force ***F*** and the perpendicular distance $\boldsymbol{r}_\perp$ from the line of action of the force to the point.

$$\tau = F(r_\perp)$$



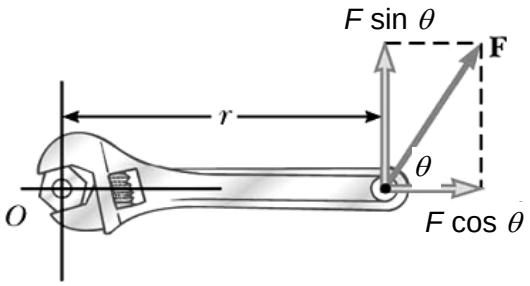
To illustrate the concept of perpendicular distance and line of action, let us consider the example of a spanner being used to loosen a bolt. A force *F* is applied at an acute angle θ to the length of the spanner and at a distance *r* away from the point *O* as shown in the diagram below.



Method 1: Finding the perpendicular distance

The line of action of a force is the line along which the force acts. The perpendicular distance *d* is the length of the line that joins the point *O* to the line of action perpendicularly. Using simple trigonometry, it can be deduced that $d = r \sin\theta$. The torque of the force *F* about the point *O* can then be calculated as

$$\tau = (F)(r_\perp) = (F)(r\sin\theta) = Fr\sin\theta\,.$$



Method 2: Finding the perpendicular component of the force

Another method is to resolve the force into two components: perpendicular and parallel to the spanner length. Only the perpendicular component $F \sin\theta$ produces the turning effect. The line of action of the parallel component $F \cos\theta$ passes through the reference point O and produces no turning effect. The torque of the force F about the point O is then given by

$$\tau = (r)(F_{\perp}) = (r)(F\sin\theta) = Fr\sin\theta .$$

Since both approaches lead to the result $\tau = Fr\sin\theta$, one could treat the equation as either

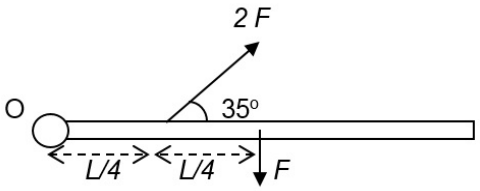
1. magnitude of force × perpendicular distance (between the reference point and the line of action),

or

2. distance (from reference point to point of application of force) × perpendicular component of the force.

Example 6

A uniform rod of weight F and length L is pivoted on one end O. A force $2F$ is applied at a quarter of its length from O. Find the magnitude of the net moment about point O.



[Answer: 0.213*FL*]

4.5 Rotational Equilibrium and Principle of Moments

A rigid object is said to be in rotational equilibrium when the net moment of the object about any point is zero. Note that we may freely choose the reference point to be anywhere (even outside the object) when calculating the net moment of the object. If the object is in rotational equilibrium, regardless of where we choose reference point to be, the net moment about that point will be zero.

Condition for Rotational Equilibrium:
The net moment on the body ***about any point*** is zero, i.e. $\Sigma\tau = 0$.

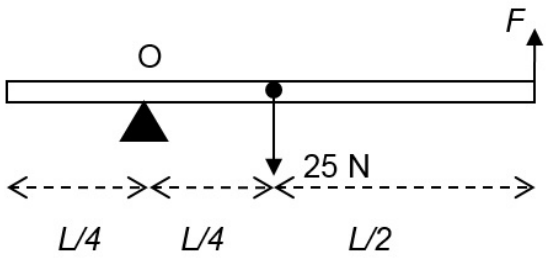
When all lines of action of the forces are in the same plane (the forces are *coplanar*), the condition for rotational equilibrium is simplified to the principle of moments.

For rotations confined to a single plane:

Principle of moments states that for a body to be in rotational equilibrium, the sum of clockwise moments about any point equals the sum of anticlockwise moments about that same point.

Example 7

A uniform rod of weight 25 N and length *L* is balanced on a knife edge at O. Find the unknown force *F* required to maintain rotational equilibrium.



[Answer: 8.33 N]

4.6 Static Equilibrium for Rigid Extended Bodies

Equilibrium is a concept to describe the state of a body that is either at rest or moving at a constant velocity. The conditions for equilibrium are the same, regardless whether the body is at rest or moving at a constant velocity. However, if the body is at rest, we say that the body is in *static* equilibrium.

Conditions for (Static) Equilibrium of a rigid extended body:
The net external force acting on the body is zero. $\Sigma F_{ext} = 0$.
The net moment on the body *about any point* is zero. $\Sigma\tau = 0$.

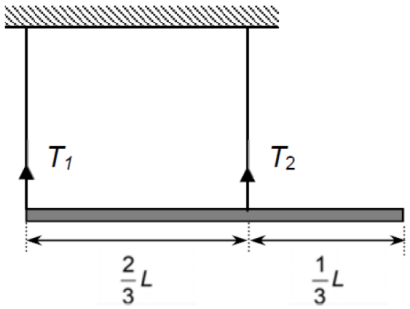
TIPS **Solving Static Equilibrium problems**

1. Isolate the body of interest.
2. Draw the free-body diagram and identify **all** forces acting on the body.
3. Set up the torque equation (condition for rotational equilibrium)
   - Choose a "good" reference point.
   - A "good" reference point is a point which lies on the lines of action of one or more of the unknown forces, thereby reducing the number of unknown terms in the equation.
   - Apply the condition $\sum\tau = 0$,
     sum of clockwise moments = sum of anti-clockwise moments.
4. Set up the two force equations (*x* and *y* components) (conditions for translational equilibrium)
   - Choose two perpendicular directions *x* and *y*.
   - Resolve all forces into *x* and *y* components.
   - Apply the condition $\sum F_x = 0$,
     sum of leftward forces = sum of rightward forces
   - Apply the condition $\sum F_y = 0$,
     sum of upward forces = sum of downward forces
5. Solve the equations to find the unknown forces.

Example 8 (J95/I/5)

A uniform beam of length *L* and mass *M* is supported by two vertical cords as shown in the diagram.

What is the ratio $\frac{T_1}{T_2}$ of the tensions in these cords?



[Answer: 0.333]

4.7 Three-Force Systems

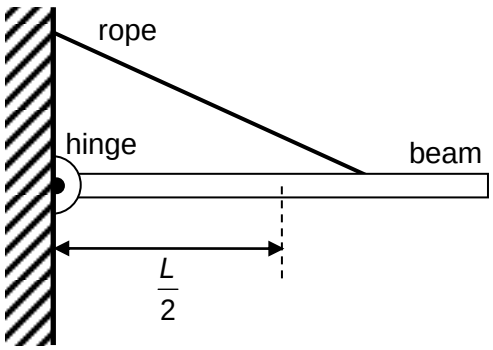
A solid body submitted to three forces whose lines of action are not parallel is in equilibrium, if the following three conditions apply:

1. The lines of action are coplanar (⇒ all forces lie in one and the same plane)
2. The vector sum of the forces is equal to the zero vector (⇒ they can form a closed vector triangle)
3. The lines of action are *convergent* (⇒ their lines of action cross through a single point).

Reference: The 'Three forces must intersect at the same point' Rule, (xmphysics)
https://www.youtube.com/watch?v=u5bF80D0D2c

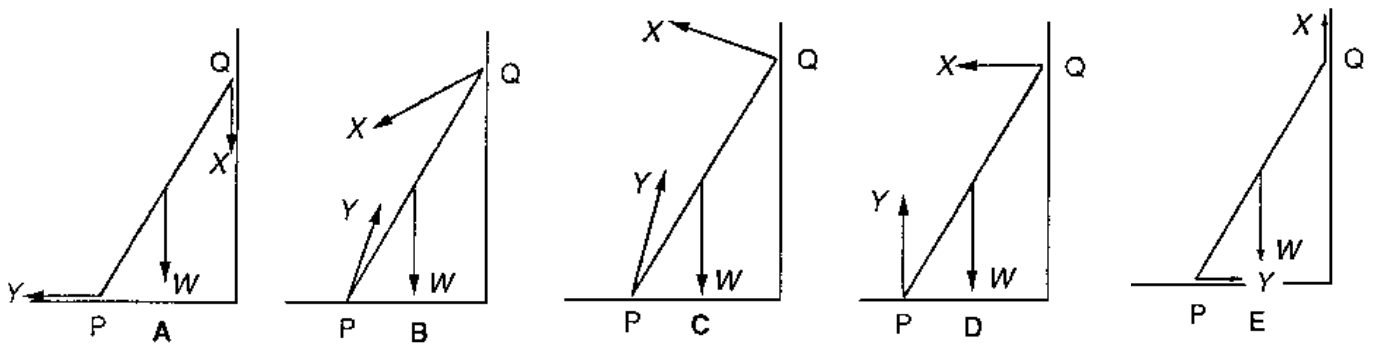
Example 9

A uniform beam balance of length *L* is hinged to a wall and is supported by a rope as shown. Show the direction of the force exerted by the hinge (attached to the wall) on the beam.



Example 10 (N84/II/1)

A ladder PQ, resting on a rough floor and leaning against a rough wall, is on the point of slipping. It is of weight *W* and the contact forces exerted on the ladder by the wall and the floor are *X* and *Y,* respectively. Which one of the following diagrams correctly shows the directions of these forces?



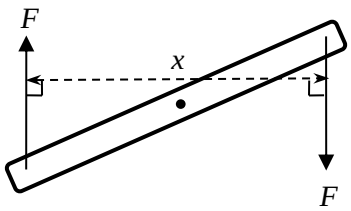
4.8 Couple

Definition of a Couple

A pair of forces, which are equal in magnitude but opposite in direction, whose lines of action do not coincide.

Couples produce **only rotation** but no translation.

An example of a couple is shown in the figure below.



Taking moments about the centre of mass,

the torque due to the couple $=\left(\frac{1}{2}x\right)F+\left(\frac{1}{2}x\right)$

$=xF$ (in the clockwise direction),

where F is the magnitude of the forces and x is the *perpendicular distance* between the lines of action.

In fact, taking moments about **any point** would produce the same result.

Torque of a couple is the product of the perpendicular distance between the lines of action of the forces and the magnitude of one of the forces.

Example 11

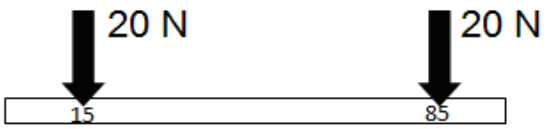
The following diagrams show forces acting on uniform metre rules. The rulers are lying on a horizontal smooth table top and shown as seen from above.

Which of the diagrams do not show a couple?

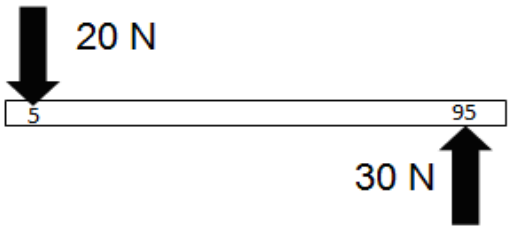
For the other diagrams, what is the torque of the couple?

Assume that for each option, the two forces are parallel to each other.

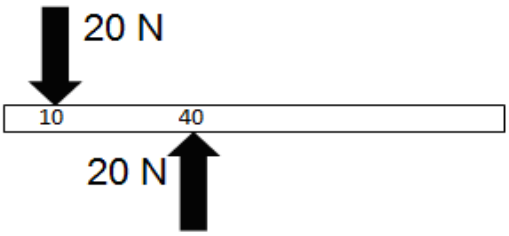
A



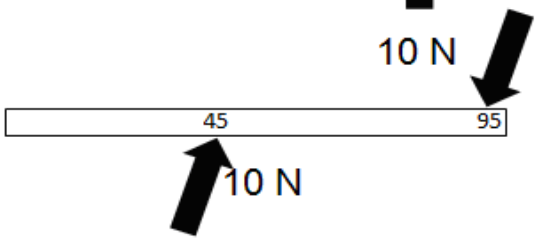
B



C



D



Appendix I

Contact force and Frictional force

When two bodies are in contact such as a stone at rest on a road, the two rough surfaces actually make close contact only at relatively few places. Where contact is made, the road will exert a force on the stone as shown in Fig. (a). The vector sum of these forces is represented by the single force known as the contact force which the road exerts on the stone as shown in Fig. (b).

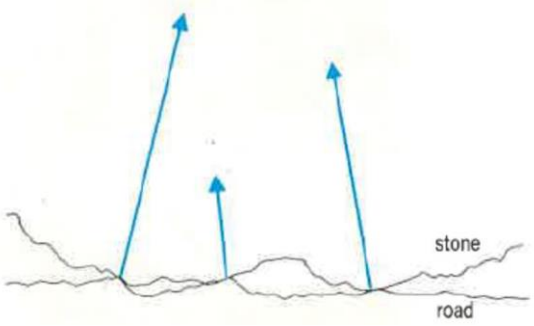


Fig. (a)

Fig. (b)

[on a different scale from fig. (a)]

If the stone slides across the road to the left say, the forces exerted by the road on it are likely to be both upwards and to the right as shown in Fig. (c). The resultant of all these forces is hence tilted to the right (Fig. (d)). The horizontal component of this force is known as the frictional force and the vertical component is known as the normal contact force. Friction which is in the opposite direction to the direction of motion, is usually a component of a contact force. For two surfaces where friction can be neglected, the only force between them will be the normal contact force.

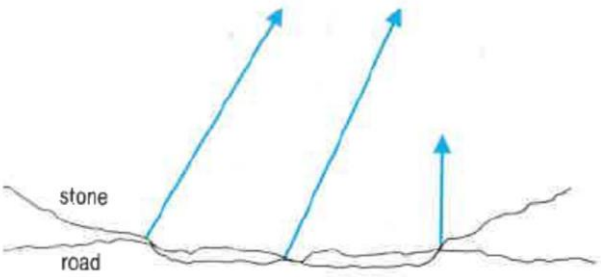


Fig. (c)

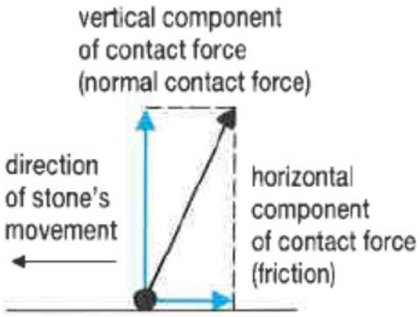


Fig. (d)

[on a different scale from Fig. (c)]

(Ref: Physics, Robert Hutchings)

Appendix II

Viscous Force (air resistance and air drag)

A viscous force refers to the resistance force which a body experiences when it moves through a fluid (a gas or liquid). As the moving body exerts a force on the fluid to push it out of the way, the fluid would in turn exert an equal and opposite force on the body (by Newton's third law).

Two things to note on viscous forces:

(1) The direction of the fluid resistance acting on a body is always opposite the direction of the body's velocity relative to the fluid.

(2) Magnitude of the fluid resistance *f* increases with *v* the speed of the body through the fluid.

(a) For small objects moving at very low speeds, the air resistance is proportional to v, ($f \propto v$)

(b) For larger objects (such as airplanes) moving at very fast speeds (through air), the fluid resistance also known as air drag is proportional to v^2 , ($f \propto v^2$)

[refer to kinematics notes, section 2.6 Viscous Drag Force (pg 17-18) on other factors affecting the viscous force]

(Ref: Sears & Zemansky's University Physics with Modern Physics, Hugh D. Young, Roger A. Freedman)

(image: https://www.straitstimes.com/singapore/rsaf-gain-first-hand-experience-with-australian-air-force-s-f-35as-in-training-exercise , Straits Times, Nov 24, 2022)

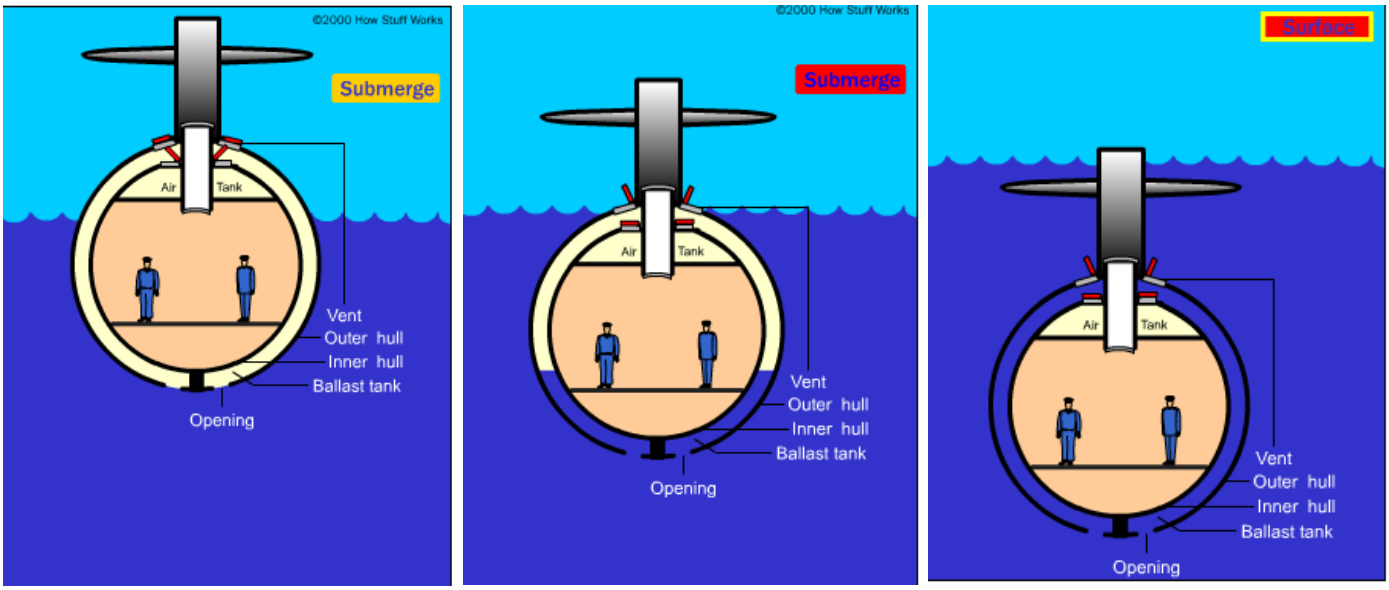
Appendix III

Buoyancy of a submarine

A submarine or a ship can float because the weight of the water it displaces is equal to the weight of the ship. This displacement of water creates an upward force called the buoyancy (upthrust). Unlike a ship, a submarine can control its weight, thus allowing it to sink and surface at will.

To control its weight, the submarine has ballast tanks and auxiliary, or trim tanks, that can be filled with water or air. A supply of compressed air is maintained aboard the submarine in air flasks for life support and for use with the ballast tanks.

When the submarine is on the surface, the ballast tanks are filled with air and the submarine's overall density is less than that of the surrounding water. As the submarine dives, the ballast tanks are flooded with water and the air in the ballast tanks is vented from the submarine until its overall density is greater than that of the surrounding water and the submarine begins to sink. To keep the submarine level at any set depth, the submarine maintains a balance of air and water in the trim tanks so that its overall density is equal to the surrounding water.



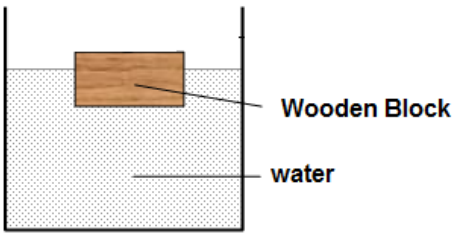
Tutorial 4: Forces

Self-Review Questions

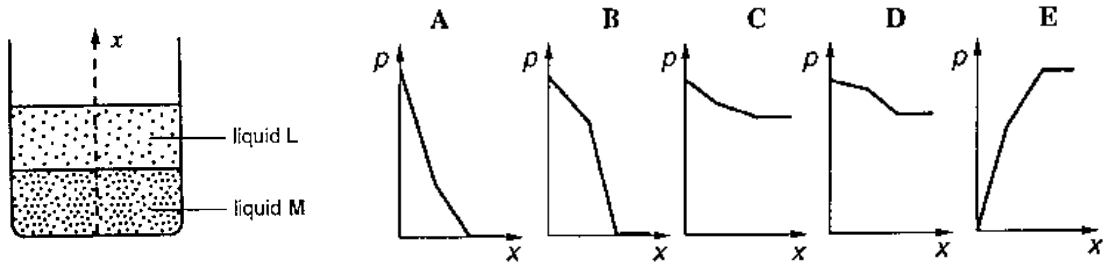
Use these questions to test your familiarity with the concepts. These questions should be sufficiently easy such that you can solve them on your own, with a little bit of thinking, without help from the tutors. The solutions are made available on the cohort Google Classroom for self-check.

S1. A wooden block, with a density of 400 kg m^{-3} and a volume of 0.030 m^3, is floating in water. The density of water is 1000 kg m^{-3}.

(a) What is the net force acting on the wooden block?
(b) What is the weight of wooden block?
(c) What is the upthrust acting on the block?
(d) What is the volume of the block beneath the surface of water?



S2. (J93/P1/Q23) A tall container, which is open to the atmosphere, contains a layer of liquid L, floating on liquid M. Liquid M has a density which is twice as great as that of liquid L. Which graph shows how the pressure *p* at a point varies with its height *x* above the base of the container?

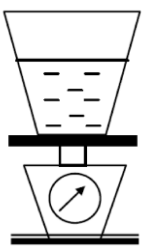


S3. An inextensible string supports a solid iron object of mass 180 g, totally immersed in a liquid of density 800 kg m^{-3}. The density of iron is 8000 kg m^{-3}. Calculate the tension in the string.

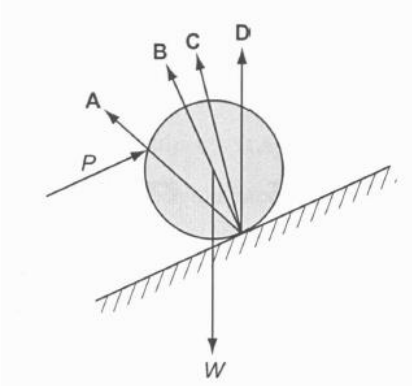
S4. A cup, half-filled with water, is resting on a weighing scale which registers a reading *W*. When a boy dips his finger in the water without touching the base, the reading of the weighing scale changes to *W′*.

Which of the following statements is correct?

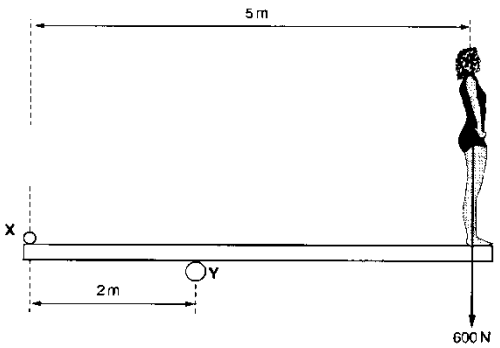
A *W′* = *W* because the water has not overflowed from the beaker.
B *W′* < *W* because the water exerts an upthrust on the boy's finger.
C *W′* > *W* because the water exerts an upthrust on the boy's finger, and thus the boy's finger exerts a force back on the water.
D *W′* > *W* because the weight of the boy's finger is added to that of the water in the beaker.



S5. (2010/H1/P1/Q8) A full barrel of weight *W* is being rolled up a ramp. The force *P* is required to hold the barrel at rest on the ramp. Friction between the barrel and the ramp stops the barrel from slipping.
Which arrow represents the resultant force the ramp exerts on the barrel?



S6. (J/96/P1/Q5) The diagram shows a light diving-board held in position by two rods X and Y.
Which additional forces do these rods exert on the board when a diver of weight 600 N stands on the right-hand end?



	At X (downwards)	At Y (upwards)
A	400 N	1000 N
B	600 N	1200 N
C	900 N	600 N
D	900 N	1500 N

S7. (2009/H2/P2/Q3) Fig.3.1 shows a force diagram that represents a boat that is being lifted by two ropes so that the boat remains horizontal and travels vertically upwards at a constant speed after leaving the water.

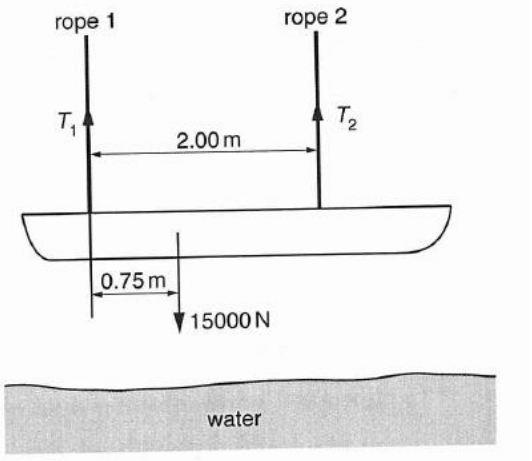


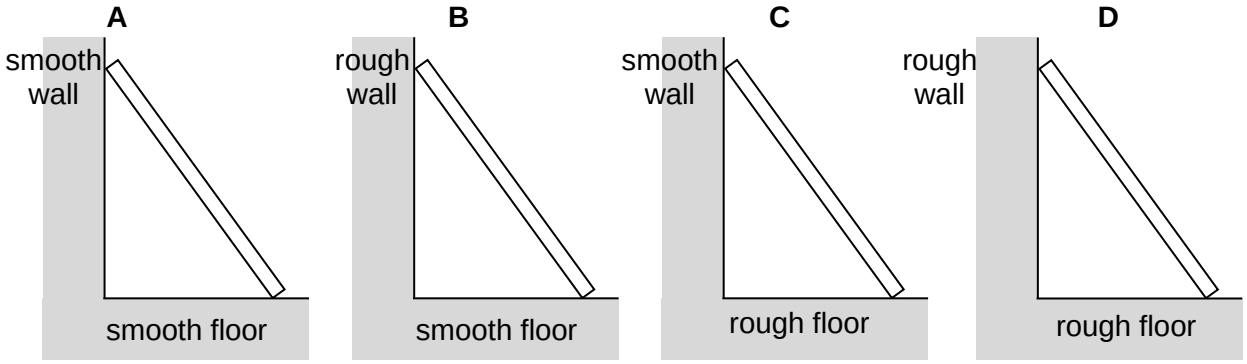
Fig. 3.1

The weight of the boat is 15 000 N and the tensions in the ropes 1 and 2 are T_1 and T_2, respectively.

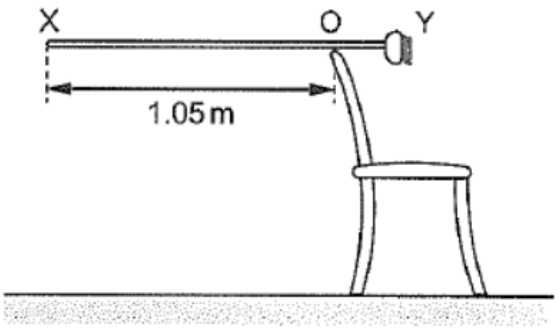
(a) The position of the centre of gravity of the boat is not at its midpoint. Suggest what this implies about the distribution of mass in the boat.
(b) Explain two conditions required for the boat to be in a state of equilibrium while it is moving upwards.
(c) Use the principle of moments to determine the tensions in the two ropes.

Discussion Questions

D1. Consider a uniform ladder leaning against a wall. By drawing and labelling the forces acting on the ladder for each scenario, determine the scenario(s) in which the ladder will definitely slip.



D2. (N/18/P1/Q5) A student balances a long-handled broom XY horizontally across the top of the back of a chair. The top of the chair O is at a distance 1.05 m from the end X of the handle.
A 200 g mass is tied to the handle by a thread at a distance of 0.10 m from the end of X of the handle. The broom now needs to be moved 0.27 m to the right to balance horizontally.
What is the mass of the broom?



A 500 g
B 580 g
C 700 g
D 830 g

D3. (N/16/P2/Q3)
A non-uniform bar AB makes an angle of 60° with a horizontal surface, as shown in Fig. 3.1.

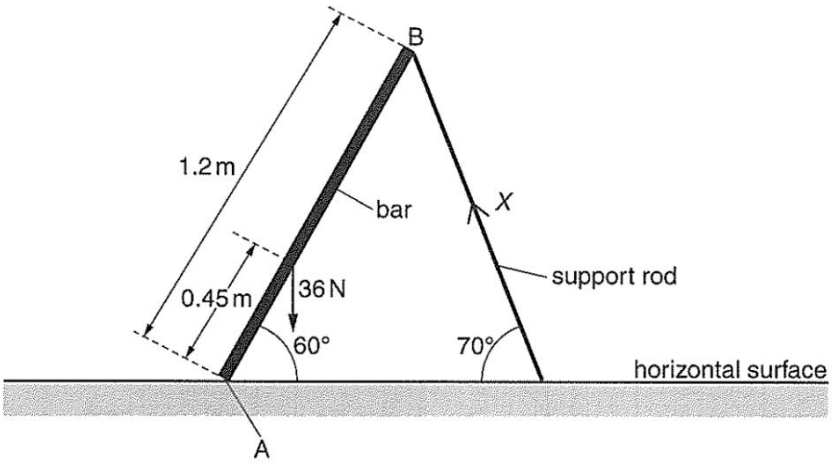


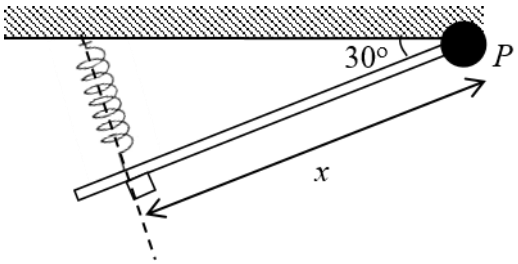
Fig. 3.1

The bar is hinged at A and is supported by a rod at B. The force *X* produced by the rod at B acts at an angle of 70° to the horizontal.
The bar has a length of 1.2 m and a weight of 36 N.
The centre of gravity of the bar is 0.45 m from A.

(a) Use the principle of moments to show that the magnitude of X is 8.8 N.

(b) A force F acts on the bar at A.
(i) Explain why a force is required to act on the bar at A to keep the bar in equilibrium.
(ii) Calculate the magnitude of F.
(iii) On Fig. 3.1, draw an arrow to show the approximate direction of F.

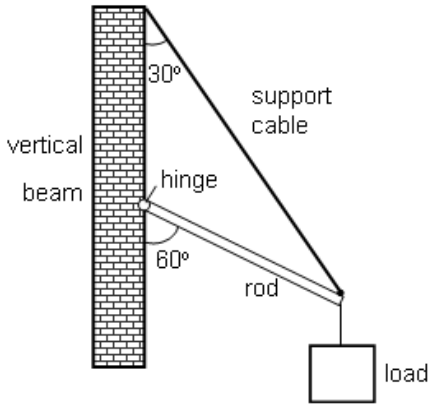
D4. A uniform rod of mass 8.0 kg and length 0.50 m is pivoted about point P. It is held in position by a spring of original length 0.20 m, spring constant 5.0 x 10^2 N m^{-1}, and negligible mass. Find the length x.



D5. RJC/2009/Prelim/P3/Q1
(a) Define moment of a force.

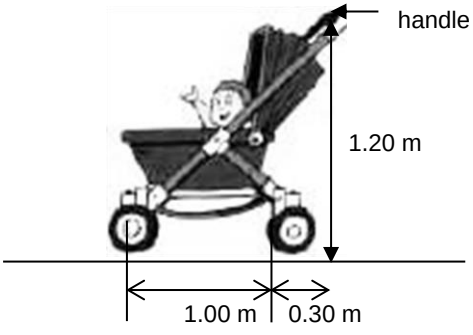
(b) A uniform rigid rod of weight W_1 = 400 N is attached to a vertical beam by a hinge as shown. The other end of the rod is fastened to a support cable. The structure is used to support a load of weight W_2 = 2000 N.

(i) Show that the tension T in the support cable is 3810 N.
(ii) Determine the magnitude and direction of the horizontal and vertical components of the force acting on the rod by the hinge.

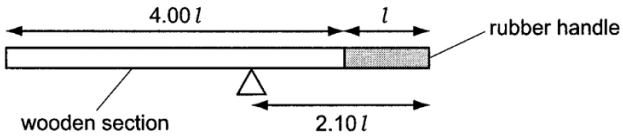


D6. (HCI/BT/05/P2/Q3(b))
The diagram shows the side view of a baby stroller. The combined mass of stroller and the boy is 22.0 kg.

(i) Assuming that the centre of gravity of the boy and stroller lies 0.40 m in front of the hind wheels and 0.35 m above the ground, determine the forces experienced by the front wheels F_1 and that experienced by the hind wheels F_2 from the ground.
(ii) As the stroller is not easy to manoeuvre, it is common to see parents hanging their groceries at the handle to free their hands. Determine the maximum load that can be placed at the handle before the stroller topples over.
(iii) It is still extremely dangerous to hang groceries at the handle even though it may be less than the maximum load calculated in (ii). Suggest why this is so.



D7. (2008/P1/Q8) A uniform rod has a wooden section and a solid rubber handle, as shown.



The length of the handle is ℓ and the length of the wooden section is 4.00 ℓ. The rod balances a distance 2.10 ℓ from the rubber end. What is the ratio $\frac{\text{density of rubber}}{\text{density of wood}}$?

A 1.71
B 2.25
C 2.50
D 3.27

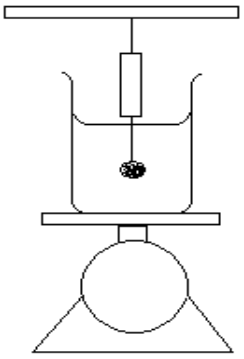
D8. A piece of solid weighs W_1 in air and W_2 when totally immersed in a liquid. Given that the density of the liquid is ρ, determine the volume of the solid in terms of W_1, W_2, ρ and g.

D9. (2012/H2/P1/Q7) A small air bubble in some water is rising to the surface with constant velocity. The volume of the bubble is 2.370×10^{-8} m^3. The density of water is 1000 kg m^3. The density of air is 1.290 kg m^{-3}. What is the magnitude of the viscous force on the bubble?

A 2.367×10^{-5} N
B 2.373×10^{-5} N
C 2.322×10^{-4} N
D 2.328×10^{-4} N

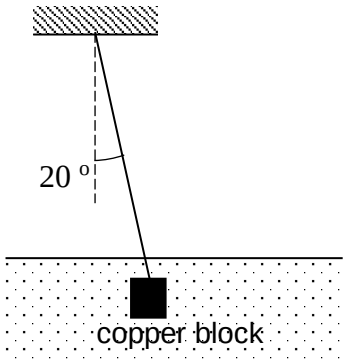
D10. A cube of wood 20 cm on a side and having a density of 0.65×10^3 kg m^{-3} floats on water. Take the density of water to be 1.0×10^3 kg m^{-3}.
(a) What is the distance from the top surface of the cube to the water level?
(b) Find the mass of lead that has to be placed on top of the cube so that the cube's top is just level with the water level.

D11. The weight indicated on a weighing machine is X when a beaker of water is placed on it. A solid object has weight Y in air and displaces weight Z of water when completely immersed. The figure shows the object suspended from a spring balance and completely immersed in the beaker of water. What are the readings on the spring balance and the weighing machine in the given arrangement?



	Spring balance	Weighing machine
A	$Y - Z$	X
B	$Y + Z$	$X + Y - Z$
C	$Y + Z$	$X + Y$
D	$Y - Z$	$X + Z$

D12. A copper block of mass 0.50 kg is hung from the end of a light, inextensible thread and immersed in a fast-flowing river, as shown. The other end of the thread is suspended from a fixed point above water. The river currents caused the copper block to be displaced to the right such that the thread makes an angle of 20 ^{o} with the vertical. The density of copper is 9.00 x 10^{3} kg m^{-3} and the density of the river water is 1.00 x 10^{3} kg m^{-3}.



Calculate the magnitude of the horizontal force exerted on the copper block by the river currents.

D13. (HCI/BT/08/P2B/Q1)
A spherical steel ball bearing of diameter 1.40 cm is placed on a wooden block of mass 50.0 g, such that the block and ball bearing are floating in a cylinder of oil, as shown in Figure 1.

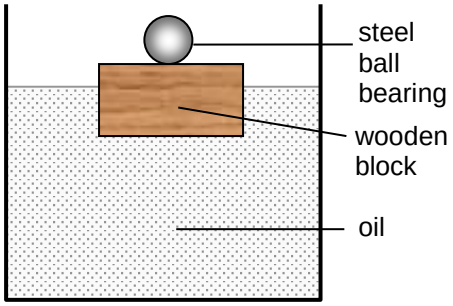


Figure 1

In the following parts of the question, you may assume that:
density of steel, ρ_{steel} = 7.85 g cm^{-3}
density of oil, ρ_{oil} = 0.83 g cm^{-3}

i. Consider the ball bearing and the wooden block as a single system, as in Figure 2. Draw labelled arrows to indicate the external forces acting on this system.
ii. Determine the *upthrust* acting on the system in (i).
iii. Hence, calculate the volume of oil displaced by the wooden block to keep the system afloat.
iv. The ball bearing now rolls off the piece of wooden block and falls into the oil. Explain how the oil level in the cylinder will change.

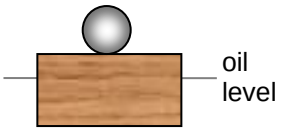


Figure 2

Numerical Answers to Tutorial 4: Forces

D3 (b) (ii) 27.9 N
D4 0.471 m
D5 (b) (ii) 900 N & 1910 N
D6 (b) (i) F_1 = 86 N, F_2 = 129 N; (ii) 288 N
D8 (W_1 – W_2) / (ρg)
D10 (a) 7 cm; (b) 2.8 kg
D12 1.59 N
D13 (ii) 0.601 N; (iii) 73.8 cm^3

FORCES SUMMARY

Types of Forces

exerted between bodies in physical contact → **Contact Forces**

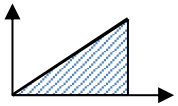
action at a distance → **Non-Contact Forces**

- Gravitational force between masses
- Electric force between charges
- Magnetic force between currents

Spring/Elastic force

For springs that **obey Hooke's Law:**

- External force F_{ext} pulling on extending or compressing spring,
  $F_{ext} = kx$
  where
  k is the spring constant;
  x is the extension or compression



- Restoring force in spring
  $F_{spring} = kx$

Elastic potential energy stored in spring,
U = Area under $F - x$ graph
$= \frac{1}{2} Fx = \frac{1}{2} kx^2$

Pressure due to Fluid

(a fluid is any liquid or gas)

- is the **force acting per unit area** by the fluid on a body submerged at a depth in the fluid.
- **Hydrostatic Pressure** or the *pressure due to a fluid at a depth* h is given by:
  $p = h\rho g$
  where
  h is depth
  p is the hydrostatic pressure at this depth
  ρ is the density of the fluid
  g is the gravitational acceleration

Upthrust or Buoyant Force

- is the **net upward force exerted by a fluid** on a body floating/submerged in the fluid.
- **Archimedes' Principle**
  (used to compute *the magnitude* of the upthrust):
  Upthrust (U) on object = Weight of fluid displaced by object
  $U = m_f g = \rho_f V_{dis} g$
  where
  U is the upthrust acting on the object
  m_f is the mass of the fluid displaced
  g is the gravitational acceleration
  V_{dis} is the volume of the fluid displaced
  ρ_f is the density of the displaced fluid

For an **object floating in equilibrium** in a fluid (**Principle of Flotation**):

Weight of object = Upthrust on object

Drag/Viscous Force

- Force resisting an object **moving relative to a fluid**. (e.g. air resistance)
- Always **opposes** motion (hence, no motion, no force)
- Magnitude is **dependent on the velocity** of the object **relative to** the fluid, e.g.,
  laminar (streamline) flow: $F_D = kv$
  turbulent flow: $F_D = kv^2$
  where
  F_D is the drag force acting on the object
  k is a constant dependent on the dimensions of the object the type of fluid
  v is the velocity of the object relative to the fluid
  (relationship usually provided in exam questions)

Equilibrium of Forces

2 conditions for a rigid body to be in static equilibrium:

1) **Translational equilibrium: net external** force acting on the body is zero.

$$\sum F = 0$$

$$\Rightarrow \sum F_x = 0 \;\&\; \sum F_y = 0 \text{ (\& in fact, } \sum F_z = 0)$$

2) **Rotational equilibrium: net torque** on the body about **any point** is zero.

$$\sum \tau = 0$$

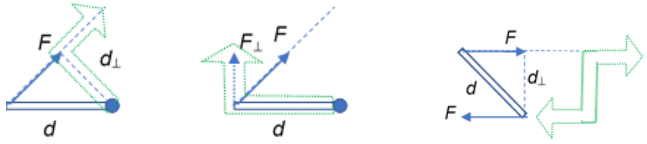
$\Rightarrow$ sum of clockwise moments = sum of anticlockwise moments

Note:

- For 3-forces system in static equilibrium, the 3 forces form ***a closed vector triangle***.
- acting on an ***extended body in static equilibrium***, the line of action of the 3 forces must ***intersect at a common point*** unless the 3 forces are //.

Turning Effects of Forces

The ***moment of a force*** about a point is the **product** of the **magnitude of the force** F and the $\perp$ **distance** $d_\perp$ of the **line of action** of the force to the point. $\tau = Fd_\perp$

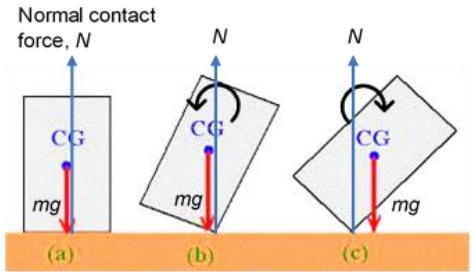


A ***couple*** consists of 2 parallel forces which are equal in magnitude and opposite in direction

The ***torque of a couple*** is the **product** of the **magnitude of one of the forces** of the couple and the **perpendicular distance** between the forces

Centre of gravity (c.g.) of a body

is the point at which the weight of the body appears to act. E.g.



Lower c.g. increases stability of object.

Chapter 5

Work, Energy and Power

Pictures (from top left and bottom left to right):
credits to www.knxradio.com/guardian.co.uk/fluidmechanic.co.uk/aafaq.genistra.com

"There is a fact, or if you wish, a law governing all natural phenomena that are known to date. There is no known exception to this law – it is exact so far as we know. The law is called the conservation of energy.

It states that there is a certain quantity, which we call "energy" that does not change in the manifold changes that nature undergoes. That is a most abstract idea, because it is a mathematical principle; it says there is a numerical quantity which does not change when something happens.

It is not a description of a mechanism, or anything concrete; it is a strange fact that when we calculate some number and when we finish watching nature go through her tricks and calculate the number again, it is the same.

It is important to realize that in physics today, we have no knowledge of what energy "is." We do not have a picture that energy comes in little blobs of a definite amount. It is not that way. It is an abstract thing in that it does not tell us the mechanism or the reason for the various formulas."

Richard Feynman
The Feynman Lectures on Physics (Volume 1)

Contents

H2 Physics Syllabus 9749 3
5.1 Work 4
5.1.1 Work done by a constant force 4
5.1.2 Work done by a variable force 6
5.2 Energy 7
5.2.1 Kinetic energy, E_k 7
5.3 Potential energy 9
5.3.1 Gravitational Potential Energy (GPE), E_p or U 10
5.3.2 Relationship between Gravitational Force and Gravitational Potential Energy 12
5.3.3 Elastic Potential Energy 13
5.3.4 Work done by External Force on a Deformed Material 14
5.4 Principle of Conservation of Energy 15
5.5 Power 17
5.5.1 Relationship between Power, Force and Velocity 18
5.6 Efficiency (of energy conversion), η 19
Appendix 1 Potential Energy (U) and Conservative Force (Good to know) 20
Self Review Questions 22
Discussion Questions 24

Videos of Lecture Examples can be found at
https://youtube.com/playlist?list=PL_b5cjrUKDlYIoMzYi383KRNFmeWtS5Mh

H2 Physics Syllabus 9749

Topic 5: Work, Energy and Power

Content

- Work
- Energy conversion and conservation
- Potential energy and kinetic energy
- Power

Learning Outcomes

Candidates should be able to:

(a) define and use work done by a force as the product of the force and displacement in the direction of the force.

(b) calculate the work done in a number of situations including the work done by a gas which is expanding against a constant external pressure: $W = p\Delta V$ **(*will be covered in Thermal Physics)**

(c) give examples of energy in different forms, its conversion and conservation, and apply the principle of energy conservation.

(d) show an appreciation for the implications of energy losses in practical devices and use the concept of efficiency to solve problems.

(e) derive, from the equations for uniformly accelerated motion in a straight line, the equation $E_k = ½ mv^2$.

(f) recall and use the equation $E_k = ½ mv^2$.

(g) distinguish between gravitational potential energy, electric potential energy and elastic potential energy.

(h) deduce that the elastic potential energy in a deformed material is related to the area under the force-extension graph.

(i) show an understanding of and use the relationship between force and potential energy in a uniform field to solve problems.

(j) derive, from the definition of work done by a force, the equation $E_p = mgh$ for gravitational potential energy changes near the Earth's surface.

(k) recall and use the equation $E_p = mgh$ for gravitational potential energy changes near the Earth's surface.

(l) define power as work done per unit time and derive power as the product of a force and velocity in the direction of the force.

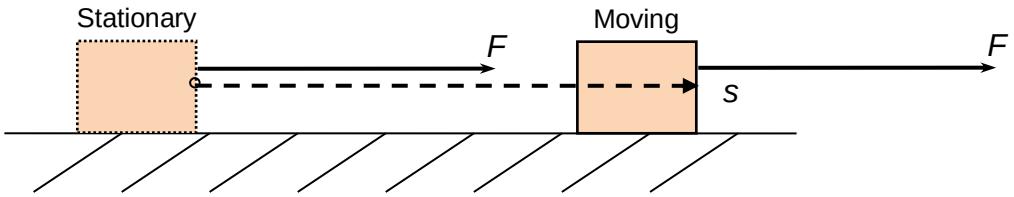
5.1 Work

The meaning of 'work' in physics differs from everyday usage. Work done by a force acting on a body refers to the amount of energy that was transferred to the body by the agent that exerts the force.

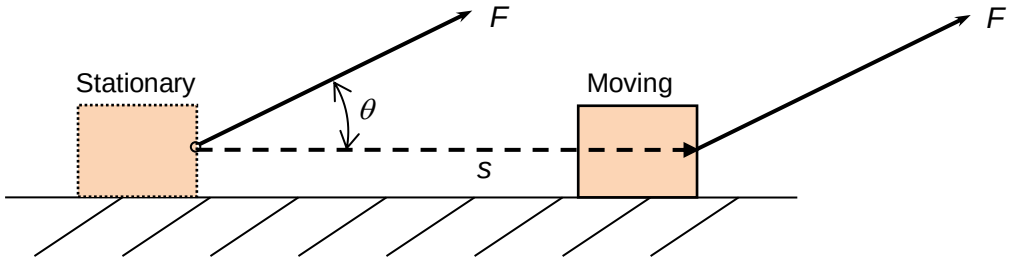
A person holding a heavy book at rest may feel that he is doing work as he feels tired after a while but, although he is exerting a force on the book, there is no energy transferred to the book as a result and hence the work done on the book by him is zero.

5.1.1 Work done by a constant force

Consider a constant force *F* acting on a body over a period of time, during which the object has a displacement *s* in the direction of the force.



The work done BY the force ON the body is defined as the **product of the magnitude of the force *F* and the displacement *s* in the direction of the force.**



For the general case when the displacement is not in the direction of the constant force, the work done by the force is the product of the displacement and the component of the force in the direction of the displacement, or, equivalently, the product of the force and the component of the displacement that is in the direction of the force. As can be seen below, both have the same mathematical expression.

$$W = F s_{//} = F (s \cos \theta)$$

Alternatively $$W = F_{//} s = (F \cos \theta)\, s$$

where *F* is the magnitude of constant force
s is the displacement
θ is the angle between the force and displacement
* *subscript // refers to its vector component parallel with respect to some other vectors*

Notes:

- Work is a **scalar** quantity. The unit of work is the **joule** (J).
- It is assumed that the object is *not rotating*, or *changing shape* (i.e. it is rigid), so all parts of the object move with the same displacement. In this case, the work done by a force can be calculated by using a point to represent the object, with the force acting on that point
- Where there are more than one force acting on a body, work done by the individual forces can be calculated separately.

Concept Check 1

As the man is walking around at a constant speed and holding up the load, is any work being done on the load? Explain why.

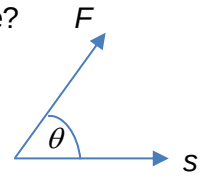
Since F and s are perpendicular.

Work done by Man exerting force F on the load **= F s cos 90° = 0 J**

Similarly work done by gravitational force is also zero.

Concept Check 2

Under what circumstance is the work done by a force acting on a body negative?



$W = Fs\cos\theta$

Work W is negative
when $180^{o} > \theta > 90^{o}$, cos θ *is negative.*
e.g. When $\theta = 180^{o}$, cos θ = -1

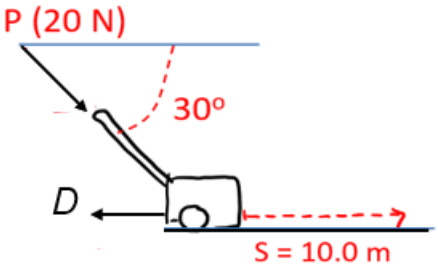
Qn: What does this mean physically? The force is resisting the motion and doing negative work to reduce the kinetic energy of the body.

Example 1

A lawn mower is pushed by a force P of 20.0 N at 30° to the horizontal, against a horizontal resistive force D of 8.0 N. It moves a horizontal distance of 10.0 m. Calculate

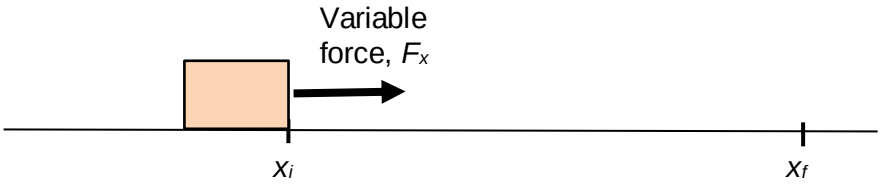
(i) the work done by the force P on the lawn mower.
(ii) the work done by the resistive force D.
(iii) the work done by the normal contact force from the ground and the weight of the lawn mower.
(iv) the change in kinetic energy of the mower.

[173 J, - 80.0 J, 0 J, 93 J]

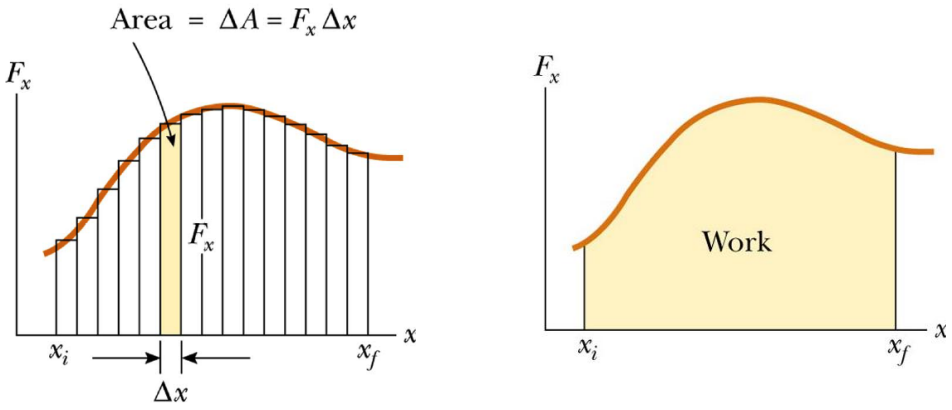


5.1.2 Work done by a variable force

Suppose an object is being displaced along the x-axis under the action of a variable force F_x that acts in the x direction. The object is displaced in the direction of increasing x from $x = x_i$ to $x = x_f$.



In order to calculate work done by the ***variable force*** F_x, we first need to know how the force F_x vary with displacement. Let's imagine we were able to use a force sensor and a data logger to measure this variation and that we were able to get the plot shown below.



The total displacement x can be regarded as the sum of many very small steps Δx. During each small step, the force is practically constant.

The total work done = sum of work done for each step

$$W = \Delta W_1 + \Delta W_2 + \Delta W_3 + \ldots\ldots$$

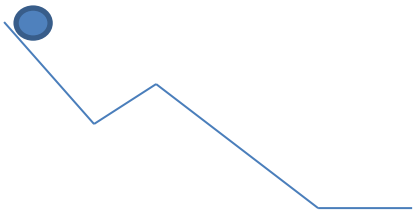
$$= F_1\Delta x + F_2\Delta x + F_3\Delta x + \ldots..$$

As the width of the small steps, Δx, gets really small, the sum of area of these strips will just be the area under the graph of F_x against x.

$$W = \int_{x_i}^{x_f} F_x \, dx$$

In other words, **the work done by variable force acting on an object that undergoes a displacement is equal to the area under the force–displacement graph**.

5.2 Energy



Suppose a ball is allowed to roll down a hill and we would like to calculate its speed by the time it reaches the bottom of the hill. The ball does not follow a straight path as it is not a straight slope. It will be very tiresome to try and calculate the speed of the ball using what we learnt in kinematics.

An alternative and fast approach to solving this problem is to apply the principle of conservation of energy. If we ignore friction, there are only two forms of energy we need to concern ourselves with, the kinetic energy of the ball by virtue of its motion and the gravitational potential energy stored in the ball-earth system.

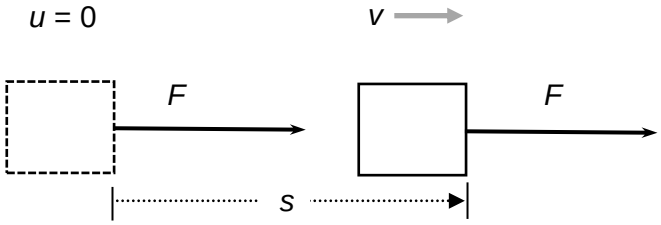
The ball might gain or lose some of its kinetic energy as it rolls down the slope but if we realise that all its kinetic energy gained must be balanced by a corresponding decrease in the gravitational potential energy of the ball-earth system (assuming the earth is stationary), we could easily find the kinetic energy it had at the bottom of the slope by simply comparing the change in the gravitational potential energy of the ball-earth system.

5.2.1 Kinetic energy, E_k

Kinetic energy of a body is a measure of the energy possessed by the body **by virtue of its motion.**

Derivation of $E_k = \frac{1}{2}mv^2$

Consider an object, initially at rest, is being displaced along the *x*-axis under the action of a constant force,



We can infer the formula for kinetic energy from the amount of work that is done by an external force to bring a body from rest to its state of motion.

$$E_k = F\,s$$

By Newton's second law: $E_k = (ma)s$

Using the equation of motion for uniform acceleration: $v^2 = u^2 + 2as$,

Re-arrange: $as = \frac{1}{2}\left(v^2 - u^2\right)$

Since $u = 0$, $as = \frac{1}{2}v^2$

Thus, the kinetic energy of the object gained through work done by the constant force:

$$E_k = m(as)$$
$$= m\left(\frac{1}{2}v^2\right)$$
$$\therefore\ E_k = \frac{1}{2}mv^2$$

Work- Energy Theorem (good to know)

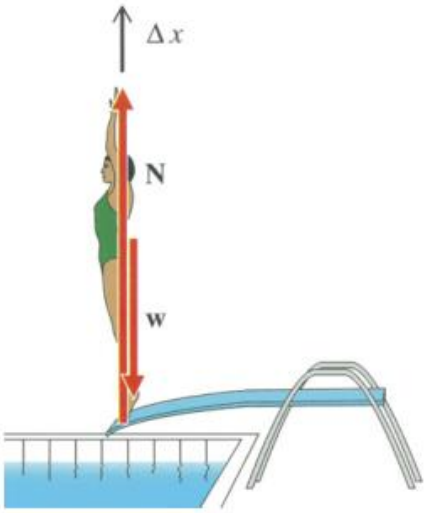
More generally, if the object does not start from rest ($u \neq 0$), and is acted by several forces, then the net work done on the object (i.e. the work done by the net force),

$$W_{net} = mas = m \times \frac{1}{2}\left(v^2 - u^2\right) = E_{k\ final} - E_{k\ initial}$$

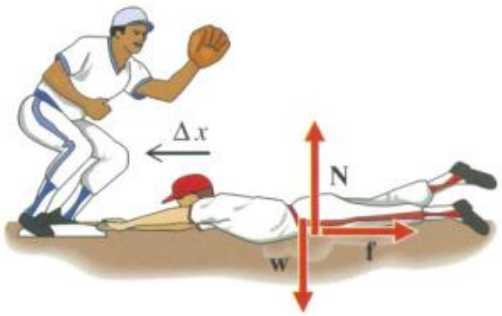
$$\therefore W_{net} = \Delta E_k$$

When there is no net work done on an object (i.e. work done by net force is zero), the object's change in kinetic energy is zero. Referring back to Concept Check 1 (p.5), no work is done on the load, so its kinetic energy remains constant. In Fig.(a), A diver's kinetic energy increases as she springs upward because positive work is done on her, since normal contact force **N** is greater than her weight **w**. In Fig. (b), a baseball player's kinetic energy decreases as he slides because friction does negative work on him.

Work-energy theorem is valid even for varying net force. One useful example is when an object at rest is pushed from one position to another position, back to rest, we conclude that the net work done is zero since there is no change in kinetic energy in the process.



(a) $\Delta K = W_{net} > 0$



(b) $\Delta K = W_{net} < 0$

5.3 Potential energy

We know that two bodies will tend to attract each other due to gravity. This means that if we want to separate the two bodies, we need to exert forces on them. The energy we use to do this is stored in the two-body system as gravitational potential energy. If we now release the two bodies, they will begin to move towards each other. The gain in the kinetic energies of the two bodies comes from the gravitational potential energy that was stored earlier.

Similarly, we know that two positively charged bodies tend to repel each other due to electrostatic forces. We can try to force them together, and in so doing, we need to exert forces on them and the energy we expend is stored in the form of electrical potential energy of the two-body system. If we now release the two bodies, they will repel each other and move apart at increasing speed. The gain in their kinetic energies comes from the electrical potential energy that is stored earlier.

The same principle applies to a spring. When loading a spring gun, we expend energy in compressing the spring. This energy is stored in the form of elastic potential energy and when fired, the spring exerts a force on the bullet and the work done by this force, which is the gain in the bullet's kinetic energy, came from the stored elastic potential energy.
In all these examples, the amount of stored potential energy in the system came from the work done by an external agent who set up the system in the first place.

There are three kinds of potential energy (PE) you will commonly encounter. They are summarized in the table below.

Elastic PE	Gravitational PE (GPE)	Electrical PE
The amount of elastic PE stored in a spring depends on the extent to which the spring is stretched or compressed.	The amount of GPE stored in a system depends on the relative position of the masses in the system.	The amount of EPE stored in a system depends on the relative position of electrical charges in the system.
When a spring is stretched, attractive forces act to pull the parts of the spring together and when the spring is compressed, repulsive forces act to push the parts of the spring apart. Elastic PE increases when the spring is stretched or compressed.	Attractive forces act between any two masses. In a two-mass system, GPE increases when two masses are separated.	Attractive forces act on two bodies of opposite charges and repulsive forces act on two bodies of like charges. In a two charged body system, EPE increases when two opposite charges are separated or when two like charges are brought closer together

5.3.1 Gravitational Potential Energy (GPE), E_p or U

Derivation of $E_p = mgh$

Consider a person lifting a box of mass m at constant speed to a height h.

Since the box is moving at constant speed, there is no change in its kinetic energy. Net work done on the box must be zero. In the process, however, the person exerting a constant upward force on the box does positive work on the box while the gravitational force acting downwards does negative work. The work done by the person on the box is stored in the form of gravitational potential energy (GPE), E_p.

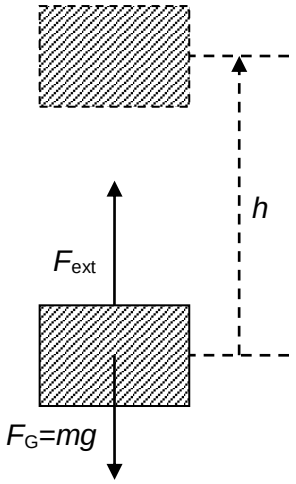
Since velocity is constant,

there is no acceleration $F_{net} = ma = 0$, thus $F_{ext} = F_G = mg$.

We can form an equation as follows.

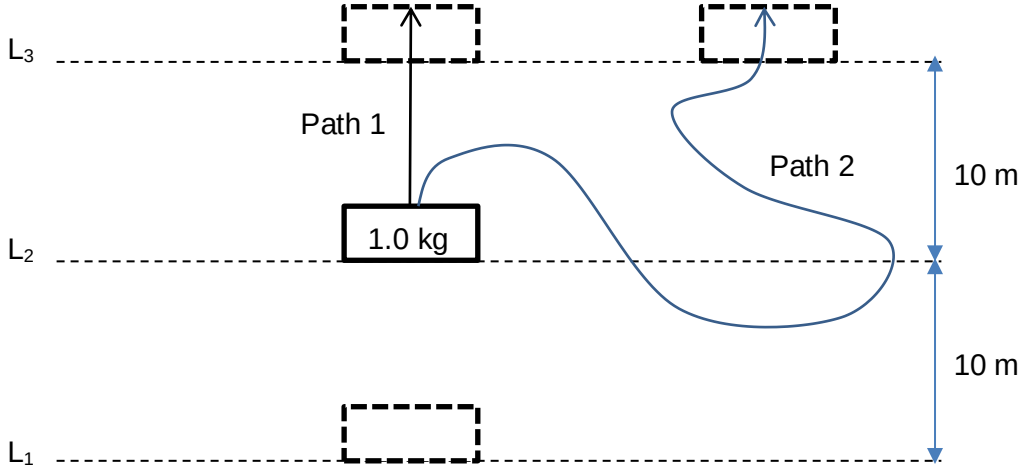
Increase in E_p = Work done by external force on the box by person, W
$= F_{ext}\, h$
$= mgh$

Hence, Increase in $E_p = mgh$



Note:

- This derivation applies only to objects that stay near to the surface of the Earth during the motion, and h is small relative to the diameter of the Earth, so that the gravitational acceleration g on the body is practically constant during the motion. (
- The potential energy is associated with the gravitational force between the Earth and the mass m. Hence the gravitational potential energy E_p belongs to the mass-Earth system and not simply to the mass m alone. It is a property of the system as a whole. In general, a system is one or more objects that we choose to study.
- E_p is a relative and not an absolute quantity. It is always taken in relation to a reference level which can be chosen arbitrarily. What is more important is the change in the E_p with reference to this level (as shown in the figure below). Consider the change in potential energy ΔE_p of a 1.0 kg mass along path 1 and 2.



Choosing L_2 as the reference level, ΔE_p = (1.0)(9.81)(10) - (1.0)(9.81)(0) = 98.1 J
Choosing L_1 as the reference level, ΔE_p = (1.0)(9.81)(20) - (1.0)(9.81)(10) = 98.1 J

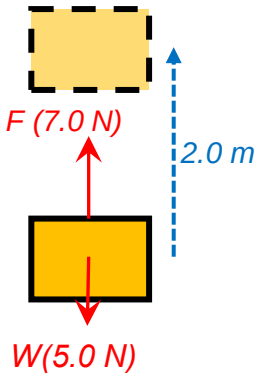
- The *work done on the mass* in moving it between two positions *is dependent only on the initial and final positions* and *independent of the path taken by the mass*.

  ΔE_p remains the same for path 1 and 2. When the starting and ending points are the same, the total work is zero, and there is no change in the gravitational potential energy.

Example 2

A mass of weight 5.0 N is lifted by a vertical force of 7.0 N through a vertical distance of 2.0 m.
(a) Calculate the work done by the lifting force.
(b) Calculate the work done by the gravitational force.
(c) Calculate the change in gravitational potential energy.
(d) Calculate the total work done on the mass.
(e) Assuming it was initially at rest. calculate the final speed of the mass.

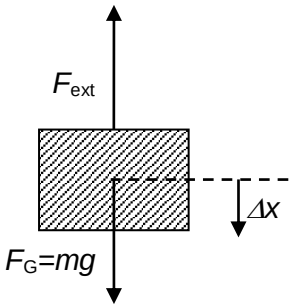
[+14 J, -10J, +10J, 4J, 4.0 m s^{-1}]



5.3.2 Relationship between Gravitational Force and Gravitational Potential Energy

Consider the same box on p.10, we define the change in GPE as the work done by the external force on the box. The change in GPE can also be expressed as *the negative of the work done on the mass by gravitational force*.

$$\Delta E_p = -W_G$$



Consider a small downward displacement, Δx, at constant velocity so that $F_{ext} = F_G$. the work done by the gravitational force, $W_G = F_G \Delta x$ is positive.
GPE decreases in the process.

Thus $\Delta E_p = -W_G = -F_G \Delta x \Rightarrow F_G = -\dfrac{\Delta E_p}{\Delta x}$

When we take the limit as $\Delta x \to 0$, $\boxed{F_G = -\dfrac{dE_p}{dx}}$

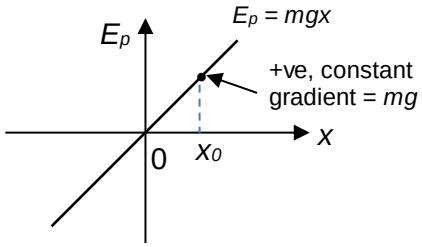
***the gravitational force is the negative of the gradient of the potential energy - position graph*.**

The magnitude of gradient gives the magnitude of the gravitational force, while the negative sign suggests that the gravitational force always acts in the direction towards lower potential energy.

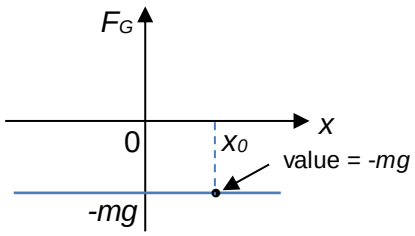
Let us now apply the above relationship in *the uniform gravitational field* near the surface of Earth, where gravitational acceleration is *g*.

Consider a mass *m* at a height x_0 from the ground, taking GPE = 0 J at the ground level.

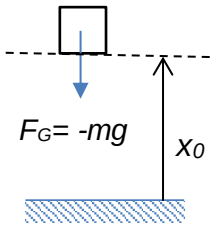
Since GPE, $E_p = mgx$, the E_p - x graph is a straight line graph passing through the origin, with gradient mg.



GPE decreases as x decreases.
Gravitational force F_G at the point x_0 is -ve of the gradient of E_P – x graph, implying F_G is a constant value acting in the direction of decreasing GPE.



For all x, F_G = -mg (constant)



At x_0, F_G = -mg (constant) is directed towards decreasing potential energy.

The gravitational force at x_0 is the negative of the gradient of E_p -x graph, or $F_G = -\dfrac{dE_p}{dx} = -mg$.

In general, the force-potential energy relationship is applicable in any conservative[1] system.

Conservative force, $F = -\dfrac{dE_p}{dx}$ is the negative of the gradient of the E_p -x graph.

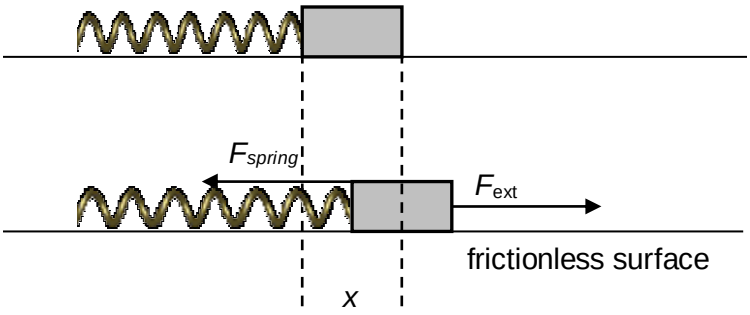
We will re-visit this relationship when we study the chapters on Gravitational Field and Electric Field later. In those chapters, potential energy is often denoted by the symbol *U.* You will encounter situations in which you are given a potential energy - distance graph and you can use this general relationship to find the corresponding force.

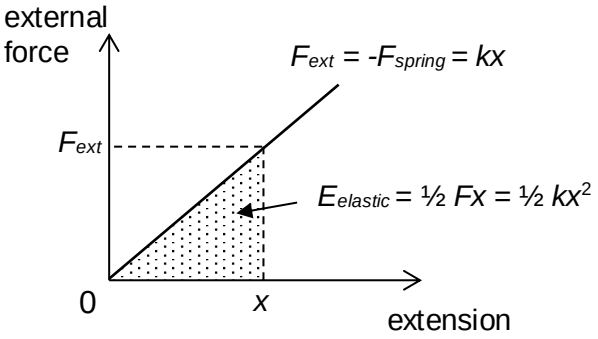
[1] See Appendix 1 for more details regarding conservative system and conservative force

5.3.3 Elastic Potential Energy

In order to stretch a spring, we must exert an increasingly larger force to pull both ends of the spring apart as we stretch the spring. This is because the force we exert must at least be equal to if not larger than the resistive spring force that the rest of the spring exerts on the ends of the spring. Indeed, the resistive spring force gets stronger the more the spring is extended or compressed.

In fact, in most situations the resistive spring force needed to stretch the spring **increases in proportion to the extension**. Whenever this is true, we say that **the spring obeys Hooke's law, i.e. the spring force F_{spring} is proportional to the spring's extension x,** i.e. $F_{spring} = -kx$ where k is the spring constant. The negative sign appears because F_{spring} is in the direction opposite to the displacement x. A stiffer spring that requires a larger force to stretch it has a larger spring constant. A stiffer spring also means that more elastic potential energy is stored per unit of extension.





Let a light spring be stretched through an extension x, very slowly (at practically zero velocity) by applying an external force F_{ext}.

To just overcome the spring force, the external force $F_{ext} = -F_{spring}$.

$$W = \int_0^x F_{ext}\, dx$$ = Area under the *force–extension graph*

$= ½ F x$

$= ½ (k x) (x)$

$E_{elastic} = ½ kx^2$

Hence, elastic potential energy $E_{elastic} = \frac{1}{2}Fx = \frac{1}{2}kx^2$

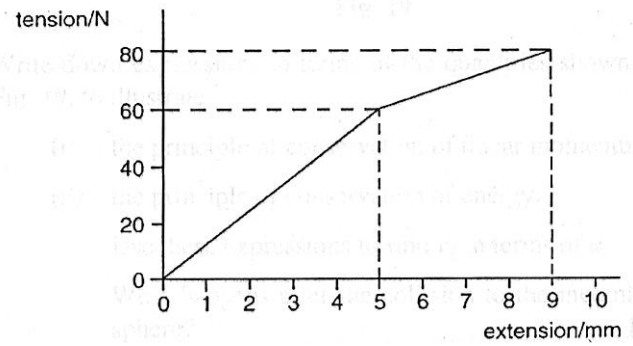
Note, the spring force is a conservative force.

$F_{spring} = -\dfrac{dE_{elastic}}{dx}$ is applicable.

Verifying: $F_{spring} = -\dfrac{d}{dx}\left(\dfrac{1}{2}kx^2\right) = -kx$

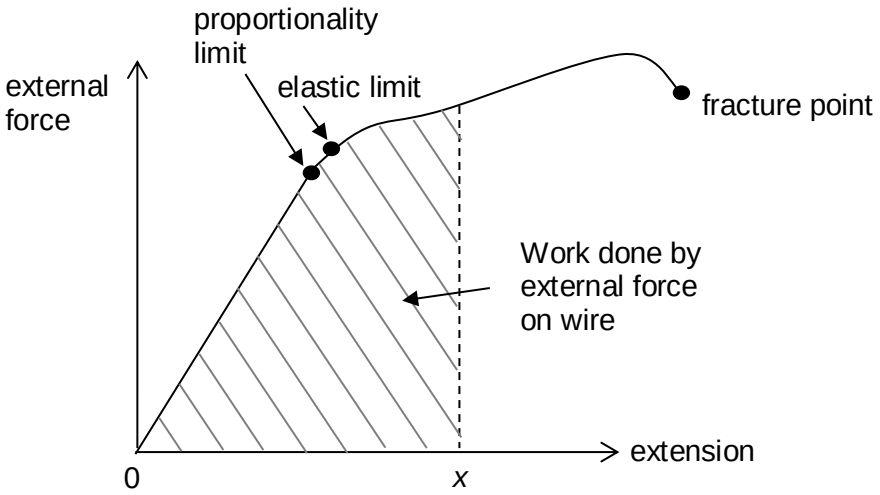
Example 3 (mod J99/I/23)

A sample is placed in a tensile testing machine. It is extended by known amounts and the tension is measured. (i) What is the work done in stretching the sample to the first point where it no longer obeys Hooke's Law. (ii) What is the work done on the sample when it is given a total extension of 9.0 mm?
[0.15 J, 0.43 J]



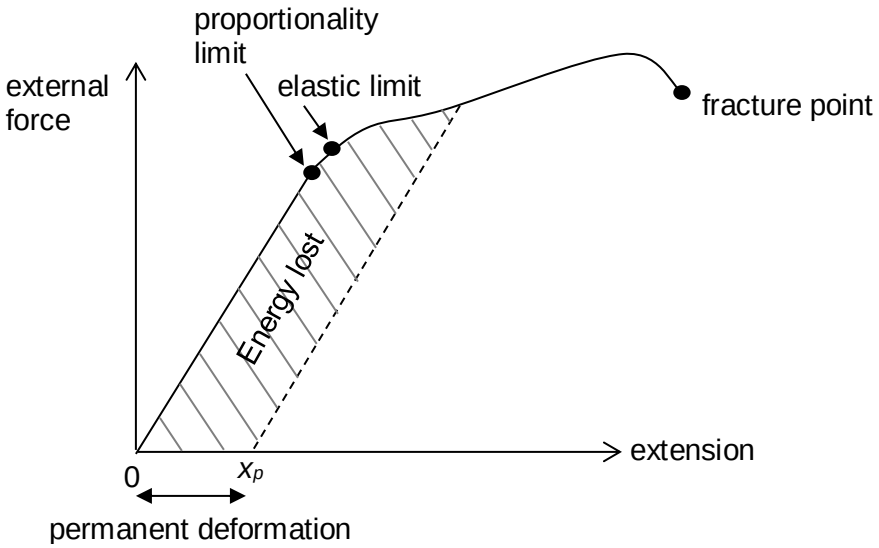
5.3.4 Work done by External Force on a Deformed Material

An elastic material (e.g. a wire) usually obey Hooke's Law closely unless the deformation is excessively large. If the material is stretched beyond its proportionality limit, it will no longer obey Hooke's Law and the force-extension graph will not be a straight line from that point onwards. If an elastic material (e.g. a wire) is stretched further **beyond its elastic limit** (also known as yield point), it will **suffer a permanent deformation** and not snap back to its original length when the external force is removed. If the wire is stretched even further, it will eventually hit fracture point and the wire will simply snap. The force-extension graph will look like the one shown in the figure below.



In such cases, to determine work done by external force on wire, one <u>cannot</u> simply use the equation $E = ½\ kx^2$ since Hooke's Law is no longer obeyed. Instead, the work done can simply be worked out by estimating the area under the force-extension graph (shaded area).

When the wire is **stretched beyond its elastic limit** and the external force is subsequently removed, the extension is only partially recovered. Work done by external force on wire will be more than the elastic potential energy. Energy is therefore lost as heat.



<u>Try this out !</u> The effect of hysteresis is usually very small for metals, but is noticeable for polythene, glass and rubber. You can easily investigate this using a rubber band. By simply stretching it and then holding it against your lips you can detect a rise in temperature!

5.4 Principle of Conservation[2] of Energy

When solving problems via the energy approach, remember that this approach is based on the principle of conservation of energy:

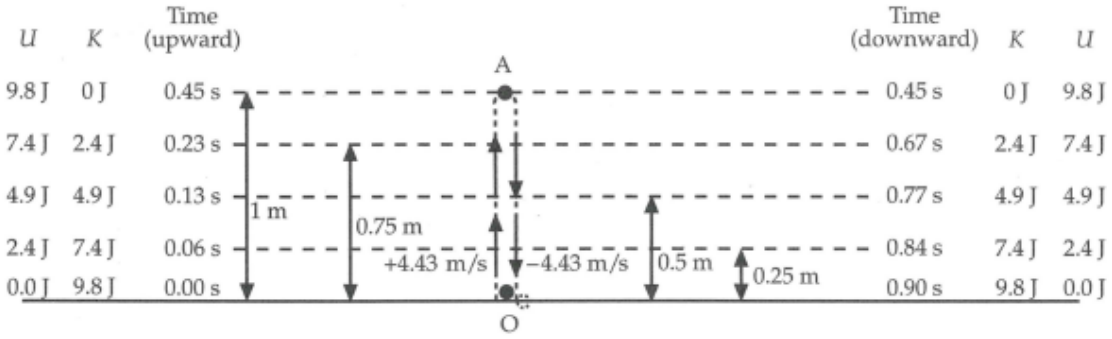
Energy can be **converted from one form to another** but **cannot be created or destroyed**.

In other words, the total energy is neither increased nor decreased in any process.

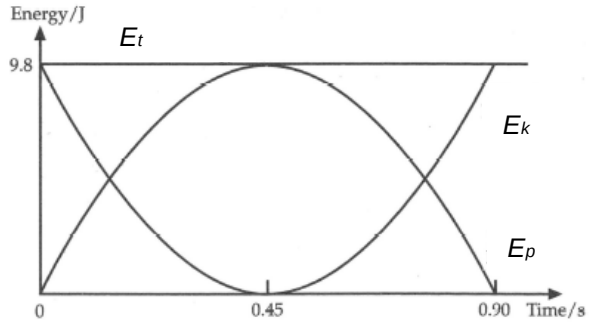
- **When the only force that does work is gravitational force in the system, the total mechanical energy *E* (i.e. sum of kinetic energy and potential energy) is constant.**

  Mathematically, $(E_k + E_p)_{initial} = (E_k + E_p)_{final}$ or $\Delta E_k + \Delta E_p = 0$

  Example: Consider the motion of a 1.0 kg mass projected vertically upwards from the ground with an initial speed of 4.43 m s^{-1}. Given that air resistance is negligible, the kinetic energy of the mass is converted to gravitational potential energy as it rises from 0 to A. As it falls back to 0, potential energy is converted into kinetic energy.



  A graph of energy against time is as shown. At any point of time, the total mechanical energy E_t, that is the sum of E_k and E_p, is the same.



- **When there is presence of external forces doing work in the system, change in mechanical energy = net work done by external forces, W_{ext}.**

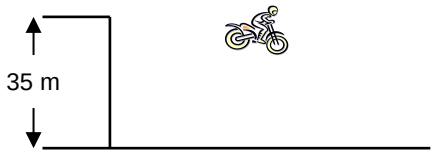
  Mathematically, $(E_k + E_p)_{initial} + W_{ext} = (E_k + E_p)_{final}$ or $\Delta E_k + \Delta E_p = W_{ext}$

Note that an external force can be doing positive work, adding to the total mechanical energy, or negative work, reducing the total mechanical energy of the system.

[2] A quantity that always has the same value is called a conserved quantity.

Example 4 (no external force, mechanical energy is conserved)

(a) A motorcyclist drives horizontally off a cliff at a speed of 38.0 m s^{-1}. Ignoring air resistance, find the speed of the motorcycle just before it reaches the ground.
[46.2 m s^{-1}]



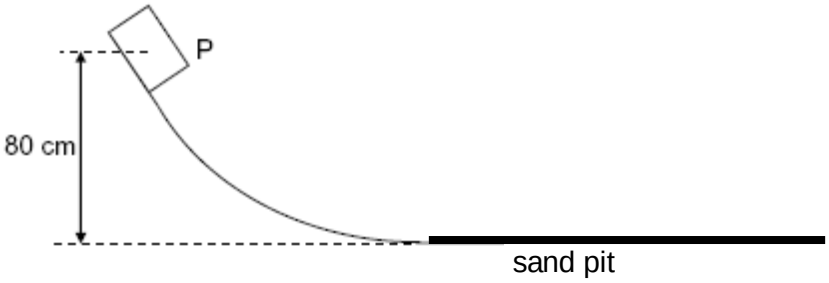
(b) Would your answer in (a) change if the motorcyclist leaps across at the same 35 m height with an initial upward angle while maintaining the same initial speed of 38.0 m s^{-1}?

Example 5 (external force present, mechanical energy is lost)

A car of mass 800 kg moving at 30 km h^{-1} along a horizontal road is brought to rest by a constant retarding force of 5000 N. Calculate the distance the car moves before coming to rest. [5.6 m]

Example 6 (external force present, mechanical energy is lost)

A block of mass 4.0 kg is released at point P on a curved frictionless track as shown. The block slides down a vertical distance of 80 cm and enters a sand pit. Determine the distance the block can move along the sand pit before coming to rest if the friction acting on the block is 85 N. [0.37 m]



5.5 Power

The same amount of work is done in raising a given body through a given height, regardless whether it takes one second or one year to do so. However*, the rate at which work is done* is greater if less time is taken.

Power is defined as the rate at which work is done.

$$\text{Instantaneous power, } P = \frac{dW}{dt}$$

$$\text{Average power, } \langle P \rangle = \frac{\Delta W}{\Delta t}$$

Notes:

- The SI unit of power is the **watt** (W). Power is a **scalar** quantity.
- Work can also be expressed in units of (power x time). This is the origin of the term *kilowatt-hour*. One *kilo-watt hour* is the work done in 1 hour by an agent working at a constant rate of 1 kW.

Example 7 (Average Power)

A 1.10×10^3 kg car, starting from rest, accelerates for 5.00 s. The magnitude of the acceleration is 4.60 m s^{-2}. Determine the average power generated by the net force that accelerates the vehicle. [5.82×10^4 W]

5.5.1 Relationship between Power, Force and Velocity

Consider a force F that acts on a body for a small time interval Δt. The body moves a small displacement Δx in the direction of the force.

Work done by the force F during Δt, $\quad \Delta W = F \Delta x$

Power delivered by that force F during the time interval Δt

$$P = \Delta W / \Delta t$$
$$= (F \Delta x) / \Delta t$$
$$= F(\Delta x/\Delta t)$$
$$= Fv$$

where v is the instantaneous velocity of the body.

When a force F acts on a body that is moving with velocity v, in the direction of the force, work is done on the body at the rate given by $P = Fv$

Notes:

- A force F that acts in the direction of the displacement Δx, does positive work on the body. It transfers energy to the body at a rate proportional to the speed of the body.
- If F or v varies with time, then P may not be constant. However, the product Fv at any moment will gives the instantaneous power of F at that instant.
- Where F acts against the motion of the body, it does negative work on the body. It extracts energy out of the body at a rate equal to the magnitude of Fv.

Example 8

A 1000 kg car experiences a total resistive force of 2000 N moving with a constant velocity of 20.0 m s^{-1} along a straight horizontal road.
What is the power developed by the engine of the car when it travels at 20.0 m s^{-1}
(a) horizontally, and
(b) up a slope with gradient of 5°?
[40 kW; 57.1 kW]

5.6 Efficiency (of energy conversion), η

No practical engine can transform energy from one form to another without some energy going to internal energy and other non-useful forms of energy.

The efficiency of any energy conversion process is the ratio of

- the useful energy output to the total energy input, or
- the useful power output to the power input

$$\text{efficiency} = \frac{\text{useful energy output}}{\text{total energy input}} \quad , \quad \eta = \frac{E_{output}}{E_{input}} = \frac{P_{output}}{P_{input}}$$

Note:
η (Greek small letter, Eta) is a ratio between two quantities of the same type. It has no unit.

Example 9
A small motor is used to raise a weight of 2.0 N at constant speed through a vertical height of 80 cm in 4.0 s. The efficiency of the motor is 20 %. What is the electrical power supplied to the motor? [2.0 W]

Appendix 1 Potential Energy (*U*) and Conservative Force (Good to know)

From the derivation on p. 10, the ***gain in gravitational potential energy*****, Δ*U*** can be taken as:

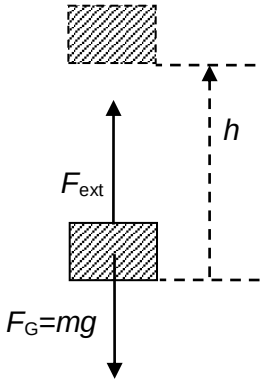
the positive work done on the mass by an external force when the mass moves (a distance *h*) to higher position.

$\Delta U = F_{ext}\, h$

or:

the negative of the negative work done on the mass by gravitational force when the mass moves to a higher position.

$\Delta U = -(-F_G\, h)$



The above is applicable only to a conservative system, or a system that involves a conservative force, such as gravitational force, electric force and spring force. An essential feature of conservative forces is that their work is always reversible. Another important aspect of conservative forces is that a body may move from point 1 to point 2 by various paths, but the work done by a conservative force is the same for all of these paths. Thus, if a body stays close to the surface of the earth, the gravitational force *mg* is independent of height, and the work done by this force depends only on the change in height. If the body moves around a closed path, ending at the same point where it started, the total work done by the gravitational force is always zero.

The work done by a conservative force always has four properties:

1. It can be expressed as the difference between the initial and final values of a potential-energy function.
2. It is reversible.
3. It is independent of the path of the body and depends only on the starting and ending points.
4. When the starting and ending points are the same, the total work is zero.

When the only forces that do work are conservative forces, the total mechanical energy $E = E_k + E_p$ is constant.

Not all forces are conservative. Consider the friction force acting on the crate sliding on a ramp. When the body slides up and then back down to the starting point, the total work done on it by the friction force is not zero. When the direction of motion reverses, so does the friction force, and friction does negative work in both directions. The lost kinetic energy cannot be recovered by reversing the motion or in any other way, and mechanical energy is not conserved. There is no potential-energy function for the friction force.

Friction is a non-conservative force that acts to reduce the mechanical energy in a system.

However, non-conservative forces do not always reduce the mechanical energy; a non-conservative force changes the mechanical energy, so a force that increases the total mechanical energy, like the force provided by a motor or engine, is also a non-conservative force.

In general,

When there are only conservative forces,

Change in total mechanical energy of the system = 0

When there are presence of conservative forces and non-conservative forces,

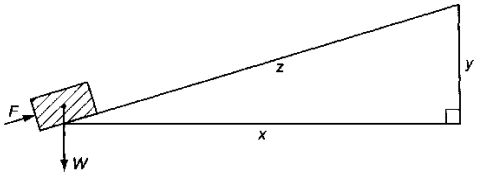
Change in total mechanical energy of the system

= work done on the system by all non-conservative forces

Topic 5 Work Energy Power
Self Review Questions

Work and Energy

S1 **[H12007P1Q14]** A mass of weight *W* rests on a rough slope of dimensions *x*, *y* and *z* as shown.



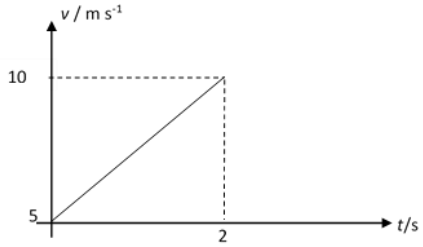
A force *F* pushes the mass all the way up the slope. The mass starts and finishes at rest.

How much work is done by the force *F*?

A *Wy* **C** *Wz*

B *Wy*+*Fz* **D** *Fz*

S2 **[NYJC 2011 Prelim P1 Q10]** The graph shows the variation with time of the velocity of a body when it is acted on by a force on a smooth horizontal surface.



If mass of the body is 2.0 kg, the work done by the force on the body is

A 25 J **C** 75 J

B 50 J **D** 100 J

S3 **[H2 2011 P1Q10]** A wire is stretched elastically by a force of 200 N, causing an extension of 4.00 mm. The force is then steadily increased to 250 N. The wire still behaves elastically.

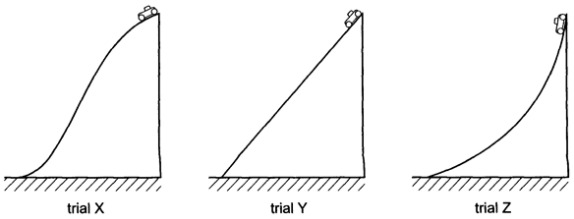
How much extra work is done in producing the additional extension?

A 0.225 J **C** 225 J

B 0.250 J **D** 250 J

Conservation of Energy

S4 **[H1 2007P1Q15]** A toy car is released from rest at the top of a length of flexible track. The track is shaped differently for three trials X, Y and Z.



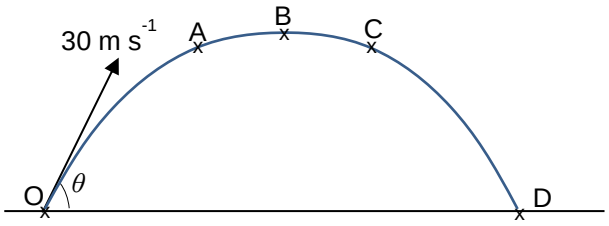
In each trial, the top of the track is the same vertical height above the ground. The speeds v_X, v_Y and v_Z at the end of the track are measured. Friction is negligible.

Which of the following is correct?

A $v_X, = v_Y = v_Z$ **C** $v_Y > v_Z > v_X$

B $v_X, > v_Y > v_Z$ **D** $v_Z > v_Y > v_X,$

S5 On a horizontal plane, a 2.0 kg particle is fired from point O at an angle of 60° above the horizontal with an initial speed of 30 m s^{-1}. When the particle is at point B, it is at its maximum height from the ground. Assuming that the effect of air resistance is negligible, at which point on the particle's trajectory is its kinetic energy minimum?



S6 **[H1/2012/1/15]** A car of mass 1000 kg travels up a hill at constant speed. It gains 87 m in height travelling a distance of 500 m. The total frictional force is 580 N. What is the work done by the driving force?

A 0.85 MJ **C** 4.90 MJ

B 1.14 MJ **D** 5.19 MJ

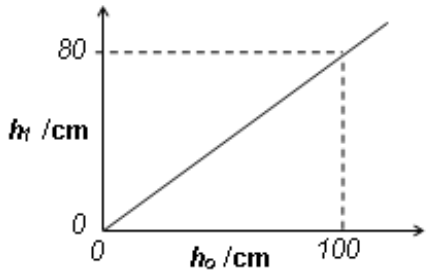
S7 **[J77/2/30]** 500 kJ of heat was produced when a vehicle of total mass 1600 kg was brought to a rest on a level road. The speed of the vehicle just before the brakes were applied was

A 0.625 m s^{-1} **D** 62.5 m s^{-1}

B 0.79 m s^{-1} **E** 625 m s^{-1}

C 25 m s^{-1}

S8 **[N81/P1/Q2]** A ball released from a height h_o above a horizontal surface rebounds to a height h_1 after one bounce. The graph that relates h_o to h_1 is shown.



If the ball (of mass m) was dropped from an initial height h and made three bounces, the kinetic energy of the ball immediately after the third impact with the surface was

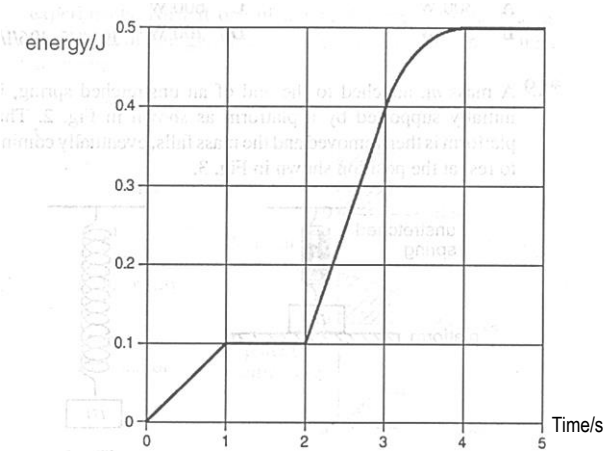
A $(0.8)^3\ mgh$ **D** $[1 - (3 \times 0.2)\ mgh]$

B $(0.8)^2\ mgh$ **E** $[1 - (0.8)^3\ mgh]$

C $0.8\ mg\ (h/3)$

Power

S9 **[J94/1/6]** A bicycle dynamo is started at time t = 0 s. The total energy transformed by the dynamo during the first 5 seconds increases as shown in the graph.



What is the maximum power generated at any instant during these first 5 seconds?

A 0.10 W **C** 0.30 W

B 0.13 W **D** 0.50 W

S10 **[NYJC/2007/Prelim/P1/Q10]** To travel at constant speed, a car engine provides 25 kW of useful power. The driving force on the car is 600 N. At what speed does it travel?

A 2.5 m s^{-1} **C** 25 m s^{-1}

B 4.2 m s^{-1} **D** 42 m s^{-1}

S11 **[H1/N2009/1/12]** The drag force acting on a car moving at velocity v through still air is proportional to v^2. When the car is travelling at 20 m s^{-1} on a level road, the power required to overcome the drag force is 4800 W.

What power is required when the car travels at 25 m s^{-1}?

A 6000 W **C** 8000 W

B 7500 W **D** 9400 W

S12 **[J93/1/6]** A small metal sphere of mass m is moving through a viscous liquid. When it reaches a constant downward velocity v, which of the following describes how the kinetic energy and gravitational potential energy of the sphere changes with time?

A K.E.: constant and equal to $\frac{1}{2}\ mv^2$
G.P.E.: decreases at a rate of mgv

B K.E.: constant and equal to $\frac{1}{2}\ mv^2$
G.P.E.: decreases at a rate of $(mgv - \frac{1}{2}\ mv^2)$

C K.E.: constant and equal to $\frac{1}{2}\ mv^2$
G.P.E.: decreases at a rate of $(\frac{1}{2}\ mv^2 - mgv)$

D K.E.: increases at a rate of mgv
G.P.E.: decreases at a rate of mgv

E K.E.: increases at a rate of mgv
G.P.E.: decreases at a rate of $(\frac{1}{2}\ mv^2 - mgv)$

Discussion Questions

Work and Energy

D1 **[2008/P3/Q1(part)]** Fig 6.1 shows the variation with displacement d of the force *F* applied to an object. *F* and *d* are in the same direction. On Fig 6.2, draw a graph showing the variation with *d* of the work done.

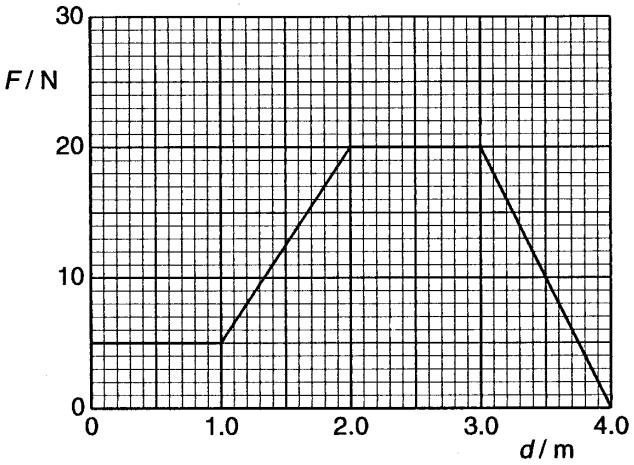


Fig. 6.1

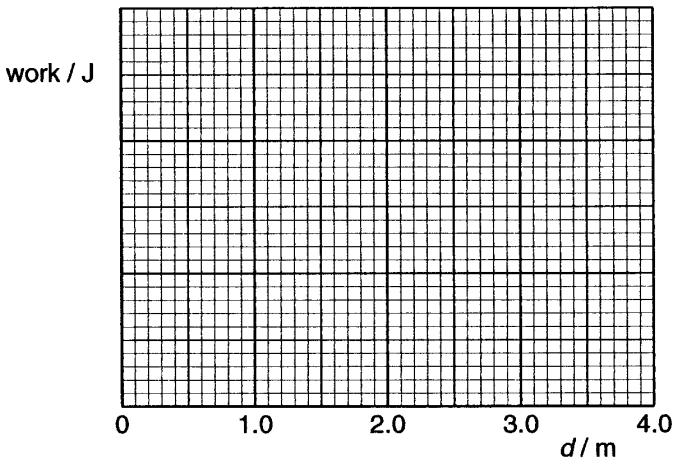
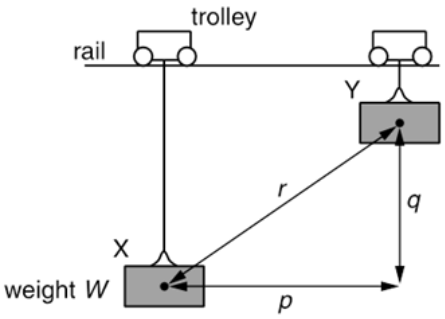


Fig. 6.2

D2 **[CIE June 2003]** A weight *W* hangs from a trolley that runs along a rail. The trolley moves horizontally through a distance *p* and simultaneously raises the weight through height *q*.

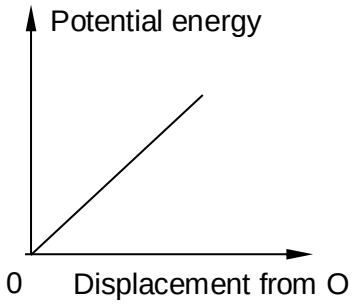
As a result, the weight moves through a distance *r* from X to Y. It starts and finishes at rest. How much work is done on the weight by the trolley during this process?



A *Wp** **B** *W(p+q)* **C** *Wq* **D** *Wr*

D3 The graph shows the way the potential energy of a body varies with its displacement from a point O.

Which feature of the graph means that the force on the body is directed towards O? Explain your answer.



A The graph is linear.

B The graph passes the origin.

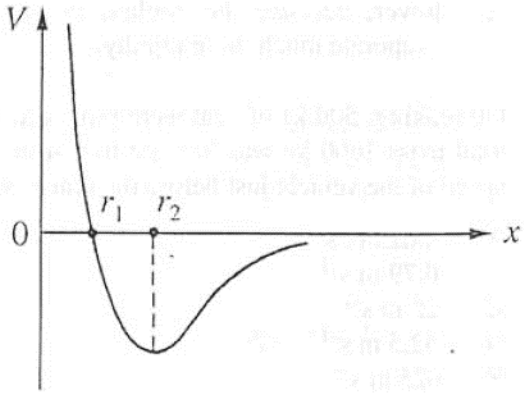
C The potential energy increases as the body moves further from O.

D The value of the potential energy is always positive.

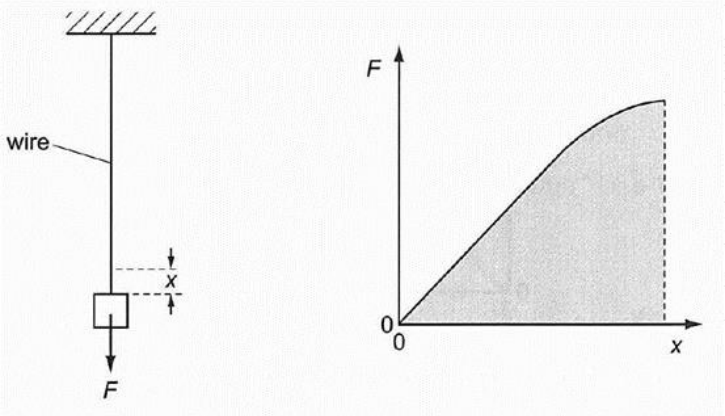
D4 [J8911/6] The potential energy V of two molecules separated by distance x is shown in the diagram.

Which of the following correctly describes the force between the molecules?

	attractive for	repulsive for
A	$x < r_1$	$x > r_1$
B	$x > r_1$	$x < r_1$
C	$x < r_1$	$x > r_2$
D	$x < r_2$	$x > r_2$
E	$x > r_2$	$x < r_2$



D5 **[H1 2009 P1 Q9 & 2015 P1 Q11]**) A wire, fixed at its upper end, is subjected to an increasing load *F* by increasing the mass attached to its lower end. A graph of *F* against the extension *x* of the wire is shown.



The wire is stretched so that it is permanently deformed. What does the shaded area on the graph represent?

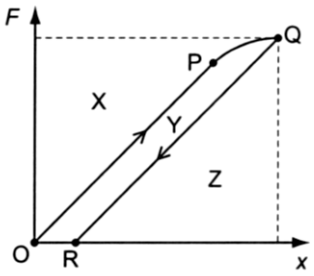
A the amount of elastic potential energy stored in the wire

B the amount of heat produced in the wire

C the loss of gravitational potential energy of the mass

D the work done by *F* on the wire

D6 **[H1 2010 P1 Q11]** A metal wire is stretched by a varying force F, causing its extension x to increase as shown by the line OPQ on the graph. The force is then gradually reduced to zero and the relation between the force and extension is indicated by the line QR.

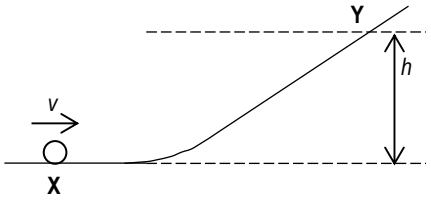


What area represents the elastic potential energy in the wire at Q?

A X **B** Y **C** Z **D** Y + Z

Conservation of Energy

D7 **[J1987/P1/Q5]** An object of mass m passes point X with velocity *v* and slides up a frictionless incline to stop at point Y which is at a point *h* above X.

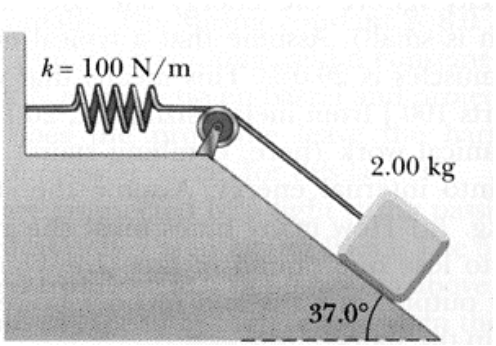


A second object of mass ½ m passes X with a velocity ½ v. To what height will it rise?

A ¼ *h* **B** ½ *h* **C** $\frac{1}{\sqrt{2}}h$ **D** *h* **E** $h\sqrt{2}$

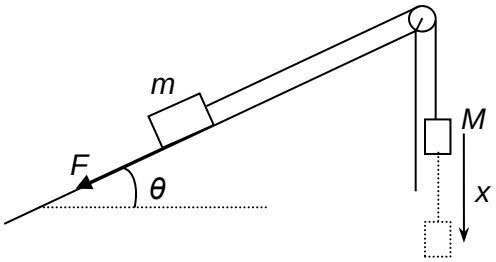
D8 **[College Physics by Serway & Faughn]**

A 2.00 kg block situated on a rough incline is connected to a spring of negligible mass having a spring constant of 100 N m^{-1}. The block is released from rest when the spring is un-stretched and the pulley is frictionless. The block moves 20.0 cm down the incline before coming to rest. Ignore air resistance.



Find the average resistive force due to friction between the block and the incline. [3]

D9 **[J88/1/6 modified]** A mass *m* moves on a rough plane inclined at an angle θ to the horizontal. It experiences a constant frictional force *F*. Mass *M* is attached to it by a light inelastic cord running over a smooth pulley. During motion, the mass *M* is allowed to fall a vertical distance *x*, causing *m* to move up the plane as shown in the diagram.



(a) How much heat is generated by the friction in this process?

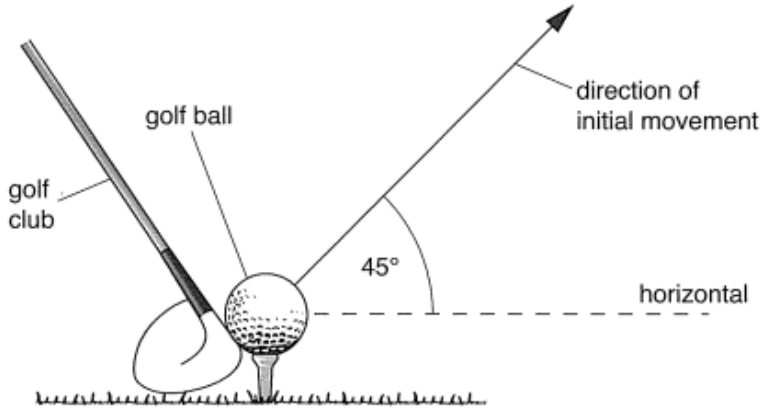
A *Fx* **C** *Mgx* sinθ **E** *Mgx* sinθ + *Fx*

B *mgx* **D** *Mgx* sinθ - *Fx*

(b) Determine the expression for the total kinetic energy gained by the two masses. Express your answers in terms of *m, M, F, g, x and* θ.

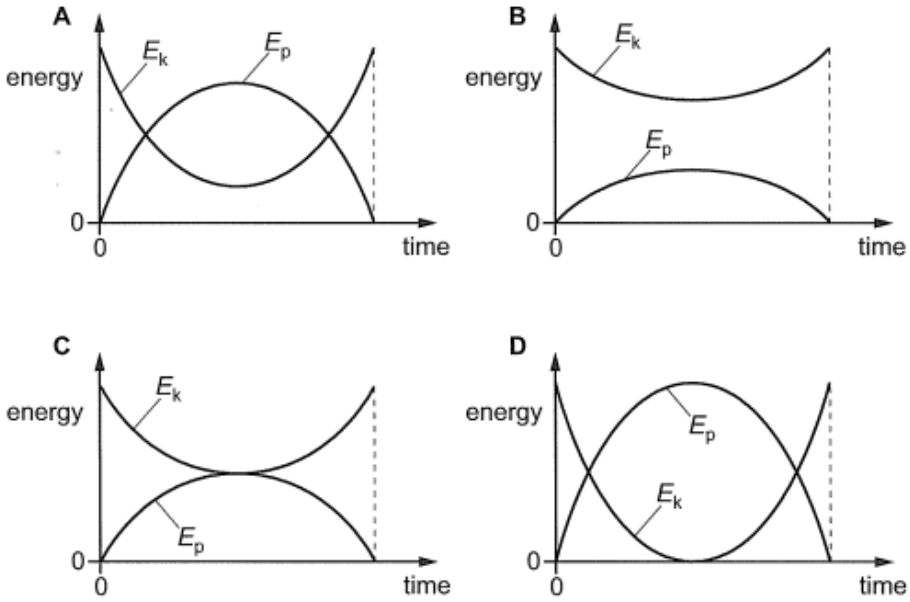
D10 [2016 Specimen P1Q6]

A golf ball, on level ground, is hit and starts to move at an angle of 45° to the horizontal.



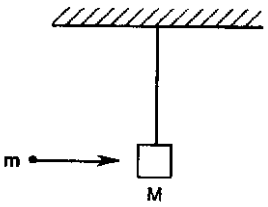
Which graph shows how the kinetic energy E_k and the gravitational potential energy E_p of the ball vary from the time of impact until just before it reaches the ground?

Ignore any effects of air resistance.



D11 [HCI Promo 1999/1/6]

A bullet of mass m = 10 g is fired into a ballistic pendulum of mass M = 2.0 kg as shown. The bullet remains in the block after the collision and the whole system rises to a maximum height of 20 cm.



Find the initial speed of the bullet in m s^{-1}.

A 2.0 **B** 28.0 **C** 398 **D** 719

D12 **[N2012/1/10, modified]** The top end of a spring is attached to a fixed point and a mass of 4.2 kg is attached to its lower end. The mass is released and after bouncing up and down several times it comes to rest at a distance 0.29 m below its starting point. Which row gives the gain in the gravitational potential energy of the mass E_p and the gain in the elastic potential energy of the spring E_s?

	E_p / J	E_s / J
A	–12	+12
B	–12	+6
C	+12	+12
D	+12	+6

D13 **[H1 2009 P2 Q7]** Fig 7.1 shows a man doing a bungee jump. The man has a mass of 75 kg and falls a distance of 41 m before the elastic rope attached to him starts to exert any force on him.

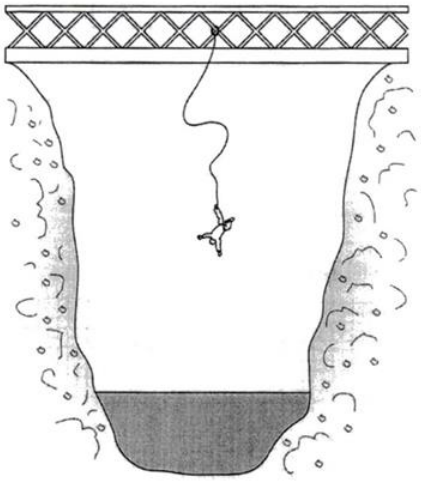


Fig. 7.1

(ai) Calculate the theoretical time taken for the fall of this distance. [2]

(ii) The actual time for this fall is 2.9 s. State the deduction you can make by comparing the actual time with your answer to (i). [1]

(b) A force-extension graph for the elastic rope used for the bungee jump is shown in Fig 7.2.

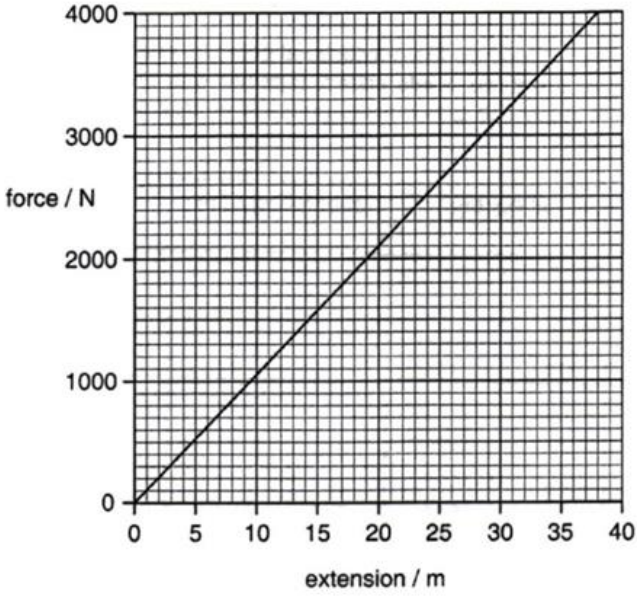


Fig. 7.2

The total distance of fall for the man before he stops for the first time is 73 m. Deduce

(i) the extension of the rope when the man stops for the first time. [1]

(ii) the elastic potential energy stored in the rope at this time. [2]

(ci) Complete the table below to show how the gravitational potential energy and the kinetic energy of the man and the elastic potential energy stored in the rope change with the distance fallen.

	at the top	after falling 41 m	after falling 73 m
GPE/J			0
EPE/J			
KE/J			

(ii) Calculate

1. how far the man has fallen from the top when he has maximum kinetic energy [2]
2. the maximum kinetic energy of the man during the fall [3]

(iii) Use your values to sketch three graphs on Figure 7.4 showing how the three different types of energy vary with distance fallen. Label each graph. [4]

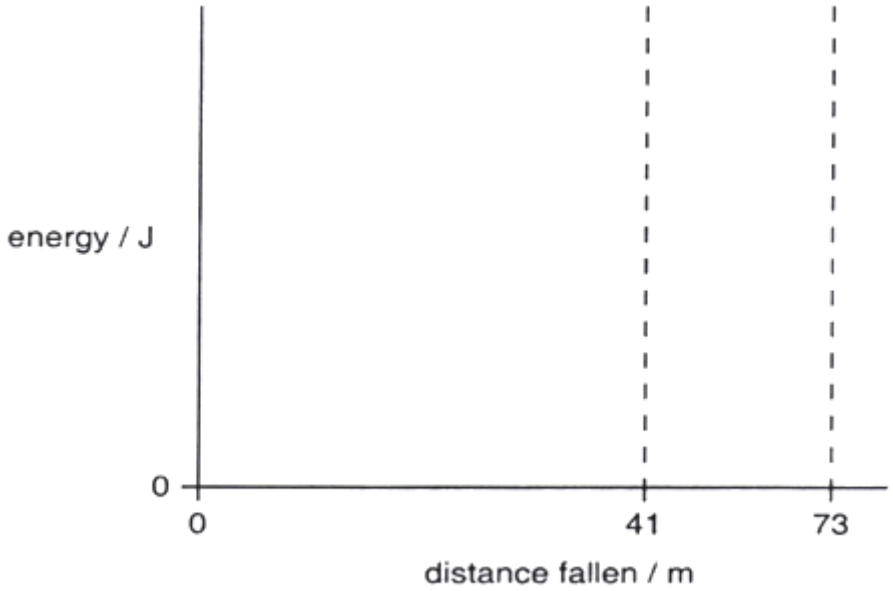
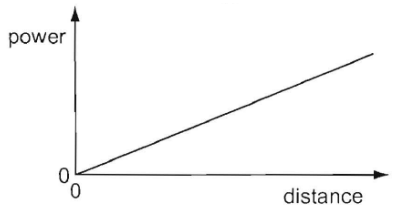


Fig. 7.4

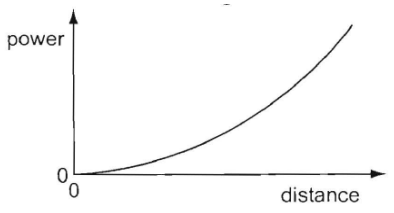
Power

D14 **[H2 2010 P1Q11]** A vehicle starts from rest and accelerates uniformly. Which graph shows how the power output of the vehicle varies with distance travelled?

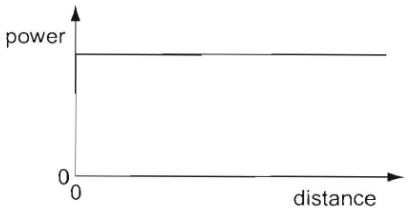
A



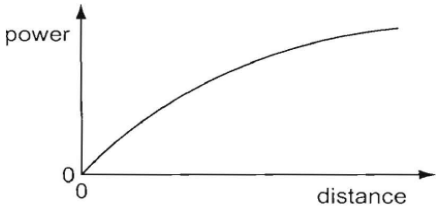
C



B



D



D15 **[2009/P1/Q12]** A driving force of 200 N is needed for a car of mass 800 kg to travel along a level road at a speed of 20 m s^{-1}. What power is required to maintain the car at this speed up a gradient in which the car rises 1 m for each 8 m of travel along the road?

A	6.0 kW	**C**	20 kW
B	7.2 kW	**D**	24 kW

D16 **[2009/P1/Q13]** A speed boat with two engines, each of power output 36 kW, can travel at a maximum speed of 12 m s^{-1}. The total drag D on the boat is proportional to the square of the speed of the boat.

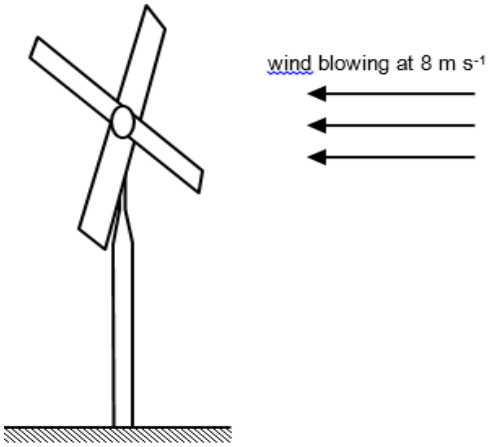
What is the maximum speed of the boat when only one engine is working?

A	3.0 m s^{-1}	**C**	8.5 m s^{-1}
B	6.0 m s^{-1}	**D**	9.5 m s^{-1}

D17 **[CIE/N2011/1/16, modified]** A man of mass *m* ties himself to one end of a rope which passes over a single fixed pulley. He pulls on the other end of the rope to lift himself up at an average speed of *v*. Which of the following expressions gives the average useful power at which he is working?

A	$\frac{1}{2}\ mgv$	**C**	$mgv + \frac{1}{2}\ mv^2$
B	mgv	**D**	$mgv - \frac{1}{2}\ mv^2$

D18 **[HCI/2007/BT2/P1/Q7]** Giant windmills have been proposed as power sources. One such windmill sweeps out a circle of 60.0 m radius. If 60% of the available energy from the wind is converted into useful energy, calculate the power generated when the wind blows at 8.00 m s^{-1} (Density of air = 1 kg m^{-3}).



A 2.90 MW **C** 1.48 MW

B 3.48 MW **D** 1.74 MW

Numerical Solutions

Self Review Questions

S1	D	**S5**	B	**S9**	C
S2	C	**S6**	B	**S10**	D
S3	A	**S7**	C	**S11**	D
S4	A	**S8**	A	**S12**	A

Discussion Questions

D8 1.81 N

D13 (a) 2.89 s
(b)(i) 32.0 m (ii) 54.4 kJ
(c) (ii) 1. 48 m 2. 32.7 kJ

Chapter 6

Circular Motion

Image credits to https://en.wikipedia.org/wiki/Singapore_Flyer#/media/File:Singapore_Flyer.JPG

The Singapore Flyer is currently one of the tallest Ferris wheel in the world. Described by its operators as an observation wheel, it reaches 42 stories high. At a total height of 165 m (541 feet) and a diameter of 150 m (492.1 feet), it was the tallest Ferris wheel at its inception, 5 m (16 feet) taller than the previous record holder, the Star of Nanchang. The Flyer was itself surpassed in 2014 by the High Roller in Las Vegas, which stands at 167.6 m (550 feet) with a diameter of 158.5 m (520 feet). The Ain Dubai in Dubai which was launched in 2021 stood at a height of 250 m (820 feet).

TOPIC 6: CIRCULAR MOTION

Circular Motion	**Learning Outcomes** Students should be able to:
Kinematics of uniform circular motion	(a) express angular displacement in radians. (b) show an understanding of and use the concept of angular velocity to solve problems. (c) recall and use $v = r\omega$ to solve problems. (d) derive, from the definitions of velocity and acceleration, equations which represent uniformly accelerated motion in a straight line.
Centripetal acceleration	(e) describe qualitatively motion in a curved path due to a perpendicular force, and understand the centripetal acceleration in the case of uniform motion in a circle. (f) recall and use centripetal acceleration $a = r\omega^2$, $a = v^2/r$ to solve problems.
Centripetal force	(g) recall and use centripetal force $F = mr\omega^2$, $F = mv^2/r$ to solve problems.

Playlist of lecture examples:
https://youtube.com/playlist?list=PL_b5cjrUKDlaAYoOvqyWUXVITtqdkKOSO

Further Readings / References

1) University Physics with Modern Physics, 14th Ed., *Young & Freedman, Chapter 9, Pg 298.*
2) Advanced Level Physics, 6th Ed., *Nelkon & Parker, Chapter 2, Pg 48.*
3) College Physics, 8th Ed., *Young & Geller, Chapter 6, Pg 161.*
4) Physics for Scientists and Engineers with Modern Physics, 6th Ed., *Serway Jewett, Chapter 6, Pg 150.*

Contents

TOPIC 6: CIRCULAR MOTION ..... **Error! Bookmark not defined.**

6.1 Introduction ……4

6.2 Kinematics of Circular Motion ……5

6.2.1 How angle is measured 5

6.2.2 Angular Displacement (θ) 5

6.2.3 Angular Velocity (ω) 6

6.2.4 Relationship between Tangential Speed v and Angular Speed ω 6

6.2.5 Period (T) and Frequency (f) 7

6.3 Uniform Circular Motion ……8

6.3.1 Derivation of the equation of centripetal acceleration..... .9

6.3.2 Centripetal Force (F_c) 10

6.4 Vertical Circular Motion ……122

6.5 Problem solving for Circular Motion ……...14

6.5.1 Further examples ……15

Tutorial 6

Self review Questions ……17

Discussion Questions ……19

Answers ……24

6.1 Introduction

Circular motion is a type of motion we often encounter in our everyday lives, - the motion of a car and its passengers when navigating a bend, the motion of the capsules of a Ferris wheel when in rotation, children on a spinning merry-go-round, and clothes being spun in a spin dryer.

On a larger scale, in astronomy, we know that the moon circles the Earth, which circles the Sun, which circles the centre of the Milky Way.

On the atomic scale, for early models of the hydrogen atom, we picture an electron orbiting around the nucleus at the centre of the atom (a single proton in hydrogen's case).

Have you ever wondered?

- Why must an airplane tilt when executing a turn?
- Why the reading of your weight would be different at the bottom of a Ferris wheel as compared to the top?
- Why do people in a roller coaster not fall out at the top of the loop? (No, it is not because of the belts!)

The exploration of the fascinating world of circular motion begins here.

A Short Revision on basic ideas of Newton's Laws of Motion

Newton's 1st Law:

Describes the motion of any object not subjected to a net external force, specifically that

- an object at rest remains at rest, or
- an object continues with constant speed in a straight line.

In the presence of a net external force, a constant mass experiences an acceleration governed by *Newton's 2nd Law*:

$$\vec{a} = \frac{\vec{F}_{net}}{m}$$

Note:

- Forces which are parallel to the motion of an object result in its motion either speeding up or slowing down, with the direction of motion remaining along the original axis.
- Forces which are perpendicular to the motion of the object cannot affect its speed, but instead will change its direction of motion.
- Forces which are neither wholly parallel nor perpendicular to the direction of motion will affect both the speed and direction of the object upon which the force acts. (Consider projectile motion.)

In this chapter you will learn that objects may move in circular paths when they experience a force or forces acting perpendicularly to their motion.

6.2 Kinematics of Circular Motion

When an extended object such as the wheel of a bicycle is rotating, different parts of the wheel will travel different distances in the same time. The outer parts of the wheel will travel larger distances and hence will have higher speeds than the inner parts of the wheel. However, all parts of the wheel will pass through the same angle in a given time. Hence the speed of rotation of a wheel is better expressed as an angular speed, as this is the same for all parts of the wheel.

6.2.1 How angle is measured

The S.I. unit for angles is not the degree but the radian. An angle measured in radians is actually the ratio of two lengths: arc length and radius. The *arc length s* is the distance travelled along the circular path, and the angle θ is said to *subtend* the arc length at the circle centre (Figure 1).

Hence, note that θ is actually a dimensionless quantity since it is a ratio of 2 lengths.

$$\theta \text{ (in rad)} = \frac{\text{arc length}}{\text{radius}} = \frac{s}{r}$$

$$s = r\theta$$

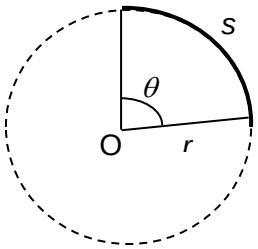


Figure 1

One radian is the angle subtended at the centre of the circle by an arc equal in length to the radius of the circle.

If the circumference of a circle $s = 2\pi r$, $\theta = s/r = 2\pi$ rad.
Since for a complete circle, the angle subtended is 360°,

$$360° \equiv 2\pi \text{ rad}$$

We can deduce that $1 \text{ radian} = \left(\frac{360}{2\pi}\right)° = 57.3°$

Conversion between degrees and radians:

$$180° \equiv \pi \text{ rad}$$

$$Z \text{ (degrees)} \equiv Z \times \frac{\pi}{180} \text{ (radians)}$$

6.2.2 Angular Displacement (θ)

The angular displacement θ is the angle an object has turned about a fixed point. In circular motion, the fixed point is taken to be the centre of the circle (Figure 2).

$$\Delta\theta = \theta - \theta_0$$

If $\theta_0 = 0$, then $\Delta\theta = \theta$.

S.I. unit of θ: radian (rad)

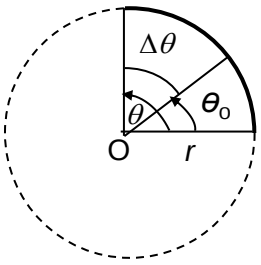


Figure 2

6.2.3 Angular Velocity (ω)

The description of circular motion in angular form is analogous to the description of linear motion. Angular velocity ω (pronounced as *omega*) is defined as the rate of change of angular displacement (Recall: In linear motion, linear velocity is defined as the rate of change of linear displacement).

Angular velocity is the rate of change of angular displacement.

$$\text{Average angular velocity, } \bar{\omega} = \frac{\text{change in angular displacement}}{\text{Elapsed time}}$$

$$\bar{\omega} = \frac{\theta - \theta_0}{t - t_0} = \frac{\Delta\theta}{\Delta t}$$

The *instantaneous angular velocity* ω is the angular velocity that exists at any given instant. Analogous to instantaneous linear velocity,

$$\text{Instantaneous angular velocity, } \quad \omega = \frac{d\theta}{dt}$$

If an object has a **constant** angular velocity, the instantaneous value and the average value are **the same** (i.e. $\omega = \bar{\omega}$).

Note:

- Magnitude of the instantaneous angular velocity is called the *instantaneous angular speed*.
- S.I. unit of ω : rad s^{-1}

6.2.4 Relationship between Tangential Speed *v* and Angular Speed ω

A particle moving in a circle has an instantaneous velocity tangential to its circular path (Figure 3). For a **constant** angular speed, the particle's **orbital or *tangential speed*** *v* is also constant. (*v* is also known as ***linear speed***.) At any instant, it is directed tangentially to the circular path at the specific point.

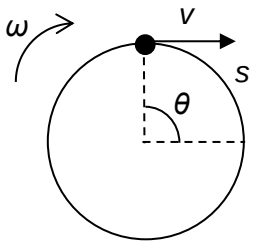


Figure 3

Since $\theta = \frac{s}{r}$ i.e. $s = r\theta$

Taking the time derivative of above and noting that *r* is constant:

$$\frac{ds}{dt} = r\frac{d\theta}{dt} \rightarrow \boxed{v = r\omega}$$ where *v* is the tangential speed and ω is the instantaneous angular speed.

Comparing the tangential speed and the angular speed of two circular motion with different radii:
Angular Velocity – xmPhysics

6.2.5 Period (*T*) and Frequency (*f*)

Period (*T*) is the time it takes for an object in circular motion to make one complete revolution, or cycle.

Frequency (*f*) is the number of revolutions, or cycles, made per unit time. The unit of frequency is s^{-1}, which is called the *hertz* (Hz) in the SI.

$$\text{frequency, } f = \frac{1}{T}$$

Note:

- When asked to *define* frequency, students often give it as the number of revolutions *per second*. This is incorrect, as a frequency can also be given as e.g. per minute or per year (frequency of the Earth).

Since an angular displacement of 2π rad is travelled in 1 period,

$$\text{angular speed, } \omega = \frac{2\pi}{T} = 2\pi f$$

Example 1

A Blu-ray player is spinning a standard Blu-ray Disc (120 mm diameter) at 900 revolutions per minute (RPM). Determine

i) the frequency of the rotation (in revolutions per second)
ii) the period of the rotation
iii) the angular speed

Example 2

Dishes are placed on a Lazy Susan at a reunion dinner during Chinese New Year. When the Lazy Susan is spun around to serve the guests, what can be said about the dishes in the inner circle compared to those in the outer circle in terms of their speed and angular velocity?

6.3 Uniform Circular Motion

Uniform circular motion is the motion of an object travelling at a **constant** (uniform) speed in a circular path.

An object moving in a uniform circular motion (Figure 4) has a *constant angular velocity* (ω). The ***magnitude*** of the velocity vector (or the linear speed, v) is ***constant*** but the ***direction*** of the velocity vector is ***changing***. Thus, the *linear velocity vector is continuously changing* ($\boldsymbol{v_1}$, $\boldsymbol{v_2}$ & $\boldsymbol{v_3}$) as the object moves around the circle. Since there is a change in the velocity with respect to time, the object can be considered to be undergoing acceleration. So the object is travelling at constant speed but is still accelerating!

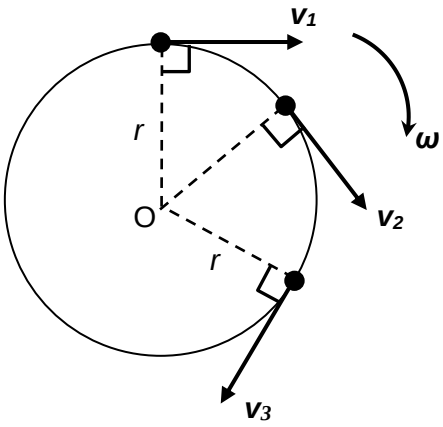


Figure 4

In the next section we derive an expression for the magnitude of this acceleration in terms of the speed and radius of the motion. We also find out the direction of this acceleration which is quite a surprising result!

Centripetal acceleration of a body moving with constant speed in circular motion:
https://xmphysics.com/2022/12/22/6-1-3/

6.3.1 Derivation of the equation of centripetal acceleration

In this derivation we only consider the kinematics of uniform circular motion to arrive at a general expression for the acceleration an object is experiencing when it moves in a circle at constant speed.
Consider a particle moving in a circle of radius *r* with uniform angular speed (Figure 5a). At one instant the particle is at A, and its instantaneous velocity is v_A in the direction AA'.

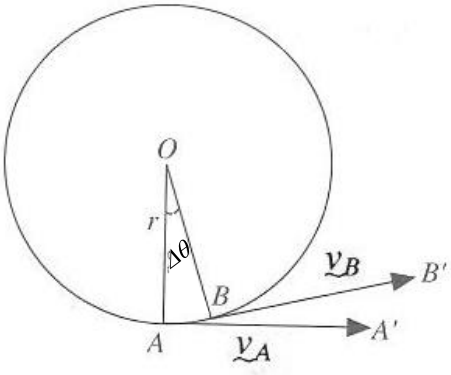


Figure 5a

A short time Δt later, the particle has moved to B, a distance $r\Delta\theta$ along the arc, where $\Delta\theta$ is a small angle, and the particle's velocity has changed to v_B in the direction BB'.

With reference to the vector triangle in Figure 5b, which is rotated 90° with respect to Figure 5a, the arc length can be approximated to a chord. For small angle $\Delta\theta$,

$$\Delta v \approx v\Delta\theta \quad \text{(note, magnitude of } v_A = v_B = v\text{)}$$
$$\Delta v/\Delta t \approx v\Delta\theta/\Delta t$$
$$\frac{dv}{dt} = v\frac{d\theta}{dt}$$
$$a = v\,\omega$$

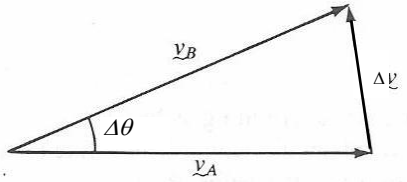


Figure 5b

Similarly,
$$a = (r\omega)\omega = r\omega^2$$
$$a = v(\frac{v}{r}) = \frac{v^2}{r}$$

As Δt (or $\Delta\theta$) (Figure 5b) becomes smaller, the angle between Δv and either v_A or v_B will tend to 90°. Since v_A and v_B are *tangent* to the circular path, this means Δv will point *towards the centre* of the circular path. From the derivation above, we see that the acceleration *a* points in the same direction as Δv, so the acceleration also points towards the centre of the circular path. This type of acceleration in uniform circular motion is called ***centripetal acceleration*** (centripetal means centre-seeking).

The centripetal acceleration is directed **radially inward** towards the centre of the circular path. The direction of the centripetal acceleration is thus **continuously changing.**

For an object in ***uniform*** *circular motion*, there is **no** acceleration component in the tangential direction (i.e. angular acceleration), or else the magnitude of the velocity vector (tangential or linear speed) would change.

The centripetal acceleration is indicated by the subscript *c*. Hence,

Centripetal acceleration, $a_c = \frac{v^2}{r} = r\omega^2 = v\omega$

Deriving the formula for centripetal acceleration $a = \frac{v^2}{r}$:
Derivation of Formula for Centripetal Acceleration v^2/r - YouTube

Conceptual Questions:

(a) Is an object travelling at non-uniform speed accelerating?

Yes, there is a change in velocity.

(b) Can an object travelling at constant speed be accelerating?

Yes, an example is an object moving in circular motion.

6.3.2 Centripetal Force (F_c)

In the last section we discovered both the magnitude and direction of the acceleration associated with circular motion. We now move into the dynamics of circular motion, quantifying and identifying forces that can constrain objects to follow such circular paths.
To provide acceleration, there must be a net force. Thus, **to produce a centripetal (inward) acceleration, there must be a resultant force towards the centre of the circular motion**. As the resultant force is centre-seeking, we call it the centripetal force, also indicated by the subscript *c*.

From Newton's second law ($\sum F = ma$),

$F_c = ma_c = \frac{mv^2}{r} = mr\omega^2 = mv\omega,$ and it is **always directed toward the centre** of the circle.

Note: In general, when a force is continuously applied **at an angle of 90° to the direction of motion** (which in this case centripetal force is), ***only the direction*** of the velocity changes.

Conceptual Question: Is there work done by a centripetal force?

No, the centripetal force is always perpendicular to motion. ($W = Fs\cos\theta$, where $\theta = 90°$)

Note: **For uniform circular motion,**

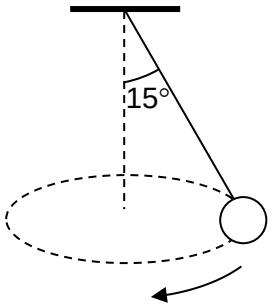
- Centripetal force is not a new type of force, it is provided by the **net force** pointing toward the centre of the circular path, and this net force is the **vector sum of all the force components that point along the radial direction**.
- Forces labelled as "centripetal" or labelled as $\frac{mv^2}{r}$ should **never** appear in a free-body diagram!

Some forces that provide the centripetal force necessary for a body to undergo a circular motion:
https://xmphysics.com/2023/01/02/6-2-1-centripetal-force/

Example 3 - Conical pendulum

In a conical pendulum system, a small pendulum bob of mass 0.50 kg is rotating in a fixed horizontal plane. The string is 30 cm long and makes an angle of 15° to the vertical.

Calculate the (a) tension in the string;
(b) linear speed of the bob;
(c) period of rotation of the bob.

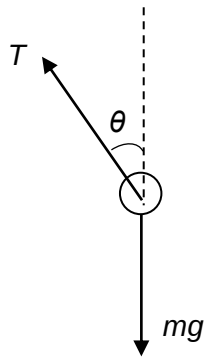


A conical problem explained:
https://www.youtube.com/watch?v=7CDwVWMvoGc

Example 4

Explain, with the aid of a diagram, why the mass at the end of a light inelastic string cannot be whirled in uniform circular motion in such a way that the string is horizontal.

.



6.4 Vertical Circular Motion

When an object is set to travel in a vertical circular path, (without the use of a motor*) its speed will not be constant as some kinetic energy will be lost to account for potential energy the object gains as it climbs the circle. This results in **non-uniform circular motion**. For non-uniform motion, *the net force is no longer equal to the centripetal force*. We can still resolve the net force into its tangential and radial components. The tangential component changes only the speed of the motion, while the radial component changes only the direction of the motion. Calculations-wise, the equation for centripetal acceleration is still $\frac{v^2}{r}$, just that centripetal acceleration must now vary as the speed changes.

Typical examples of vertical non-uniform motion, are a roller coaster going through a loop-de-loop and an object whirled in a vertical circle at the end of a string. Both will be further explored in Examples 5 and 6 below. **We will assume in both cases, that the object's total energy is conserved**. At the top of the circular motion, the object's gravitational potential energy is larger than at the bottom of the circular motion. By the principle of conservation of energy, the object will be moving faster at the bottom than at the top as it will possess more kinetic energy to compensate for having less potential energy.

Depending on the context, there are two approaches to attack problems involving vertical, non-uniform motion.

In the **FORCES APPROACH,** we follow our **general strategy for solving dynamics problems**: drawing a free-body diagram, carefully identifying the individual forces experienced by the object undergoing circular motion, and determining the net force acting on the object. Subsequently, we resolve the net force into its *tangential* and *centripetal* components.

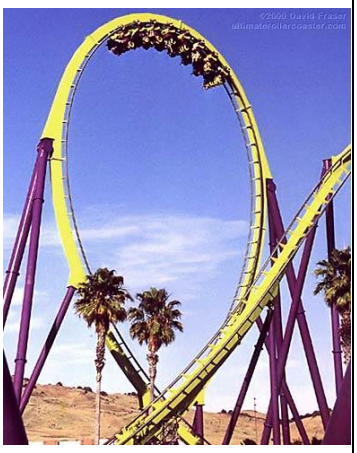
The **ENERGY APPROACH,** involving vertical, non-uniform circular motion invokes the **principle of conservation of energy**. We know how the gravitational potential energy changes as an object goes from a higher point to a lower point. Hence, if we know the difference in height between two points, we also know the difference in kinetic energy and, thus, the difference in speed.

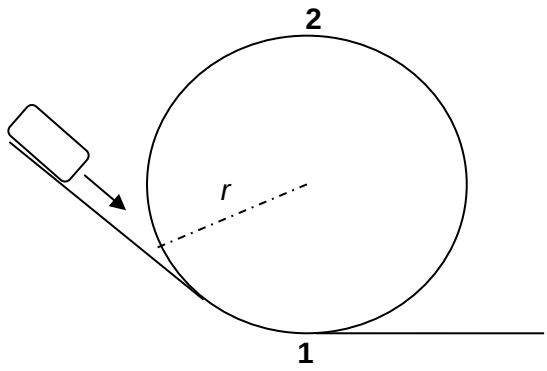
.

*Take the Singapore Flyer as an example, the carriages are moving in a uniform circular motion although it is a vertical circular path.

Example 5 - Roller coaster

Some roller coasters have several loops along the track. The picture on the right illustrates such a coaster executing a loop-the-loop.





In the question below, assume that the total mass of one car plus its passengers is 170 kg and that friction can be ignored.

(a) A passenger car for a roller coaster enters a loop of radius 19 m at position 1, as indicated in the diagram, with a speed of 33 m s^{-1}. Determine the normal contact force the track exerts on the car at
   i) the bottom (position 1) and
   ii) the top (position 2) of the loop.

(b) Find the minimum speed at which the passenger car must travel while it is at the top of the loop, in order to clear the loop safely.

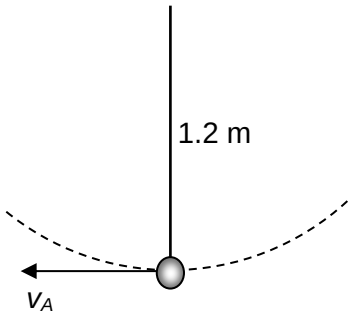
Analysing the motion of the roller coaster:
https://www.youtube.com/watch?v=bEoLhtPif2E

Example 6 – String and bob

A stone of mass 800 g is tied to one end of a string and is whirled in a vertical circle. The string is inextensible and of length 1.2 m. The stone has a certain speed v_A at the lowest point, as shown below.

(i) Determine the minimum speed the stone must have at the top of the circular motion if the string is to be taut at that instant.

(ii) Hence, show that the stone can complete a vertical circular motion if $v_A = 8.0$ m s^{-1}.



Analysing vertical circular motion:
https://www.youtube.com/watch?v=N0ByVmuoIfU

In some other instances of vertical circular motion, the speeds of the objects are kept constant (e.g. Ferris wheel). In these cases, the objects are undergoing uniform circular motion. You will encounter some of these applications in the tutorial.

6.5 Problem solving for Circular Motion

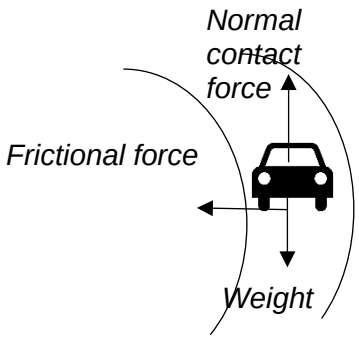
1. Identify the object undergoing circular motion (and the relevant known quantities such as m, v, r, ω…).
2. Draw a **free-body diagram** of the object, carefully identifying the individual forces *experienced by* the object, to determine the net force acting on the object.
3. **Resolve the net force** into relevant perpendicular "directions", i.e., the tangential and centripetal components.
4. The **tangential component** is usually zero. If it is not, applying Newton's 2nd Law *to the tangential component only* will tell you how fast the object is slowing down or speeding up.
5. The **centripetal component** will be equal to $\frac{mv^2}{r}$. You can use this to determine any further unknowns in the question.

Note that you have not learnt a new way of solving physics problems. The main strategy is still to draw a free-body diagram and to resolve the net force into perpendicular components. In the previous topics, it was convenient to take the horizontal and the vertical component, or the component along some slope and the component perpendicular to it. Now it is convenient to choose components parallel and perpendicular to the motion.

6.5.1 Further examples

Example 7 - Car going around a bend

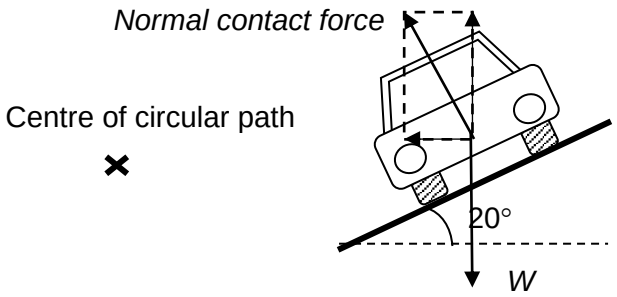
(a) A bend in the road has a 50 m radius of curvature. A car of mass 600 kg takes the bend at 45 km h^{-1}.
i) What is the centripetal acceleration of the car?
ii) What is the centripetal force experienced by the car and what provides it?
iii) What will happen to the car if driver decides to take the bend at 60 km h^{-1} instead? (Given that the maximum friction between the tyres and road surface is 3000 N.)



(b) When the car negotiates a corner on horizontal ground, the frictional force between the tyres and the ground is the only force providing the centripetal force. As there is a limit to this frictional force for a particular road surface, there is a maximum speed which the car can make the turn safely, beyond which skidding will occur.

Hence, some corners (especially at race-tracks) have raised embankments to increase the maximum speed at which a vehicle can turn the corner than if on a level road. It does so by making the **normal contact force** contribute a component to the centripetal force.

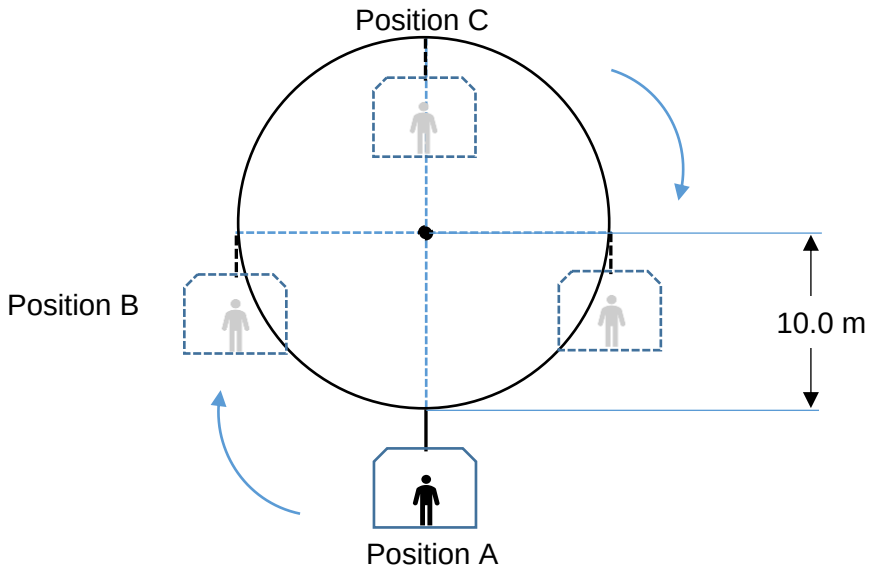
For an embankment inclined at 20° to the horizontal, find the speed at which the normal contact force is able to **completely provide the centripetal force** (i.e. no frictional force is required).



Banking of a slope and the effect on the turning of a car:
https://www.youtube.com/watch?v=KEEqE93nfe8

Example 8 - Person on a Ferris wheel

The figure below is a simplified diagram of a Ferris wheel. A single carriage with a passenger is shown.



The Ferris wheel takes 0.500 mins to make one complete revolution. Each carriage goes through a circular path of radius 10.0 m. The passenger's mass is 65.0 kg.

Determine the normal contact force exerted by the carriage on the passenger at
(a) Position A,
(b) Position B, and
(c) Position C
(d) If the passenger were to be standing on a weighing scale, what can you infer about the readings at Positions A, B and C respectively?

Circular motion of a Ferris Wheel:
https://www.youtube.com/watch?v=XwlMqWEV-pc

Tutorial 6 Circular Motion

Self-Review Questions

S1 (a) An object cannot move in a circle unless there is a resultant force acting _____ the centre of the circle. This is called a _____ force. If this force is removed, the object will continue moving in a _____ line, because of Newton's _____ law.

(b) The centripetal force causes a centripetal _____. The larger the linear _____ (in m s^{-1}) and the smaller the _____ of the circle, the larger the acceleration.

(c) The number of revolutions in one second is known as the _____. This is measured in _____. The time taken for one complete revolution is called the _____.

S2 *(N96/I/9)* A disc is rotating about an axis through its centre and perpendicular to its plane. A point P on the disc is twice as far from the axis as a point Q.

At a given instant what is the value of $\frac{the\ linear\ velocity\ of\ P}{the\ linear\ velocity\ of\ Q}$?

A 4 **B** 2 **C** ½ **D** ¼

S3 *(edited N91/I/9)* A particle travels in uniform circular motion. Which of the following correctly describes the linear velocity, angular velocity and centripetal acceleration of the particle?

	Linear velocity	*Angular velocity*	*Centripetal acceleration*
A	constant	constant	varying
B	constant	constant	zero
C	constant	varying	constant
D	varying	constant	varying
E	varying	varying	constant

S4 (*N93/I/6)* Which of the following statements is correct for a particle moving in a horizontal circle with constant angular velocity?

A The linear momentum is constant but the kinetic energy varies.

B The kinetic energy is constant but the linear momentum varies.

C Both the kinetic energy and linear momentum are constant.

D Both speed and linear velocity are constant.

E Neither the linear momentum nor the kinetic energy is constant.

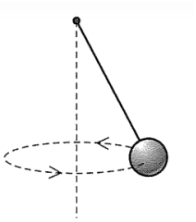
S5 *(N2015/I/11)* The minute hand of a large clock is 3.00 m long. What is the magnitude of its angular velocity?

A 1.39 x 10^{-4} $rads^{-1}$ **B** 1.75 x 10^{-3} $rads^{-1}$ **C** 5.24 x 10^{-3} $rads^{-1}$ **D** 1.05 x 10^{-1} $rads^{-1}$

S6 *(J79/II/1)* A mass of 2 kg rotates at constant speed in a horizontal circle of radius 5 m and the time for one complete revolution is 3 s. The force, in N, acting on the mass is

A $\frac{2\pi^2}{9}$ **B** $\frac{4\pi^2}{9}$ **C** $\frac{40\pi^2}{9}$ **D** $\frac{100\pi^2}{9}$ **E** $\frac{400\pi^2}{9}$

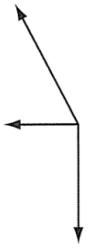
S7 *(N2013/I/11)* A small ball suspended from a light thread moves in a horizontal circle at a constant speed.



A student draws the forces acting on the ball but fails to label them.

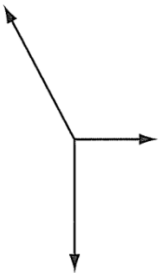
Which diagram shows the correct forces?

A



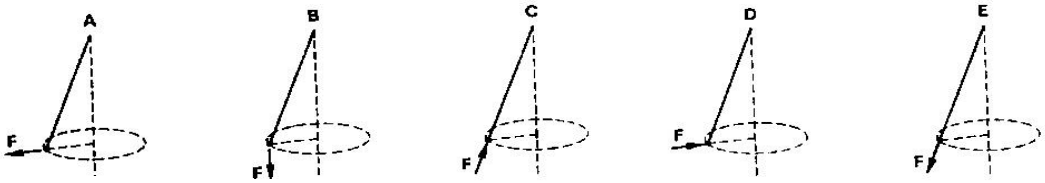
B

C



D

S8 (*J91/I/8*) A mass on the end of a string is set in motion so that it describes a circle in a horizontal plane. Which diagram shows the direction of the resultant force acting on the mass at an instant in its motion?



S9 (*J82/II/6; N85/I/4*) A car of mass *m* moving at a constant speed *v* passes over a humpback bridge of radius of curvature *r*. Given that the car remains in contact with the road, what is the net force *R* exerted by the car on the road when it is at the top of the bridge?

A $R = mg + \frac{mv^2}{r}$ **B** $R = \frac{mv^2}{r}$

C $R = mg - \frac{mv^2}{r}$ **D** $R = mg$

S10 (*J88/I/7*) An artificial satellite travels in a circular orbit about the Earth. Its rocket engine is then fired and produces a force on the satellite exactly equal and opposite to that exerted by the Earth's gravitational field. The satellite would then start to move

A along a spiral path towards the Earth's surface.
B along a tangent to the orbit.
C in a circular orbit with a longer period.
D in a circular orbit with a shorter period.

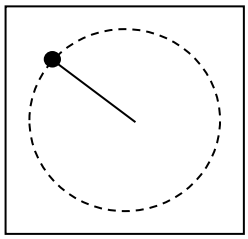
Discussion Questions

D1 A small coin of mass 2.0 g is placed on a flat horizontal turn-table. The turn-table is observed to make three revolutions in 3.14 s

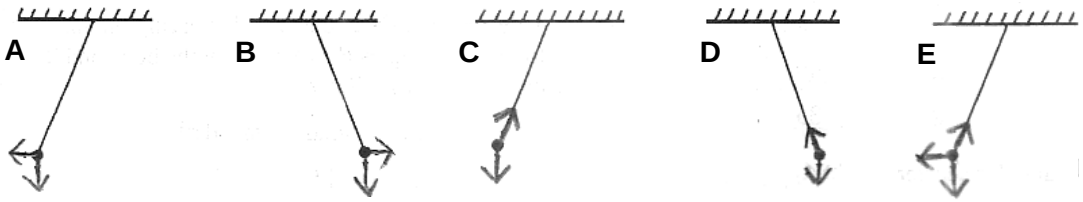
a) What is the angular displacement of the coin after it has rotated for 2.0 seconds?
b) What is the speed of the coin when it rides without slipping at a distance 5.0 cm from the centre of the turn-table?
c) What is the acceleration of the coin in part (b)?
d) What is the frictional force acting on the coin in part (b)?

D2 *(J87/I/9)* A mass of 0.050 kg is attached to one end of a piece of elastic of unstretched length 0.50 m. The force constant of the elastic is 40 N m^{-1}. The mass is rotated steadily on a smooth table in a horizontal circle of radius 0.70 m as shown. What is the approximate speed (in m s^{-1}) of the mass?

A 11
B 15
C 20
D 24
E 28



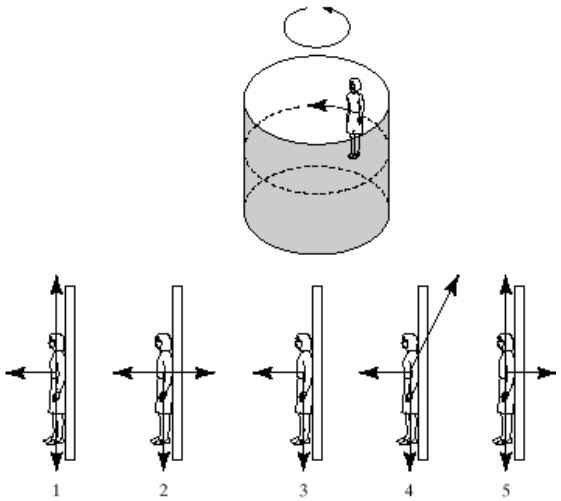
D3 (*J81/II/7)* A passenger is sitting in a railway carriage facing in the direction in which the train is travelling. A pendulum hangs down in front of him from the carriage roof. The train travels along a circular arc bending to the right. Which one of the following diagrams shows the position of the pendulum as seen by the passenger and the directions of the forces acting on it?



D4 *(J97/I/8)* The maximum safe speed of a car rounding a unbanked corner is 20 m s^{-1} when the road is dry. The maximum frictional force between the road surface and the wheels of the car is halved when the road is wet, what is the maximum safe speed (in m s^{-1}) for the car to round the corner?

A $\frac{20}{4}$ **B** $\frac{20}{2\sqrt{2}}$ **C** $\frac{20}{2}$ **D** $\frac{20}{\sqrt{2}}$

D5 In a popular amusement park ride known as the Rotor, people stand against the wall of a cylinder that is rotated. When rotating fast enough, the floor drops away, leaving the riders "pressed" against the wall in a vertical position as shown in the picture below.



For a cylinder of radius 3.00 m rotating at 5.00 rad s^{-1},

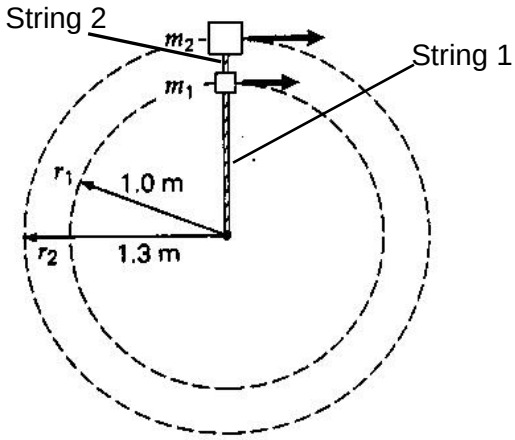
(a) identify which of the diagrams in the figure on the right correctly identifies the forces acting on the person, and suggest what keeps the person from sliding down. What physical force provides for the centripetal acceleration?

Calculate the

(b) centripetal acceleration and

(c) centripetal force, experienced by a 60 kg person.

D6 Suppose that two masses, m_1 = 2.5 kg and m_2 = 3.5 kg, respectively, are connected by light strings and are in uniform circular motion on a horizontal frictionless surface as shown on the right, where r_1 = 1.0 m and r_2 = 1.3 m. The tensions in string 2 and 1 are T_2 = 2.9 N and T_1 = 4.5 N respectively.

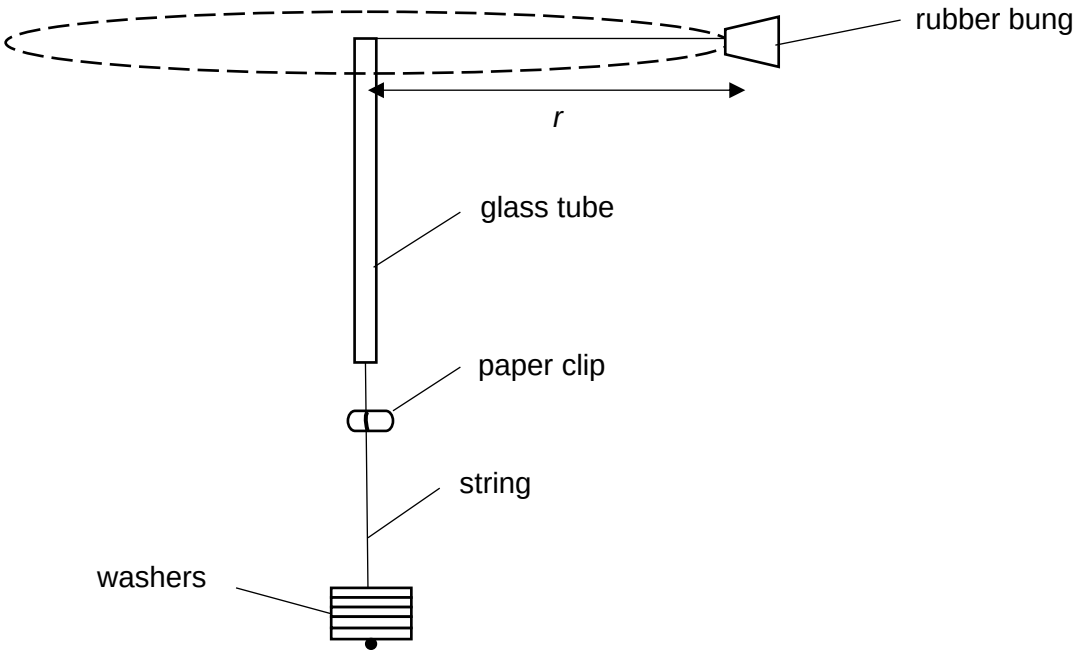


Find the

(a) centripetal accelerations,

(b) magnitude of the tangential velocities of the masses.

D7 *(2017 C2 BT2 P2 Q2)* The setup shown in the figure below is used to investigate how the force required to keep a rubber bung of mass *m* moving in uniform horizontal circular motion at a set radius, *r* varies with the speed of the motion. The rubber bung is set and maintained in motion by the experimenter gripping and twirling the glass tube.



(a) With a set radius *r* of 57.0 cm, washers of total mass 35.0 g allow the bung to perform 20 revolutions in a time of 18.2 seconds.
Show that the mass of the rubber bung is 12.6 g. [3]

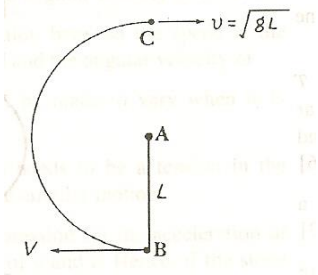
(b) Suggest the purpose of the paper clip in the experimental set up. [1]

(c) (i) Washers of mass 95.0 g are now fixed to the string and the radius of rotation is maintained at 57.0 cm. Calculate the new speed of the bung. [2]

(ii) Explain why in reality, it is not possible for the bung to rotate in a purely horizontal circle. [2]

(iii) Hence, when the bung is performing circular motion with washers of mass 95.0 g attached, calculate the angle θ, the string makes with the horizontal. [2]

D8 (*N81/I/1*) A particle is suspended from a point A by an inextensible string of length *L*. It is projected from B with a velocity *v*, perpendicular to AB, which is just sufficient for it to reach the point C.

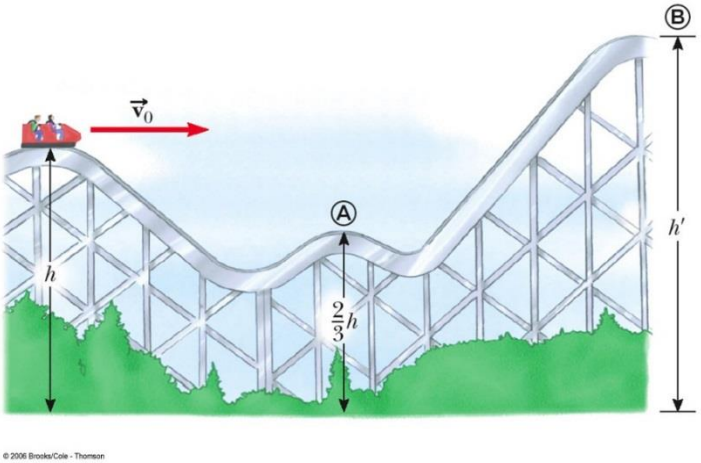


(a) Show that, if the string is just to be taut when the particle reaches C, its speed there is $\sqrt{gL}$.

(b) Find the speed *v* with which the particle should be projected from B.

D9 **Serway and Faugh. (7E) Page 223. P7.59.**

A frictionless roller coaster is given an initial velocity of v_o at height *h*, as in the figure below. The radius of curvature of the rack at point A is *R*.

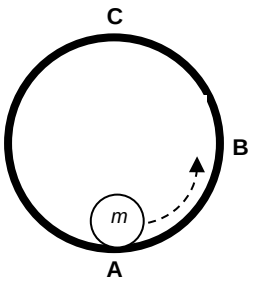


(a) Find the maximum value of v_o so that the roller coaster stays on the track at **A** solely because of gravity.

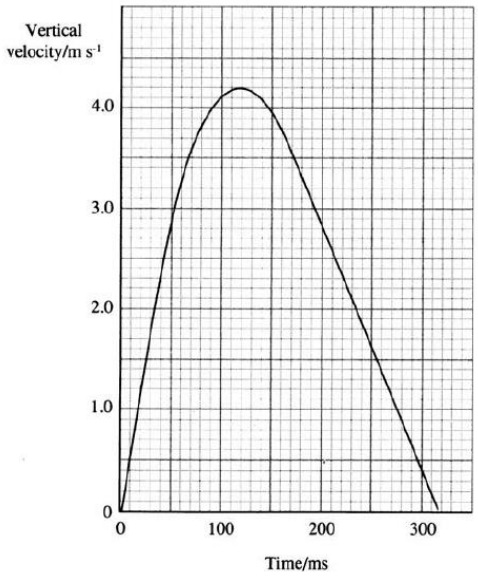
(b) Using the value of v_o calculated in (a), determine the value of *h'* that is necessary if the roller coaster just makes it to point B.

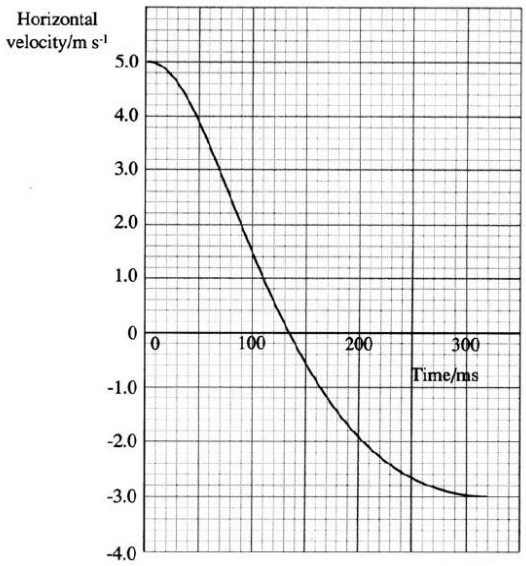
(c) Consider solution in (b), why do we not use the equations of motion that we learnt in the earlier chapter in kinematics to solve this question?

D10 *(MJC Prelims 10/I/12)* A particle of mass *m* performs vertical circular motion as shown in the diagram.



The following two graphs show the vertical and horizontal components of the velocity of the particle along path **ABC**.





Calculate the centripetal acceleration at point **C**.

A zero **B** 4.91 m s^{-2} **C** 9.81 m s^{-2} **D** 22.1 m s^{-2}

D11 Sasha's favourite ride at the fair is the Ferris wheel that has a radius of 7.0 m.

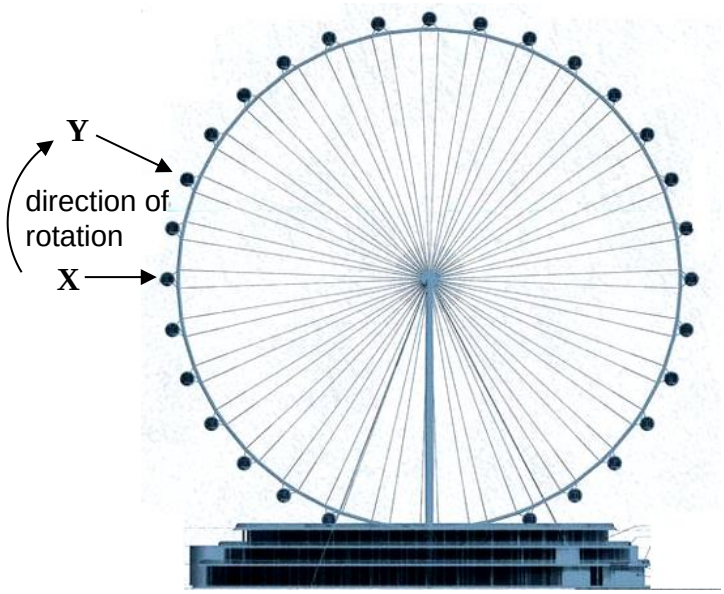
(a) If the ride takes 20.0 s to make one full revolution, what is the linear speed of the wheel?

(b) What is the magnitude of the centripetal force acting on Sasha's 50.0 kg body?

(c) In order for Sasha to feel weightless at the top of the ride, at what linear speed must the Ferris wheel turn?

(d) At this speed, how much will she appear to weigh at the bottom of the Ferris wheel?

D12 The Singapore Flyer (as shown in the figure below) is a giant observation wheel that is Asia's most visible iconic visitor attraction, providing breath-taking, panoramic views of Singapore and beyond. Completed in Mar 2008, it is one of the world's largest man-made moving land objects. It has a height of 178 m, a diameter of 150 m and sits on a 28 m high three-storey terminal building.



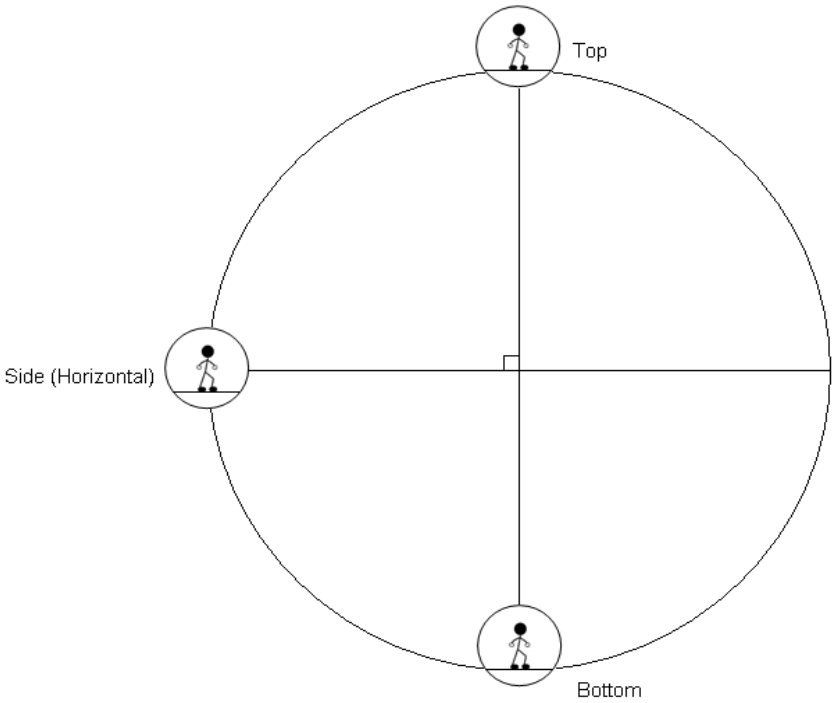
During the testing phase, the Singapore Flyer went through a series of continuous revolutions without stopping. Assume that it revolves at constant angular speed ω and each complete revolution takes 37 minutes.

(a) Calculate the angular speed of the Singapore Flyer. [2]

(b) Hence or otherwise, determine the speed of a passenger capsule which is located at the circumference. [2]

(c) Given that there are 32 equally spaced capsules on the Flyer, find the time taken for a capsule to go from position X to position Y. [2]

(d) The Singapore Flyer now revolves at a much faster rate such that the centripetal force on the passenger becomes about half his weight. Draw and label the forces acting on a passenger standing inside a capsule at the three positions in the figure below. [4]



Answers:

D1 **(a)** 5.7 rad
(b) 0.30 m s^{-1}
(c) 1.8 m s^{-2}
(d) 3.6×10^{-3} N

D5 **(b)** 75.0 m s^{-2}
(c) 4500 N

D6 **(a)** a_1 = 0.64 m s^{-2}, a_2 = 0.83 m s^{-2}
(b) v_1 = 0.80 m s^{-1}, v_2 = 1.0 m s^{-1}

D7 **(c)(i)** 6.49 ms^{-1}
(c)(iii) 7.62°

D8 **(b)** $\sqrt{5gL}$

D9 **(a)** $v_o = \sqrt{g(R - \frac{2h}{3})}$
(b) $h' = \frac{R}{2} + \frac{2h}{3}$

D11 **(a)** 2.20 m s^{-1}
(b) 34.6 N
(c) 8.29 m s^{-1}
(d) 981 N

D12 **(a)** 2.83 x 10^{-3} rad s^{-1}
(b) 0.212 m s^{-1}
(c) 139 s

CIRCULAR MOTION

Definitions:

- Angular displacement, θ: angle swept from a reference position
  Unit: rad
- Period, T: time taken for one complete revolution
- Angular velocity, ω: rate of change of angular displacement wrt time $\left(\omega = \frac{d\theta}{dt}\right)$
  Unit: rad s^{-1}
- Linear/Tangential speed v

Conditions for Uniform Circular Motion:
Constant ω and constant r

Important formulae for uniform circular motion:

- $\omega = \frac{2\pi}{T} = 2\pi f$
- $v = r\omega$
- $a_c = \frac{v^2}{r} = r\omega^2 = v\omega$

Using Newton's Laws to explain why an object moving in a constant speed in a circle experiences a resultant force towards the centre of the circle

- Since object experiences a constant change in direction of motion, by N1L, there must be a resultant force on it.
- Given the linear speed remains constant, by N2L, the resultant force must act perpendicular to the velocity, hence in the radial direction, towards the centre of the circle, to keep changing the directions of the linear velocity.

The resultant force causing the circular motion is called the centripetal force.

$$F_c = \sum F = ma_c = m\left(\frac{v^2}{r}\right) = m(r\omega^2) = m(v\omega)$$

always state "*the _____ (force) provides the centripetal force/acceleration*"

Horizontal Uniform Circular Motion

Conditions:

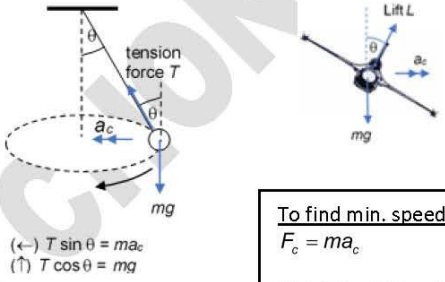
- $\sum F_{radial} = F_{centripetal} = ma_c$
- $\sum F_{tangential} = 0$ (seldom use)

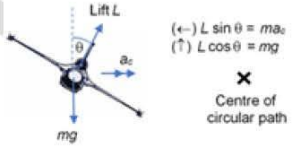
Strategy:

- Draw the FBD for the object undergoing circular motion
- Remember not to draw the centripetal force in the FBD.
- Resolve the forces radially in the horizontal plane
- **Substitute a_c** depending on the variables given.
- Balance the forces in the vertical plane.

See examples below.

For uniform circular motion, there is NO WORK done by the centripetal force since the direction of the force is always perpendicular to the direction of displacement.
Also, kinetic energy of the body remains unchanged. By work-energy theorem, no net work is done on it.





To find min. speed at the top:

$F_c = ma_c$

$N + mg = ma_c = m\left(\frac{v^2}{r}\right)$

v is minimum when N→ 0, hence

$mg = m\left(\frac{v_{min}^2}{r}\right) \Rightarrow v_{min} = \sqrt{gr}$

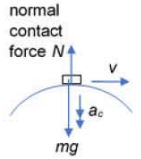
Conservation of energy:
$(KE + GPE)_{top} = (KE + GPE)_{bottom}$
If there is no frictional forces

Vertical Circular Motion

- Most examples of vertical circular motion are non-uniform, since the linear speed and ω are not constant. Eg roller coaster. If resistive forces are negligible, we use PCOE to find change in *v*. (see example below)
- For non-uniform circular motion, except when the object is at the top or bottom of the circular motion, the resultant force no longer points towards the centre of the circle due to presence of tangential acceleration.
- There are situations where ω are fixed eg Ferris wheel, to force vertical circular motion uniform, then extra work has to be done on the system.

Strategy (analyse forces acting at top/ bottom):

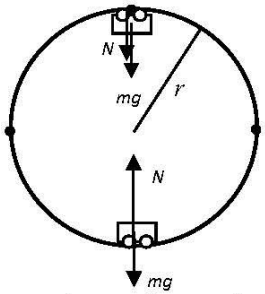
- Draw the FBD for the object in circular motion
- Do not draw the centripetal force in the FBD.
- Write the equation of motion for the object:
  $\sum F_y = F_c = ma_c$
- Substitute a_c using the variables given.
- Solve.



Car on a hump
$mg - N = ma_c$

Example of roller coaster:

At the top of the circular motion:



$F_c = ma_c$
$N + mg = ma_c$
$N = m(a_c - g)$

At the bottom of the circular motion:

$F_c = ma_c$
$N - mg = ma_c$
$N = m(a_c + g)$

Normal contact force, *N*, decreases as the object goes up and is min. at the top of the circular motion, max at the bottom.
Speed of the roller coaster also decreases as it gains height. There is a minimum velocity at the top that allows the roller coaster to just remain in contact with the track and complete the circular motion.

HWA CHONG INSTITUTION

Chapter 7

Gravitational Field

Playlist of Lecture Examples and Concept Videos can be found at
https://youtube.com/playlist?list=PL_b5cjrUKDlbcII2_xwdbabhxq0fs40Xw.

Alternative resource (with videos) at xmphysics:
https://xmphysics.com/xmGravitation/

Chapter 7: Gravitational Field

H2 Physics Syllabus 9749

Content

- Gravitational field
- Gravitational force between point masses
- Gravitational field of a point mass
- Gravitational field near to the surface of the Earth
- Gravitational potential
- Circular orbits

Learning Outcomes

Candidates should be able to:

(a) show an understanding of the concept of a gravitational field as an example of field of force and define the gravitational field strength at a point as the gravitational force exerted per unit mass placed at that point.

(b) recognize the analogy between certain qualitative and quantitative aspects of gravitational and electric fields (will be done in the chapter on electric fields).

(c) recall and use Newton's law of gravitation in the form $F = \frac{Gm_1m_2}{r^2}$.

(d) derive from Newton's law of gravitation and the definition of gravitational field strength, the equation $g = \frac{GM}{r^2}$ for the gravitational field strength of a point mass.

(e) recall and apply the equation $g = \frac{GM}{r^2}$ for the gravitational field strength of a point mass to new situations or to solve related problems.

(f) show an appreciation that near the surface of the Earth, gravitational field strength is approximately constant and equal to the acceleration of free fall.

(g) define the gravitational potential at a point as the work done per unit mass in bringing a small test mass from infinity to the point.

(h) solve problems by using the equation $\phi = -\frac{GM}{r}$ for the gravitational potential in the field of a point mass.

(i) analyse circular orbits in inverse square law fields by relating the gravitational force to the centripetal acceleration it causes.

(j) show an understanding of geostationary orbits and their application.

7.1 Newton's Law of Universal Gravitation

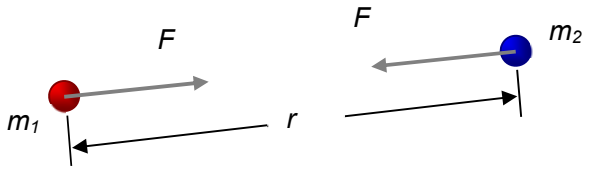
Newton's Law of Universal Gravitation states that:

> Every **point mass attracts** every other **point mass** with a **force** that is directly **proportional to the product of their masses** and **inversely proportional to the square of the distance between them**.

Consider two point masses of m_1 and m_2 separated by a distance r. Each will exert a force F on the other, and its magnitude is given by

$$F = G\frac{m_1 m_2}{r^2}$$

where
- F : gravitational force between the two point masses.
- m_1, m_2 : masses of the two point masses.
- r : distance between the two point masses.
- G : gravitational constant



The two forces form an **action-reaction pair** and have the following characteristics,

- are equal in magnitude,
- are opposite in direction,
- act on different bodies
- are of the same type (gravitational force).

G is called the *gravitational constant* (or constant of universal gravitation), which has been measured experimentally to be: $\boldsymbol{G = 6.67 \times 10^{-11}\ N\ m^2\ kg^{-2}}$. It is given in the list of constants provided on the second page of all tests and exams.

<u>Important points to note about Newton's Law of Gravitation</u>

1. Newton's Law of Gravitation is a **universal** law. It applies everywhere in the universe.

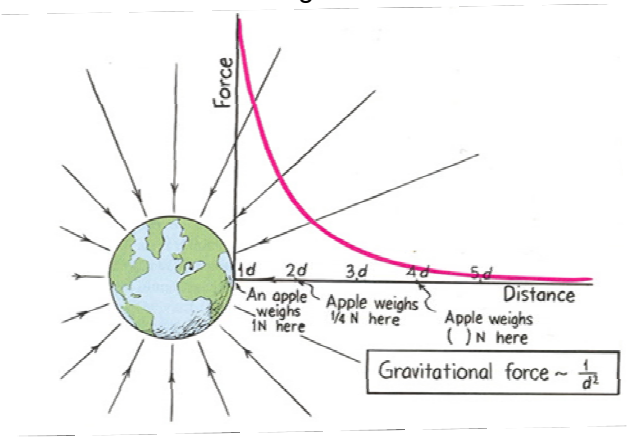
2. **Attractive nature of gravitational force:** Note that the masses in this case are always attracted to each other.

3. The gravitational force is a **field force** that always exists between two masses regardless of the medium that separates them and the two masses need not be in contact with other. In fact, it would still exist if there were no medium between them.

4. **Inverse-square law:**
   Newton's Law of Gravitation is an example of an **inverse-square law**, i.e., the force is inversely proportional to the square of the separation of the masses.

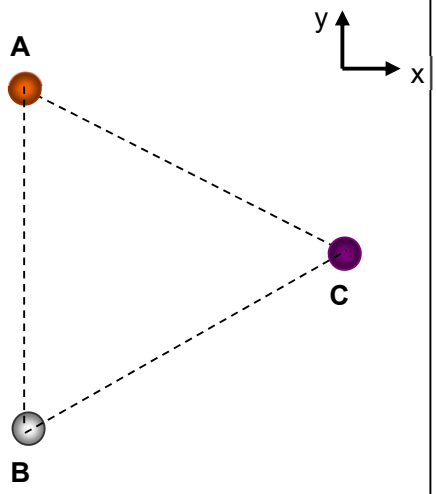
$$F \propto \frac{1}{r^2}$$

5. Newton's Law of Gravitation can be applied to spherical bodies as though they were point masses with all their mass concentrated at their geometrical centres.



Example 1: Finding Resultant Gravitational Force

Three identical masses, each of mass *m,* are located on a table at the corners of an equilateral triangle of side *d*. Determine the resultant gravitational force on mass C due to masses A and B.



Solution:

Concept: If several masses are present, each pair will experience a mutual gravitational attraction. The resultant gravitational force on a given mass is **the vector sum of the separate attractive forces acting on it due to the masses interacting with it**.

7.2 Gravitational Field Strength

The gravitational field is a **region of space** where a **mass** will experience a **gravitational force**.

Consider a particle of mass *m* placed at a point in a gravitational field (due to another mass *M*).

The particle would experience a **gravitational force**, $\boldsymbol{F_g}$, in the gravitational field, given by:

$$\boldsymbol{F_g = mg}$$

Definition: Gravitational Field Strength, *g*

The gravitational field strength *g* at a point is the **gravitational force per unit mass** acting on a small test mass placed at the point.

$$\boldsymbol{g} = \frac{F_g}{m}$$

where $\boldsymbol{g}$: gravitational field strength at that point
$\boldsymbol{F_g}$: gravitational force acting on mass *m* placed at that point in the gravitational field.

Notes:

1. S.I. unit of ***g*:** N kg^{-1} or m s^{-2}.

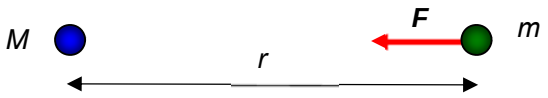
2. ***g*** is a vector quantity. Its **direction is determined by the gravitational force acting on a mass *m* placed at the point** in the gravitational field.

3. Mass *m* is a particle that is placed in the gravitational field. It is not the mass that is creating the gravitational field. Hence, mass *m* is often known as the *test* mass. The mass *M* that is creating the gravitational field is known as the *source* mass. The presence of the test mass is not necessary for the field to be present.

4. ***g*** is also commonly known as the *acceleration due to gravity* or *gravitational acceleration*.

7.2.1 Gravitational Field of a Point Mass

If a point mass *M* is placed at some point in space, it will set up a gravitational field around itself. Suppose we wish to determine the gravitational field strength ***g*** at a distance *r* away from this point mass *M*. We need to place a test mass *m* at the distance *r* away from this source mass *M*.



The test mass would experience a gravitational force ***F*** as it is in the gravitational field created by *M*.

By Newton's Law of Gravitation,

⇒ The magnitude of the gravitational force acting on *m* due to the gravitational field set up by *M*,

$$F = G\frac{Mm}{r^2}$$

From the definition of gravitational field strength,

⇒ The magnitude of gravitational field strength due to *M* at a distance *r* away from *M*,

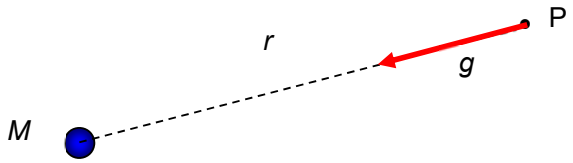
$$g = \frac{\text{Magnitude of gravitational force acting on test mass } m}{\text{Mass of test mass } m}$$

$$= \frac{F}{m}$$

$$= \frac{\left(G\frac{Mm}{r^2}\right)}{m}$$

$$= G\frac{M}{r^2}$$

Gravitational Field Strength due to a Point Mass



The magnitude of the gravitational field strength at a point due to a point mass *M* at a distance *r* away is given by:

$$g = G\frac{M}{r^2}$$

Note:
The equation is also valid for gravitational field strength outside a spherical body of uniform density or spherical shells of uniform density, where its whole mass is concentrated at its geometric centre.[1] *r* would then represent the distance between the test mass and the geometric centre of the source mass.

[1] Read about the gravitational field outside and within a sphere in Appendix A.

7.2.2 Representation of Gravitational Field

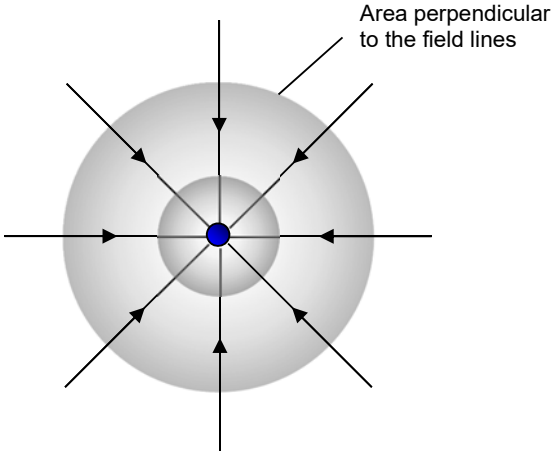
We use field lines to represent a gravitational field. The gravitational field lines are related to the gravitational field in the following manner:

1. **Direction**
   The direction of the gravitational field strength *g* is **tangent** to the gravitational field line at each point in space.

2. **Magnitude**
   The magnitude of the gravitational field strength is **proportional to the number of lines per unit area** through a surface perpendicular to the lines.
   Therefore, *g* is large when the lines are close together and small when they are far apart.

Gravitational Field Line Diagram

For a point mass



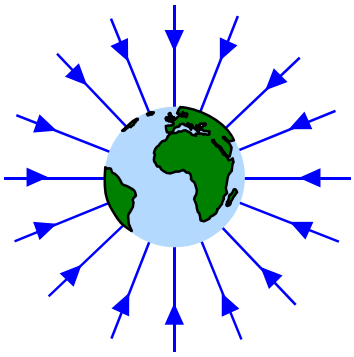
- The gravitational field in the vicinity of a point mass is always directed *radially towards* the point mass.

- The magnitude of the gravitational field is inversely proportional to the square of the distance from the point mass. Therefore, the number of field lines per unit area decreases at further distances from the point mass, indicating the diminishing strength of the field.

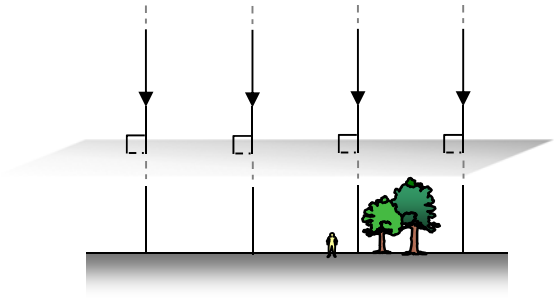
- The field is really 3-dimensional, but on paper, we can only take a 2-dimensional slice of it. This is a radial field.

- Note that
  - Fields lines end on a mass and extend back all the way to infinity.
  - Field lines never cross one another.

For Earth (spherical mass)



Far away



Close to the surface of the Earth

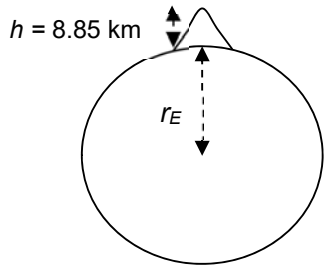
- At a large distance from the Earth (sphere), the field lines look similar to that of a point mass.

- Close to the surface of Earth,
  - The field lines are directed downwards (the direction in which a body near the Earth's surface would feel a gravitational force).
  - The field lines are *parallel and equidistant indicating that the field is constant*, or uniform.
  - The number of field lines per unit area is approximately constant for heights up to about 10 km above the Earth surface. The field strength is approximately 9.81 m s^{-2}.

Example 2: Gravitational Field Strength near the Surface of the Earth

Assume that the Earth is not rotating and is a perfect sphere of uniform density.

(i) Calculate the gravitational field strength of the Earth at its surface.
(ii) Given that Mount Everest is 8.85 km high, compute the gravitational field strength of Earth at the peak of Mount Everest.
(iii) Comment on your values computed above.



You are given the following information:
- Mass of the Earth, m_E = 5.98 x 10^{24} kg
- Average radius of Earth, r_E = 6.37 x 10^6 m

Solution:

7.2.3 Gravitational Field Strength (Acceleration due to Gravity)

If the only force acting on a mass m is the gravitational force, which is given by mg, where g is the gravitational field strength at the point, then applying Newton's second law,

$$m\boldsymbol{a} = m\boldsymbol{g}$$
$$\boldsymbol{a} = \boldsymbol{g}$$

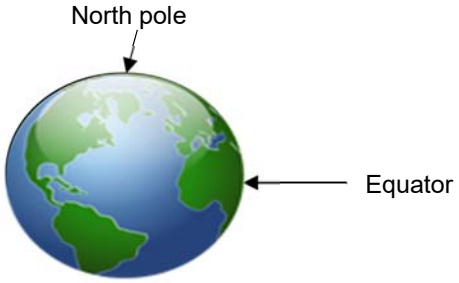
Hence the mass will have an acceleration equal to the gravitational field strength g. Thus the gravitational field strength g is also known as *acceleration due to gravity* or *gravitational acceleration*.

7.2.4 Variation of g

However, when we try to measure the value of $\boldsymbol{g}$ at various regions on Earth, the value of $\boldsymbol{g}$ varies. At the poles, $\boldsymbol{g}$ is 9.83 m s^{-2}, while at the equator, it is 9.78 m s^{-2}. In Singapore, it is about 9.79 m s^{-2}.

What accounts for this variation? There are mainly three factors affecting the value of g measured.

(1) <u>The Earth is not a perfect sphere</u>



Earth is approximately an ellipsoid, flattened at the poles and bulging at the equator. Its equatorial radius is greater than its polar radius by 21 km. Thus a point at the poles is closer to the dense core of the Earth than a point on the equator. This is one reason the gravitational field strength increases as one proceeds, at sea level, from the equator towards the poles.

(2) <u>The density of the Earth is not uniform</u>

The density of the Earth varies from region to region over the Earth's surface. The northern hemisphere, for instance, has far more land area than the southern hemisphere. Accurate measurements of the gravitational pull are used to identify different densities of materials beneath the Earth's surface. These measurements are used in prospecting for oil, coal and other minerals.

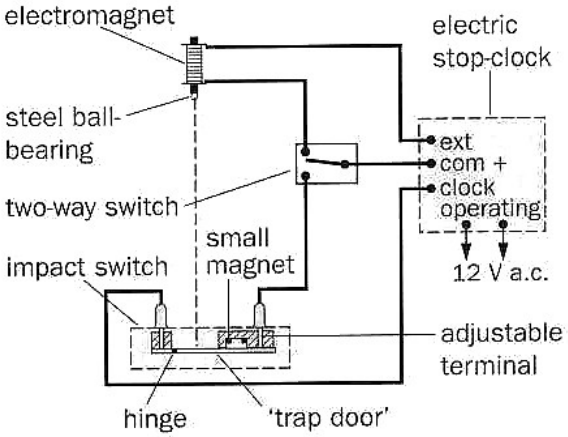
(3) <u>The Earth is rotating</u>

The Earth rotates about an axis passing through the poles. Because of the Earth's rotation, the gravitational force on a body that is not at the poles has to also provide the body with the centripetal acceleration required for rotation.

The farther the body is from the poles and the nearer it is to the equator the larger the centripetal acceleration required. Therefore, the *measured* gravitational acceleration gets smaller as one approaches the equator.

The measured acceleration is also called the acceleration of free fall (or the apparent gravitational field strength).

7.2.5 Measurement of acceleration of free fall



In this experiment, a steel ball bearing is made to fall a known distance, *s*, from rest and the time taken to fall this distance is measured using an electric stop-clock.

When the two-way switch is changed to the 'down' position, the electromagnet releases the ball and simultaneously the clock starts. At the end of its fall, the ball opens the 'trap door' of the impact switch and the clock stops.

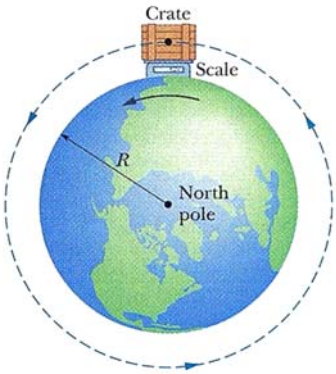
Assuming air resistance is negligible and applying the equations of motion,

$$s = ut + \frac{1}{2}at^2 \quad \Rightarrow \quad a = \frac{2s}{t^2}$$ where *t* is the time taken to fall the distance *s*.

Accurate measurements of *s* and *t* will allow the acceleration of free fall to be obtained.

Example 3: How Earth's rotation affects the weight measured

Consider a crate of mass *m* resting on a weighing scale at the equator.



Take the Earth to be a uniform sphere of mass *M* and radius *R*, spinning on its axis.

(a) On the diagram, draw all the forces acting on the crate.

(b) Write down an expression for the reading of the weighing scale supporting this mass.

(c) Write down an expression for the 'apparent' gravitational acceleration indicated by the weighing scale and determine its value.

Solution:

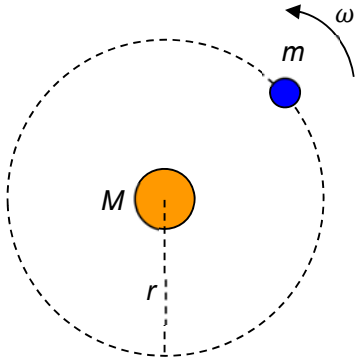
7.3 Satellites

A satellite is a moon, planet or machine that orbits a planet or star. For example, Earth is a satellite because it orbits the sun. Likewise, the moon is a satellite because it orbits Earth. Usually, the word "satellite" refers to a machine that is launched into space and moves around Earth or another body in space.

Earth and the moon are examples of natural satellites. Thousands of artificial, or man-made, satellites orbit Earth. Some take pictures of the planet that help meteorologists predict weather and track hurricanes. Some take pictures of other planets, the sun, black holes, dark matter or faraway galaxies. These pictures help scientists better understand the solar system and universe.

Still other satellites are used mainly for communications, such as beaming TV signals and phone calls around the world. A group of more than 20 satellites make up the Global Positioning System, or GPS. If you have a GPS receiver, these satellites can help figure out your exact location.

Example 4: Investigating satellite motion



Consider a satellite of mass *m* revolving around a larger mass *M* in a circle of radius *r*.

Determine the relationship between the radius of the orbit and

(a) the velocity of the satellite;

(b) the period of rotation.

Solution:

As the satellite revolves, its velocity keeps changing as the direction of motion keeps changing. Hence, the satellite is accelerating. From the topic of Circular Motion, we know that a resultant force must be providing this centripetal acceleration.

The gravitational attraction of the large mass *M* on the satellite keeps the satellite in orbit, i.e., **the gravitational force acting on the satellite (by the mass *M*) provides for the centripetal force required for the satellite to maintain circular orbit about the mass *M***.

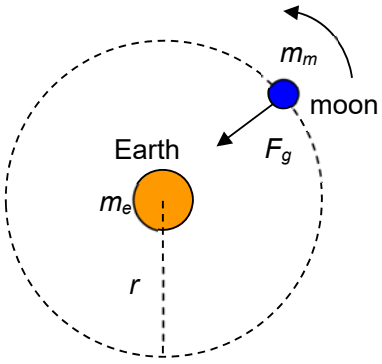
Example 5: How do you 'weigh' the Earth?

When Newton first discovered his Law of Gravitation, he had neither the value of *G* nor the mass of the Sun.

In 1798, a hundred and twenty-one years after Newton proposed the Universal Law of Gravitation, Henry Cavendish[2] conducted the first experiment to measure the force of gravity between masses in the laboratory, and obtained accurate values for the gravitational constant, *G*, and the mass density of the Earth. With the value of *G*, it was possible to obtain a value for the mass of the Earth.

By considering the motion of the moon about the Earth, show that the mass of the Earth m_e is approximately 6 x 10^{24} kg. You may assume that the distance between Earth and the moon is $r = 4.0 \times 10^8$ m.

Solution:



The **gravitational force** acting on the moon by the Earth's gravitational field **provides** for **the centripetal force** required by the moon to move in orbit about the Earth.
(NOTE: the statement above is required to explain the following steps below. Please write this in exams!)

By Newton's 2nd Law,

$$F_g = m_m a_c$$

$$\Rightarrow G\frac{m_e m_m}{r^2} = m_m r \omega_m^2 = m_m r\left(\frac{2\pi}{T_m}\right)^2$$

$$\Rightarrow m_e = \frac{4\pi^2}{G}\frac{r^3}{T_m^2}$$

Substituting for *G*, *r* and since we know that the moon takes approximately about 1 month to make one complete revolution around the Earth,

$$\Rightarrow \quad m_e = \frac{4\pi^2}{G}\frac{r^3}{T_m^2} = \frac{4\pi^2}{G}\frac{(4\times10^8)^3}{(30\times24\times60\times60)^2} = 5.64 \text{ x } 10^{24} \text{ kg} \sim 6 \text{ x } 10^{24} \text{ kg}$$

Hence, we have found the mass of the Earth!

The Cavendish Experiment - Sixty Symbols

[2] https://en.wikipedia.org/wiki/Cavendish_experiment

7.3.1 Geostationary Satellites

Geostationary satellites are satellites in orbits where the satellites are always positioned over the same geographical spot on Earth.[3] They appear stationary when seen from Earth.

Key Features of Geostationary Satellites:

1. **Period**
   - 24 hours (same period as the Earth's rotation about its own axis)

2. **Direction of rotation**
   - same direction as the rotation of the Earth, i.e. from west to east.

3. **Plane of rotation**
   - the plane that contains **the equator**.

The gravitational force on the satellite is always directed towards the centre of the Earth and so any circular orbit must have its centre at the centre of the Earth. If the orbit is not in the equatorial plane, the satellite will sometimes be over the northern hemisphere and sometimes over the southern hemisphere and so will not be geostationary.

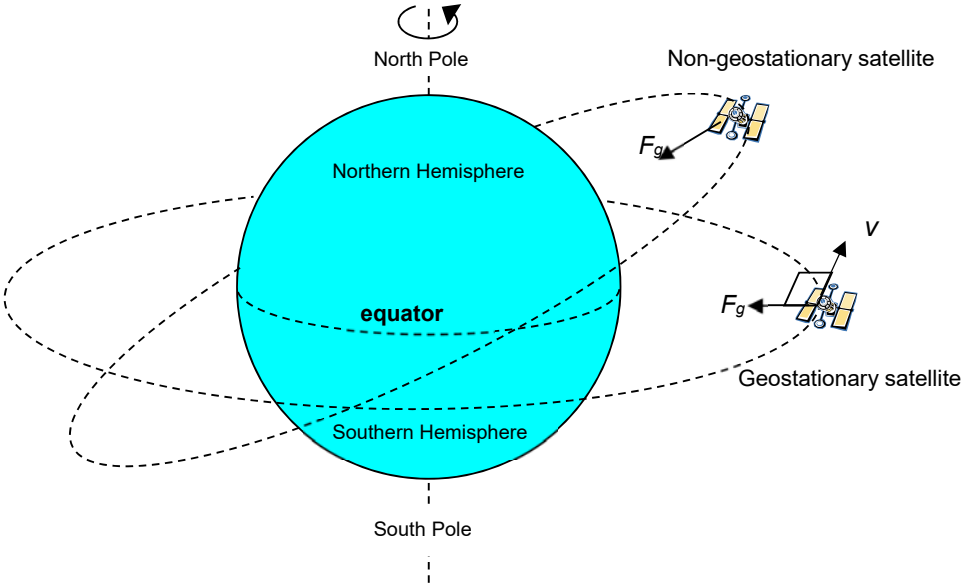


Figure not to scale

Advantages of Geostationary Satellites

1. As they remain stationary above the same point on Earth, they are ideal for use as communication satellites as they require no tracking to receive signals from Earth. This is why there is no need to keep adjusting the satellite dish to receive TV signals from a particular geostationary satellite.

[3] Besides the geostationary orbits, satellites can have other orbits as well. Check out the following sites for different types of satellites, orbits and functions by scanning the QR codes below.

NASA: Earth Observatory

Types and Uses of Satellites

2. As geostationary satellites are positioned at such a high altitude (see Example 6 for its height above the Earth's surface), they can capture almost the full-disk Earth image, rather than a small subsection. Hence, they are ideal for meteorological applications and remote imaging.

Disadvantages of Geostationary Satellites

1. As geostationary satellites are positioned at such a high altitude, the spatial resolution (i.e. the amount of detail shown) of their images tends to be not as good as satellites which are much closer to Earth.

2. Another problem with geostationary satellites is that since they are positioned above the equator, they cannot see the north or south poles and are of limited use for latitudes greater than 60-70 degrees north or south.

Example 6: Geostationary Satellite

A communications satellite of mass *m* is placed in a circular geostationary orbit.
Find its height above the Earth's surface.
(Take radius of the Earth to be 6.38×10^6 m and the mass of the Earth to be 5.98×10^{24} kg).

Solution:

The **gravitational force** on the satellite **provides** the required **centripetal force.**
(NOTE: the statement above is required to explain the following steps below. Please write this in exams!)

By Newton's second law:

7.3.2 Weightlessness

There are two types of weightlessness – true weightlessness and apparent weightlessness.

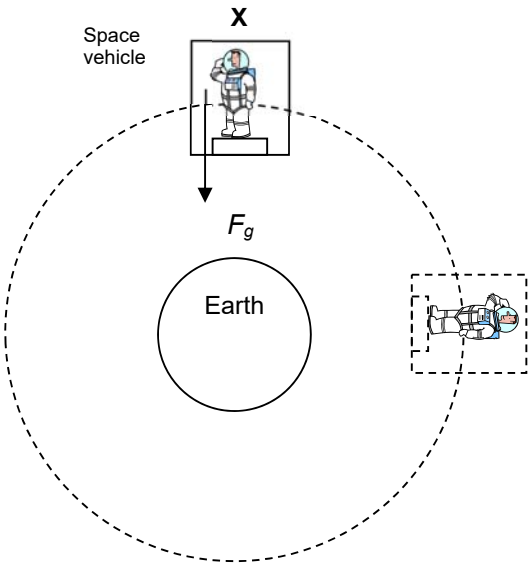
True Weightlessness

An object is truly weightless when there is no net gravitational force acting on it, such as when it is at an infinite distance away from any other body, or when it experiences opposing gravitational forces that exactly cancel one another out.

Apparent Weightlessness

An astronaut orbiting the Earth in a space vehicle with its thrusters off is often said to be 'weightless' as well. To understand why, let us try to 'weigh' this astronaut in the space vehicle.

Suppose an astronaut is standing on a weighing balance in a space vehicle orbiting the Earth. Let us first consider the space vehicle, astronaut and weighing balance as a single system. Let the mass of this system be M_{sys}.



The only force acting on this system is the gravitational force due to the Earth's gravitational field, F_g.

Since the space vehicle is in a circular orbit around the Earth,
$\Rightarrow$ F_g provides for the centripetal acceleration required for the space vehicle to move in a circular orbit.

By Newton's 2nd Law,

$$F_g = M_{sys}\, a$$

But $F_g = M_{sys}\, g$

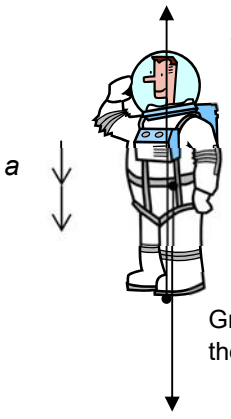
$$\Rightarrow M_{sys} g = M_{sys} a$$

Therefore, $a = g$

The space vehicle, the man and the weighing balance all have the same acceleration of g.

Now consider forces acting only on the astronaut at position X in the orbit.

.



N, Force on astronaut by weighing balance

Gravitational force on astronaut by the Earth's gravitational field, mg

By Newton's 2nd Law,

$$mg - N = ma$$

where m is the mass of the astronaut and a is the acceleration of the astronaut.

However, from above, $a = g$.

Therefore, $mg - N = mg$

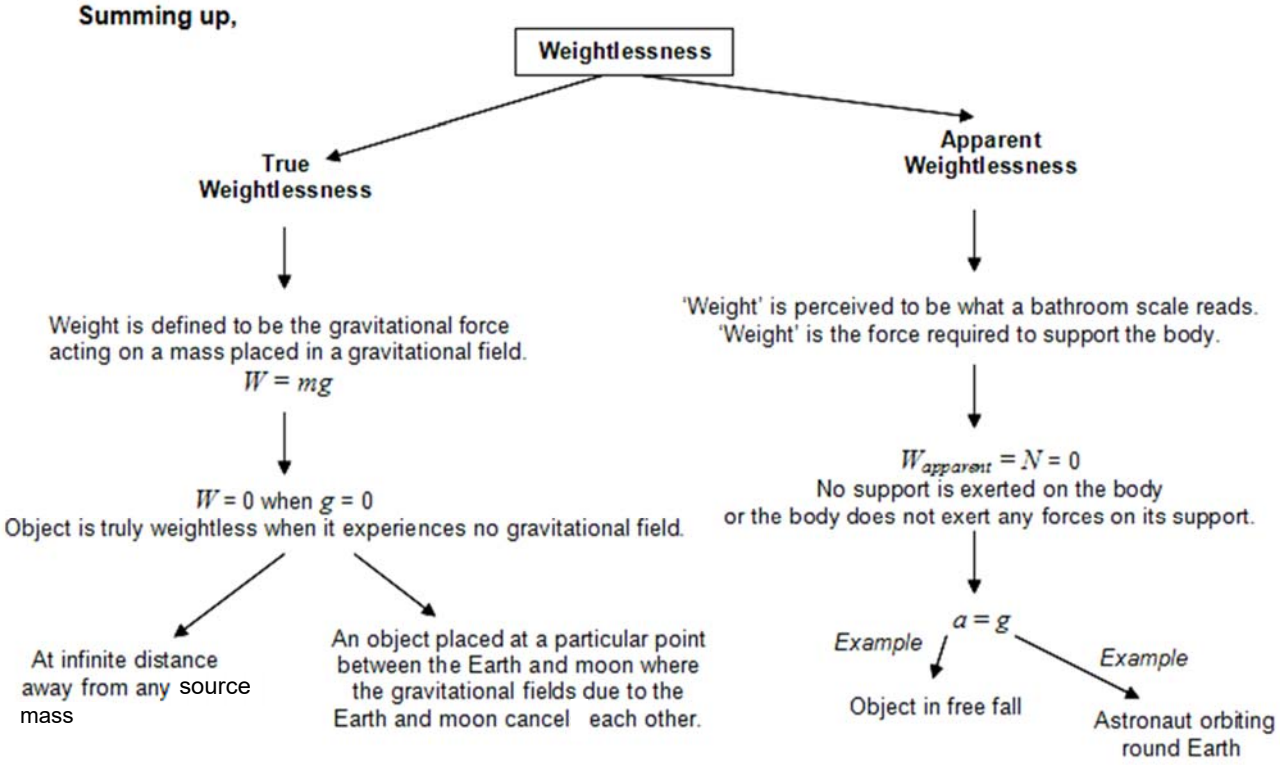
$$\Rightarrow \quad N = 0$$

Therefore, we see that the weighing balance does not exert a support force on the astronaut.

By Newton's 3rd Law, the astronaut does not exert a force on its support, the weighing balance. The reading indicated on the weighing balance is **zero**. The astronaut thus experiences the sensation of 'weightlessness'.

Hence, **an object is considered to be (apparently) weightless when it exerts no contact force on its support.** This happens when the object is accelerating with an acceleration of ***g***, e.g. an object in free fall or an object orbiting around the Earth.

Summing up,



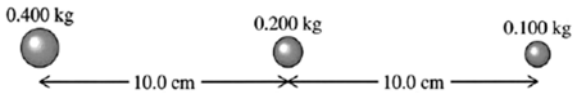
Tutorial 7A: Gravitational Force and Field

Self-Review Questions

S1 The gravitational constant G has the SI base unit

A $m\ s^{-2}$ **B** $m^{-1}\ kg^{-1}\ s^{-2}$ **C** $m^3\ kg^{-1}\ s^{-2}$ **D** $m^2\ kg^{-2}$ **E** $m^3\ s^{-2}$

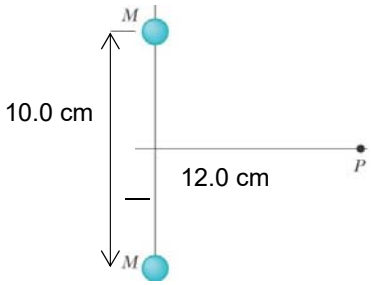
S2 **Coletta, Physics Fundamentals, Page 158, Problem 6.10.**
Find the resultant force on (a) the 0.100 kg mass and (b) the 0.200 kg mass in the diagram below (the masses are isolated from the Earth).



S3 An experimental satellite is found to have a weight W when assembled before launching from a rocket site. It is placed in a circular orbit at a height $h = 6R$ above the surface of the Earth (of radius R). What is the gravitational force acting on the satellite whilst in orbit?

A $W/6$ **B** $W/7$ **C** $W/36$ **D** $W/49$

S4 **Modified from Serway and Faugh, 6th Edition, P13.25.**
Compute the magnitude and direction of the gravitational field at a point P on the perpendicular bisector of the line joining two objects of equal mass, 20.0 kg separated by a distance 10.0 cm as shown in the figure to the right.



S5 The values of the acceleration of free fall, g, on the surfaces of two planets will be the same provided that the planets have the same

A mass **B** radius **C** mass/radius **D** mass/(radius)^2 **E** mass/(radius)^3

S6 **Coletta, Physics Fundamentals, Page 161, Problem 6.39.**
Find the radius of a planet made of solid lead, such that gravitational acceleration at the surface of the planet is 1.00 m s^{-2}. The density of lead is 1.13 x 10^4 kg m^{-3}.

S7 The gravitational field strength outside a uniform sphere of mass M is the same as that due to a point mass M at the centre of the sphere. The Earth may be taken to be a uniform sphere of radius r. The gravitational field strength at its surface is g.

What is the gravitational field strength at a height h above the surface?

A $\dfrac{gr^2}{(r+h)^2}$ **B** $\dfrac{gr}{(r+h)}$ **C** $\dfrac{g(r-h)}{r}$ **D** $\dfrac{g(r-h)^2}{r^2}$

S8 A small part of the Earth's gravitational field close to the surface of the Earth is uniform. Which of the following statements is **not** correct?

A The field lines are parallel to each other.
B The gravitational force on an object is proportional to its height above the Earth's surface.
C The units of gravitational field strength are equivalent to m s^{-2}.
D The direction of the field lines is towards the Earth.

S9 **J83/II/6 mod** Assuming Earth to be a uniform sphere rotating about an axis through the poles, the weight of a body measured on a weighing scale at the Equator compared with its weight measured on the same weighing scale at a pole would be

A greater, because the angular velocity of the Earth is greater at the equator than at the pole.
B greater, because the weight at the Equator is given by the sum of the gravitational attraction of the Earth and the centripetal force due to the circular motion of the body.
C the same, because the weight is the gravitational attraction of the Earth and for a uniform sphere, even when rotating, this is independent of the body's position on the Earth.
D smaller, because the gravitational attraction of the Earth must provide both the weight and the centripetal force due to the circular motion of the body.
E smaller, because the gravitational attraction at the pole is greater than that at the Equator.

S10 The moon remains in its orbit around the Earth rather than fall to the Earth because

A it is also attracted by the gravitational forces from the sun and other planets
B the net force on the Moon is zero
C the gravitational force exerted by the Earth on the moon provides a net force that provides the Moon's centripetal acceleration.
D the magnitude of the gravitational force from the Earth is too small to cause any appreciable acceleration of the Moon

S11 Which statement about a geostationary satellite is true?

A It can remain vertically above any chosen fixed point on the Earth.
B Its linear speed is equal to the speed of a point on the Earth's equator.
C It has the same angular velocity as the Earth's rotation on its axis.
D It is always travelling from east to west.

S12 **2009/1/6**
An object in a space capsule orbiting the Earth seems to be floating. Which statement describes the forces acting on the object?
A There are no forces on the object.
B The centrifugal force on the object is equal and opposite to its weight.
C The centripetal force on the object is equal and opposite to its weight.
D The weight of the object is the only force acting on it.

S13 **Coletta, Physics Fundamentals, Pg 159, Problem 6.23.**
Find the speed of a satellite moving in a circular orbit just above the surface of the moon. The moon has radius of 1.74 x 10^6 m and a mass of 7.36 x 10^{22} kg. Also find the satellite's orbital period.

Discussion Questions

D1 **Coletta, Physics Fundamentals, page 158. Problem 6.13.**
A particle of mass *m* is between a 1.00 x 10^2 kg mass and a 4.00 x 10^2 kg mass, which are 10.0 m apart. Find the distance of the particle from the 1.00 x 10^2 kg mass such that the resultant force on the particle is zero. (This is also known as the neutral point or null point.)

D2 **2010/1/15**
The neutral point in the gravitational field between the Sun, the Earth and the Moon is the point at which the resultant gravitational field due to the three bodies is zero. The mass of the Earth is about 80 times the mass of the Moon.

At what position is it possible for the neutral point to be? (The diagram is not drawn to scale.)

D3 **Lowe & Rounce, 3rd Edition, page 85, Ex12.2, Q2.**
The acceleration due to gravity at the Earth's surface is 9.81 m s^{-2}. Calculate the acceleration due to gravity on a planet which has
(a) the same mass and twice the radius,
(b) the same radius and twice the density,
(c) half the radius and twice the density.

D4 The acceleration of free fall on the surface of the Earth is 6 times its value on the surface of the Moon. The mean density of the Earth is 5/3 times the mean density of the moon.
What is the ratio of the radius of the Earth over the radius of the Moon?

A 1.9 **B** 3.6 **C** 6.0 **D** 10

D5 **Coletta, Physics Fundamentals, Pg 161, Problem 6.42**
Suppose the Earth rotates on its axis so quickly that an object at the equator has an apparent weight of zero. What would then be the length of an Earth day, given the radius of Earth is 6400 km?

D6 2009/3/5

(a)(i) Define gravitational field strength.

(ii) State Newton's law of gravitation and hence, using your definition in **(i)**, show that the gravitational field strength g at a distance R from a point mass M is given by $g = \frac{GM}{R^2}$

(b) A neutron star has mass 5.2×10^{30} kg and radius 1.7×10^4 m.

(i) Calculate the mean density of the star.
(ii) Suggest, with a reason, whether the density is likely to vary with the distance from the centre of the star.

(c) The mass of the star in (b) may be considered to be a point mass at its centre.

(i) Calculate the gravitational field strength at the surface of the star.
(ii) Determine the centripetal acceleration of a particle moving in a circular path of radius 1.7×10^4 m and with a period of rotation of 0.21 s.
(iii) The star rotates about its axis with a period of 0.21 s. Use your answers in (i) and (ii) to suggest whether particles on the surface of the star leave the surface owing to the high speed of rotation of the star.

D7 CIE J96/III/2(part)

(c) The Earth may be considered to be uniform sphere of radius 6370 km, spinning on its axis with a period of 24.0 hours. The gravitational field at the Earth's surface is identical with that of a point mass of 5.98×10^{24} kg at the Earth's centre. For a 1.00 kg mass situated at the Equator,

(i) calculate, using Newton's law of Gravitation, the gravitational force on the mass.
(ii) determine the force required to maintain the circular path of the mass,
(iii) deduce the reading on an accurate newton-meter (spring balance) supporting the mass.

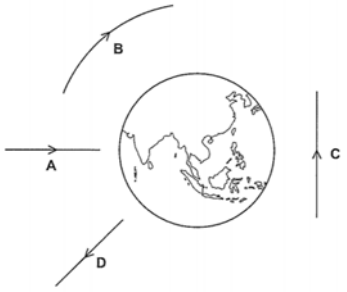
(d) Using your answers to (c), state what would be the acceleration of the mass at the Earth's surface due to
(i) the gravitational force alone,
(ii) the force measured on the newton-meter.

(e) A student situated at the Equator releases a ball from rest in a vacuum and measures its acceleration towards the Earth's surface. He then states that this acceleration is 'the acceleration due to gravity'. Comment on his statement.

D8 2018/1/11

The lines on the diagram show parts of four paths of a spacecraft moving near the Earth above any influence of the atmosphere.

Which path is not possible unless the aircraft fires its rockets as it follows the path?



D9 2021/2/4(b)

A satellite is in a circular orbit about the Earth with a period of 110 minutes.
For this satellite, the Earth may be considered to be a point mass of 6.0×10^{24} kg situated at its centre.

Determine the gravitational field strength g at the location of the satellite.

D10 2017/2/2 (part)

In July 2015, the New Horizons space probe made its closest approach of Pluto.

Charon is one of the moons of Pluto. Data, including the radius r and period of rotation about the axis T, are given for Pluto and Charon the table below.

	r/km	mass/kg	T/days
Pluto	1.20×10^3	1.31×10^{22}	6.36
Charon	0.600×10^3	1.52×10^{21}	6.36

When viewed from above, Pluto and Charon rotate in the same direction about their axes.

(a) (i) The orbital speed of Charon is 0.200 km s^{-1} and its orbital radius is 1.75×10^4 km. Show that the period of the orbit of Charon around Pluto is 6.36 days.

(ii) A space probe on the surface of Pluto is able to observe Charon over a time of several days.

Suggest what the space probe observes as a result of

1. the period of rotation of Pluto about its axis equalling the orbital period of Charon,

2. equal periods of rotation about their axes for both Pluto and Charon.

(b) Pluto may be considered to be an isolated sphere with the mass concentrated at its centre. Use data from the table to determine the gravitational field strength on the surface of Pluto.

D11 2016/3/9 (part)

A binary star consists of two stars A and B. The two stars may be considered to be isolated in space. The centres of the two stars are separated by a constant distance d, as illustrated in Fig. 9.3.

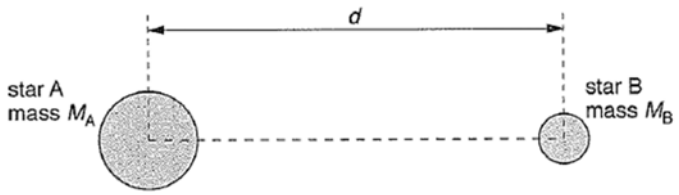


Fig. 9.3

Star A, of mass M_A, has a larger mass than star B of mass M_B, such that $\frac{M_A}{M_B} = 3.0$.

The stars are in circular orbits about each other such that the centre of their orbits is at a fixed point. Viewed from Earth over a period of time equal to the period T of the orbits, the appearance of the stars is shown in Fig. 9.4.

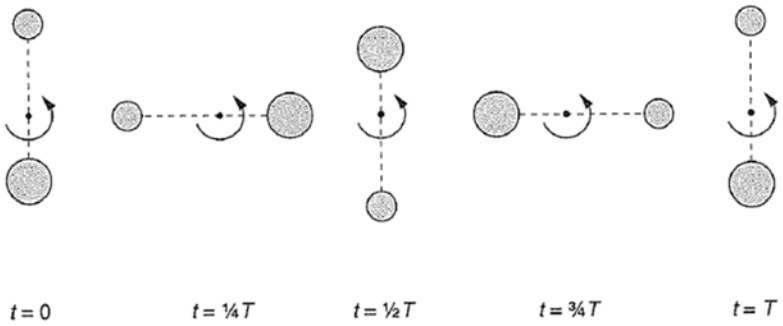


Fig. 9.4

The period T of each orbit is 4.0 years. The separation d of the centres of the stars is 3.0×10^{11} m.

(i) Explain why the centripetal forces acting on the two stars are equal in magnitude.
(ii) Calculate the angular speed ω of star A.
(iii) Determine the radius of the orbit of star A. Explain your working.
(iv) Use data and your answers from the above parts to determine the mass of each star.
(v) The plane of the orbits of the binary star in above parts is normal to the line of sight from Earth to the binary star. A second binary star has the plane of its orbits parallel to the line of sight from Earth. This binary star is so far from Earth that the individual stars cannot be distinguished. Suggest and explain what observation can be made to determine the period of the orbits of the stars.

Numerical solutions

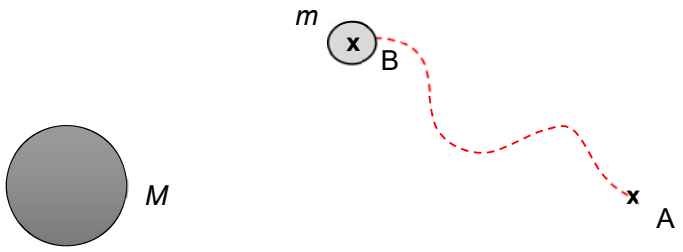
S1) C
S2) (a) 2.00 x 10^{-10} N to the left, (b) 4.01 x 10^{-10} N towards the left
S3) D
S4) 1.46 x 10^{-7} N kg^{-1},
S5) D
S6) 3.17 x 10^{5} m
S7) A
S8) B
S9) D
S10) C
S11) C
S12) D
S13) 1.68 x 10^{3} m s^{-1} and 1.81 h

D1) 3.33 m
D3) (a) 2.45 m s^{-2}, (b) 19.6 m s^{-2}, (c) 9.81 m s^{-2}
D5) 5.06 x 10^{3} s or 1.4 hrs
D6) (b)(i) 2.53 x 10^{17} kg m^{-3}, (c)(i) 1.20 x 10^{12} N kg^{-1}, (c)(ii) 1.52 x 10^{7} m s^{-2}
D7) (c)(i) 9.83 N, (ii) 0.0337 N, (iii) 9.80 N, (d)(i) 9.83 m s^{-2}, (ii) 9.80 m s^{-2}
D9) 6.9 N kg^{-1}
D10) (b) 0.607 N kg^{-1}
D11) (ii) 4.98 x 10^{-8} rad s^{-1}, (iii) 7.5 x 10^{10} m, (iv) 2.51 x 10^{29} kg, 7.53 x 10^{29} kg

7.4 Gravitational Potential and Gravitational Potential Energy

7.4.1 Gravitational Potential Energy

Consider a region of space in which a gravitational field, created by source mass *M*, is present.



Consider a mass *m* placed in this gravitational field. If no external force is applied on mass *m*, it will accelerate towards mass *M* due to the attractive gravitational force on *m* by *M*. Let's say we apply an external force on the mass to move it between two points, from A to B, without changing its kinetic energy. By conservation of energy, whatever work is done by the external force would be equal to the change in gravitational potential energy of the system of 2 masses, *M* and *m*, i.e.

work done by an external force to move *m* from A to B in the gravitational field,

$$W = \Delta U_{AB} = U_B - U_A$$

where ΔU_{AB} : Change in gravitational potential energy of the system as mass *m* moves from A to B

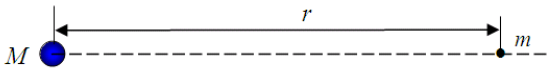
U_A : Gravitational potential energy of system when *m* is at A

U_B : Gravitational potential energy of system when *m* is at B

If point A is at an infinite distance from *M*, then the gravitational field strength due to *M* would be zero there. The gravitational potential energy is zero at that point ($U_A = U_\infty = 0$), then

$$W = \Delta U_{AB} = U_B - U_A = U_B$$

i.e. the work done by an external force to move a mass *m* from A (at infinity) to a point B without a change in kinetic energy is the gravitational potential energy of the system when mass *m* is at B.



From the concept of work done by a force, the **gravitational potential energy**[4] ***U* of a system of two point masses *M* and *m*** separated by a distance *r* is given by $\int_\infty^r \frac{GMm}{r^2} dr$. It can be shown that

$$U = -G\frac{Mm}{r}$$

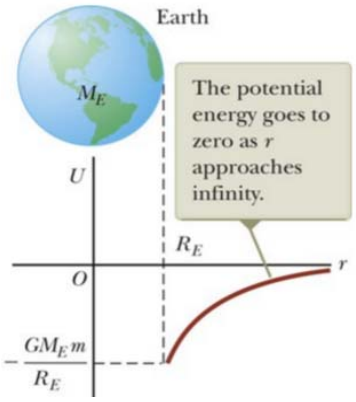
[4] A more rigorous derivation of the Gravitational Potential Energy (GPE) between 2 point masses and the gravitational potential of a point mass can be found in Appendix B.

Notes:

1. S.I. unit for *U*: joule (J).
2. The **gravitational potential energy is actually a property of the system of the two masses** rather than of either mass alone. There is no way to divide this energy and say that so much belongs to one mass and so much to the other. Therefore, we say that it is the **gravitational potential energy of the system of *M* and *m***.

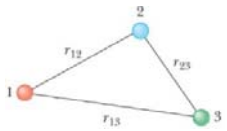
   However, if *M* >> *m*, as for the case of the Earth and the apple, we often use the phrase "the potential energy of the apple". This is because when an apple moves in the vicinity of Earth, changes in the potential energy of the Earth-apple system appear almost entirely as changes in kinetic energy of the apple, since the changes in the kinetic energy of the Earth are negligible.
3. *U* is a scalar quantity. A scalar quantity can be positive or negative in value. However, gravitational potential energy is always negative. The negative sign has nothing to do with direction.
4. **Why is *U* negative and what is the significance of the negative sign?**

   - Gravitational force is attractive in nature.
   - To bring a mass from infinity to the point in the field without changing its kinetic energy, an external force opposite to the gravitational force needs to be applied.
   - Since the external force and displacement of the mass are in opposite directions, the external force does negative work on the mass in the process.
   - As GPE at infinity is taken as zero, the GPE at any point must be negative.



Example 7: Gravitational Potential Energy of a System of 3 Masses

A system consists of three particles, each of mass *m* located at the corners of an equilateral triangle of side *x*. Determine the gravitational potential energy of the system.



Solution:

Example 8: Gravitational Potential Energy Changes near the Earth's Surface

Show that the expression $U = -G\frac{Mm}{r}$ can be reduced to the expression $mg\Delta h$ for calculations of changes in gravitational potential energy near the Earth's surface, i.e. for Δh << radius of the Earth R_E.

Solution

Example 9: Escape Speed

The minimum speed with which a mass should be projected from the Earth's surface in order to escape Earth's gravitational field is known as the **escape speed** of the body from Earth.

Show that the escape speed of the mass m placed at the surface of the Earth is $\sqrt{2gR_E}$, given that the radius of the Earth is R_E and gravitational field strength at the Earth's surface is given by g.

Solution:

Example 10: Mechanical Energies of a Satellite in Orbit

A satellite of mass *m* orbits Earth of mass *M* in a circular path. The satellite has speed *v* and the radius of its orbit about Earth is *R.*

(a) Derive an expression for the KE and the total energy of the satellite in terms of G, *M, m* and *R*.

(b) Plot a graph showing how each of the following physical quantities associated with the satellite varies with distance from the centre of Earth:
   (i) kinetic energy;
   (ii) gravitational potential energy;
   (iii) total energy.

Solution:

7.4.2 Gravitational Potential

Whenever a field is present in a region, we can always define a scalar property of the field at each point in the region. This property is called the *potential* of the field. This *potential* of a field is associated with the *potential energy per unit mass* that a test mass will possess when it is placed at the point in the field.

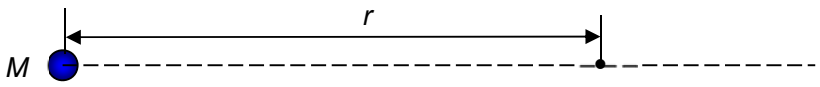
Definition: Gravitational Potential, ϕ

The **gravitational potential** at a point in a gravitational field is the work done per unit mass, by an external force, in bringing a small test mass from infinity to that point, without any change in kinetic energy.

Notes:

1. S.I. unit of gravitational potential ϕ: J kg^{-1}.
2. The gravitational potential is a **scalar quantity**.
3. The gravitational potential at infinity is taken to be zero.

Gravitational Potential due to a Point Mass :

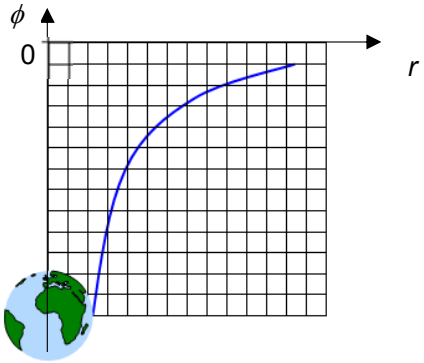


The gravitational potential ϕ at a point a distance r away from a source point mass M is given by:

$$\phi = -\frac{GM}{r}$$

Note:

1. The equation is also valid for any point outside masses of radial symmetry. This means that, for instance, we can assume the whole mass of the Earth to be concentrated at the Earth's centre. The gravitational potential *beyond the Earth's radius* can then be calculated using the formula given above.



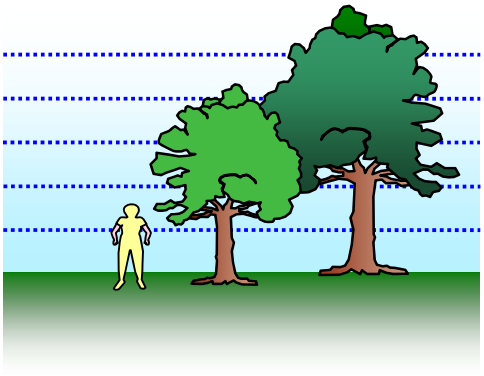
The gravitational potential at a point in the field is the work done per unit mass in bringing a test mass from infinity to that point.

Since the gravitational field is attractive and zero potential is at infinity, the gravitational potential is always negative.

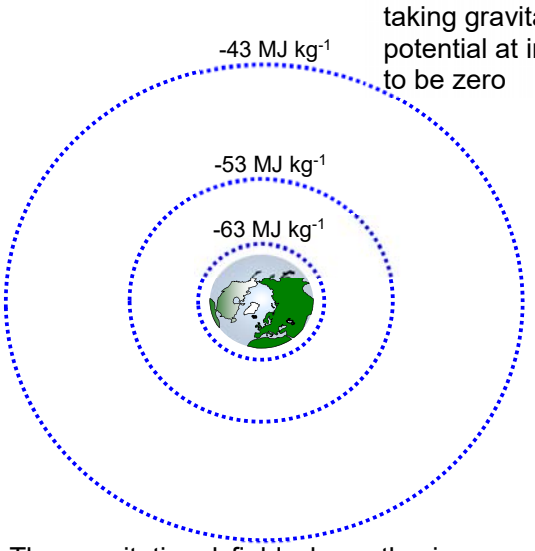
2. All points at the same distance from the centre of the Earth will have the same gravitational potential. Points of the same potential are known as points of **equipotential**.

   Equipotential lines/surfaces join points of equal potential. Below are diagrams showing the cases of uniform fields (near the surface of the Earth) and radial fields (for point and spherical masses):

Equipotential surfaces: ·····

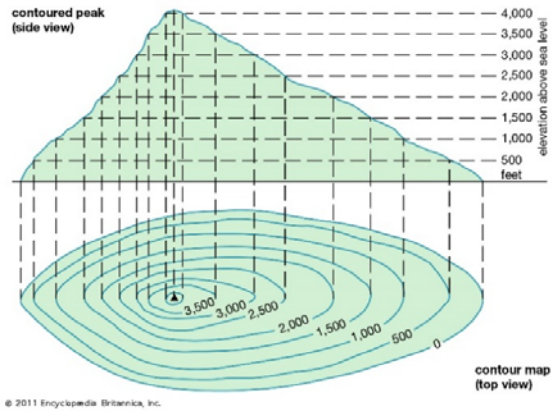


The gravitational field is uniform close to the surface of the Earth. This results in uniformly spaced equipotential lines.



The gravitational field obeys the inverse square law. Hence the distance between adjacent equipotential lines increases as distance from Earth increases.

Equipotential lines are always perpendicular to field lines. Diagrams of equipotential lines give us information about the gravitational field in much the same way as contour maps give us information about geographical heights which show regions of same heights as shown below.



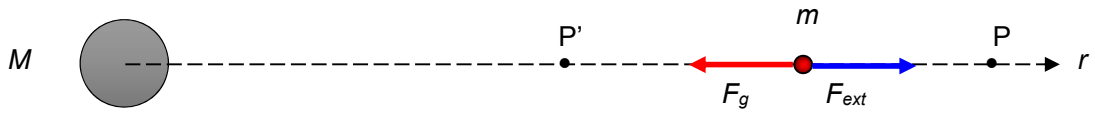
<u>Relationship between Gravitational Potential Energy *U* and gravitational potential ϕ</u>

$$U = m\phi$$

As we move a mass from one point to another, the change in gravitational potential energy is given by

$$\Delta U = m\Delta\phi = m(\phi_f - \phi_i)$$

7.4.3 Relationship between Potential and Field Strength



Consider a mass *M* fixed in space. Suppose we move a mass *m* from point P to point P' in the gravitational field of *M* by applying an external force F_{ext} such that there is no change in the kinetic energy of *m*. Then

$$\Rightarrow \qquad F_{ext} = -F_g$$

where F_g: gravitational force acting on *m*. (Note that F_{ext} and F_g vary with *r*.)

The change in gravitational potential energy of *m*, ΔU, is equal to the work done W_{ext} by external force F_{ext} in moving a mass *m* from point P to point P' (without an increase in kinetic energy) i.e.,

$$\Rightarrow \qquad W_{ext} = \Delta U$$

But $W_{ext} = \int_{r_P}^{r_{P'}} F_{ext}\, dr$

$$\Rightarrow \qquad \int F_{ext}\, dr = \Delta U$$

$$\Rightarrow \qquad F_{ext} = \frac{dU}{dr}$$

$$\Rightarrow \qquad F_g = -\frac{dU}{dr}$$

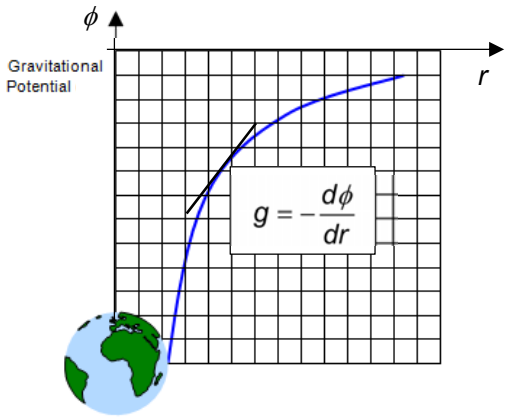
Relationship between Gravitational Force F_g and Gravitational Potential Energy *U*

$$\boxed{F_g = -\frac{dU}{dr}}$$

Dividing both sides of $F_g = -\frac{dU}{dr}$ by *m*, we obtain:

Relationship between Gravitational Field Strength *g* and Gravitational Potential ϕ

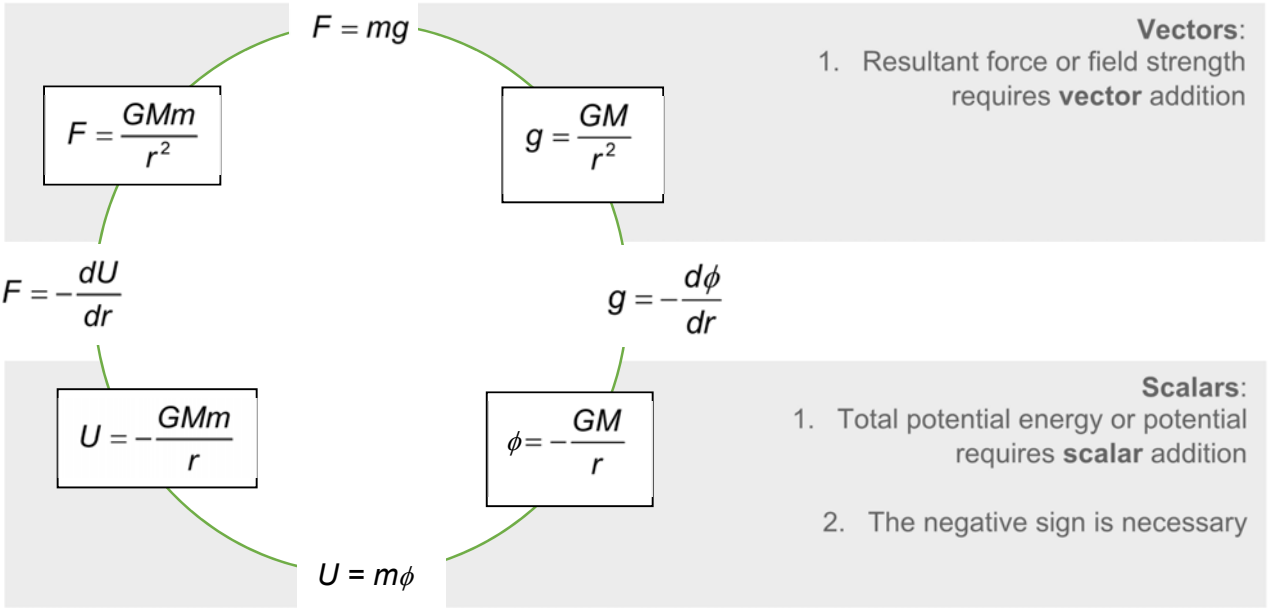
$$\boxed{g = -\frac{d\phi}{dr}}$$



Notes:

1. $\frac{d\phi}{dr}$ is known as the **potential gradient**.
2. The magnitude of the potential gradient at a point in the gravitational field gives the magnitude of the gravitational field strength at the point.
3. The **negative sign** indicates that gravitational field strength points in the direction of decreasing ϕ.

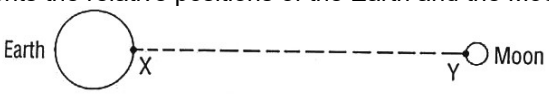
Gravitational Field Relationships



Note: The four relationships in the boxes are only valid for point masses (and spherical objects) while the other four relationships are valid for all cases.

Example 11: Gravitational Potential between 2 Masses (Mod TYS Pg 91 Qn 38 N93/I/5; N2000/I/7)

The diagram (not to scale) represents the relative positions of the Earth and the Moon.



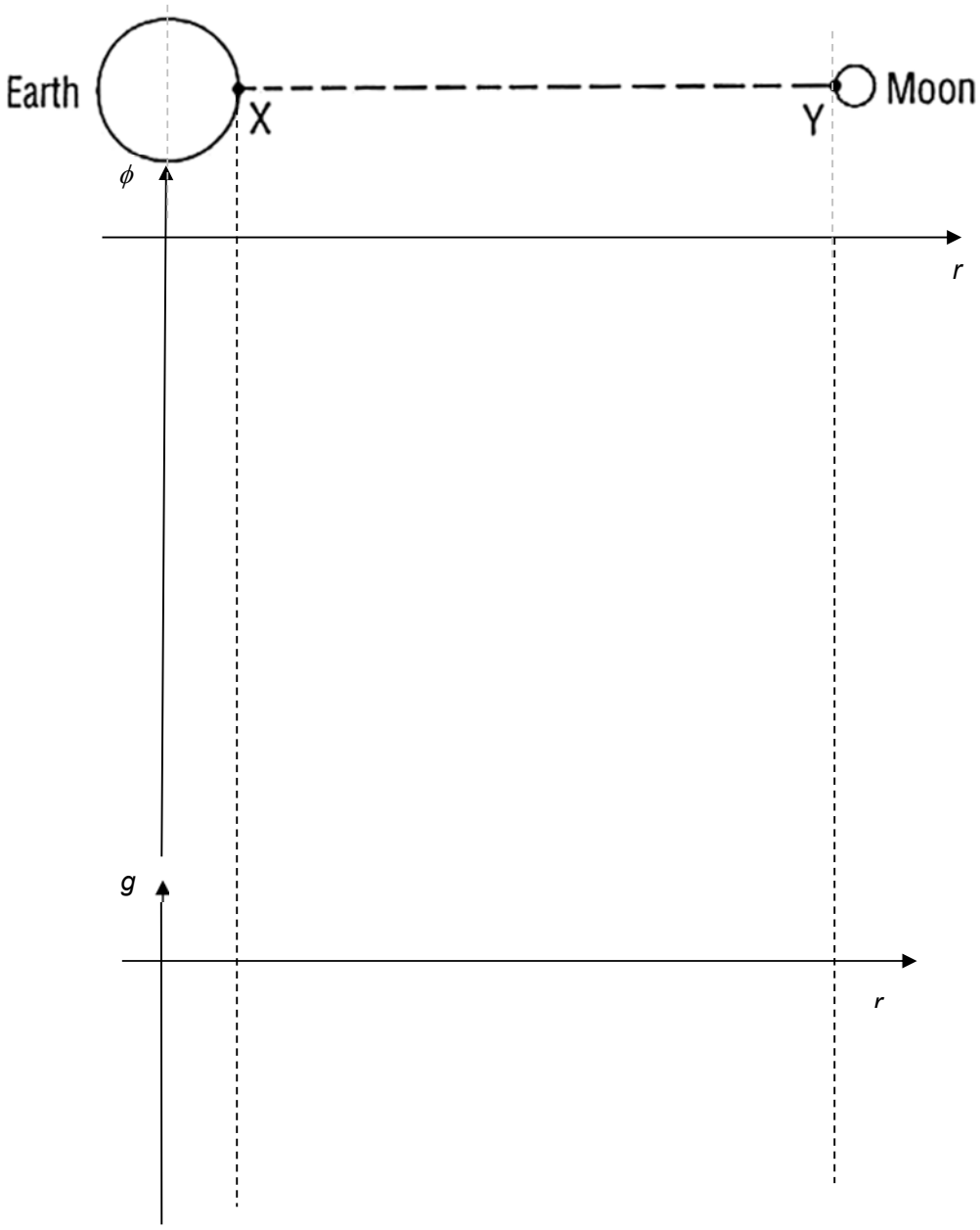
The line XY joins the surface of the Earth to the surface of the Moon.

(a) Sketch on the same axes graphs representing

(i) the gravitational potential due to the Earth between X and Y;
(ii) the gravitational potential due to the Moon between X and Y;
(iii) the resultant gravitational potential between X and Y.

(b) Sketch the corresponding variation of gravitational field strength with distance on the axis below.

Solution:



Appendices:

Appendix A: Variation of the Gravitational Field Strength above and below the Earth's Surface

Consider the Earth to be a non-rotating spherical planet of uniform density ρ. Take the radius of the Earth to be r_E, and the mass of the Earth to be m_E.

To determine the variation of g with respect to the distance from the centre of the Earth r, the following results are useful:

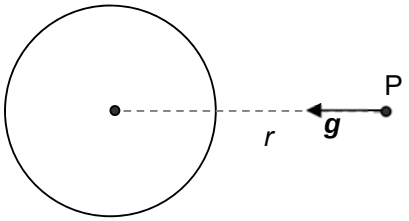
Result 1: The acceleration due to gravity **outside** a spherical body of uniform density is the same as if the entire mass of the body were concentrated at its centre.

Result 2: The acceleration due to gravity at all points **inside** a spherical shell of uniform density is zero.

With these results, we are able to obtain expressions for the acceleration due to gravity both above and below the Earth's surface.

Gravitational Field Strength above the Earth's surface ($r \geq r_E$)

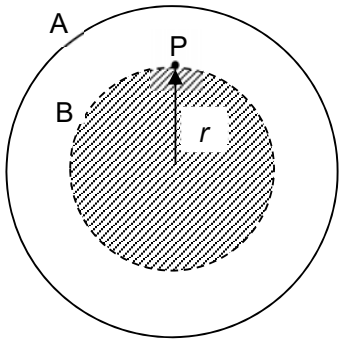
It follows from Result 1 that for $r \geq r_E$, $g = G\frac{m_E}{r^2}$; and at $r = r_E$, $g = G\frac{m_E}{r_E^2}$



Since G and m_E are constants, $g \propto \frac{1}{r^2}$.

Gravitational Field Strength below the Earth's surface ($r \leq r_E$)[5]

Consider a point P below the Earth's surface and at a distance r away from the centre of the Earth.



The gravitational field at P,

g = g at P due to spherical shell A + g at P due to spherical mass B

By Result 2,
Gravitation field at P due to shell A = 0, and

By Result 1,
Gravitation field at P due to spherical mass B
$= G\frac{M_B}{r^2}$ where M_B is the mass of the spherical mass B

[5] The gravitational field strength below the Earth's surface is not explicitly required in the syllabus, it is only introduced here for completeness.

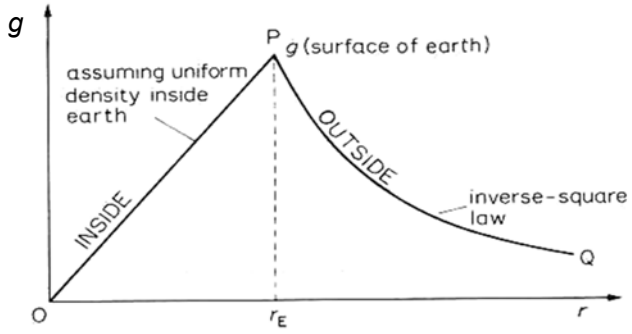
But $M_B = \rho$ (Volume of spherical mass B) = $\rho(\frac{4\pi r^3}{3})$

Hence the gravitational field at P (with r < radius of the Earth) is:

$$g = \frac{G}{r^2}(\frac{4\pi r^3 \rho}{3}) = (\frac{4\pi G\rho}{3})r$$

Since $\frac{4\pi G\rho}{3}$ is a constant, we see that for $r < r_E$, $g \propto r$.

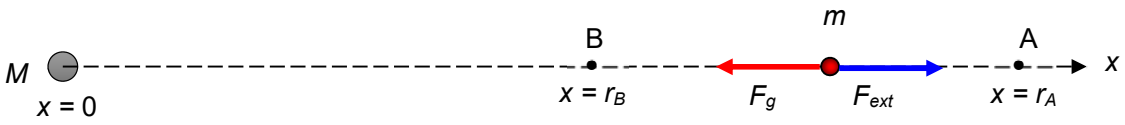
A sketch of g against r is shown below.



Appendix B: Derivation of the Gravitational Potential Energy and Gravitational Potential

Consider a particle of mass *m* moving between two points A and B in the gravitational field exerted by another particle *M.* To find an expression for the gravitational potential energy of this system, we need to first find the work done by an external force in moving the mass *m* between the two points without a change in kinetic energy of *m*.

Suppose we set up the coordinate system as below.



By Newton's Law of Gravitation, the gravitational force[6] F_g acting on the mass *m* at a distance *x* away from *M*,

$$F_g = -G\frac{Mm}{x^2}$$

The negative sign here indicates that the gravitational force is pointing in the negative *x* direction.

For the kinetic energy of *m* to remain constant, we must move *m* at a **constant velocity** (acceleration = 0).

Therefore, the external force F_{ext} acting on *m* is opposite in direction and equal in magnitude to F_g.

$$F_{ext} = G\frac{Mm}{x^2}$$

Since F_{ext} is not constant but varies with distance *x* away from *M*, to find the work done W by this external force in moving the mass *m* between A and B,

$$W = \int_{r_A}^{r_B} F_{ext}\,dx = \int_{r_A}^{r_B} G\frac{Mm}{x^2}dx = GMm\left[-\frac{1}{x}\right]_{r_A}^{r_B} = -GMm\left[\frac{1}{r_B}-\frac{1}{r_A}\right]$$

By conservation of energy, since there is no increase in the kinetic energy of the system,

W = Change in gravitational potential energy of the system, ΔU

Therefore, $W = \Delta U = U_B - U_A = -GMm\left[\frac{1}{r_B}-\frac{1}{r_A}\right] = \left[-G\frac{Mm}{r_B}\right]-\left[-G\frac{Mm}{r_A}\right]$

If the reference point A is chosen to be at an infinite distance away from M, i.e. $r_A = \infty$, the gravitational potential energy for *m* at A, U_A is equal to zero.

[6] Note that is gravitational force is an internal force to the system of *M* and *m*. To consider work done on *m*, we must always treat *M* and *m* as a single system, as *m* only experiences a force due to the gravitational field exerted by *M*.

$$U_A = U_\infty = \left[-G\frac{Mm}{r_A} \right] = 0$$

Then, $$U_B - U_A = U_B = -G\frac{Mm}{r_B}$$

Therefore, the gravitational potential energy U of a test mass m placed at a distance of r away from the source mass M, is given by

$$\boxed{U = -G\frac{Mm}{r}}$$

From the definition of the potential at a point, the potential ϕ at a point at a distance r away from the source mass M,

$$\phi = \frac{U}{m}$$

$$\Rightarrow \quad \boxed{\phi = -\frac{GM}{r}}$$

Tutorial 7B: Gravitational Potential and Potential Energy

Self-Review Questions

S1 CIE J80/II/5 (modified) *X* and *Y* are two points at respective distances *R* and 2*R* from the centre of the Earth, where *R* is greater than the radius of the Earth. The gravitational potential at *X* is –800 kJ kg^{-1}. When a 3 kg mass is taken from *X* to *Y*, the work done on the mass is

A –400 kJ **B** –1200 kJ

C +400 kJ **D** +1200 kJ

S2 N12/1/15 (modified) A satellite of mass 810 kg is to be raised from the Earth to a height of 92.0 km above the surface of the Earth. What is the necessary increase in the potential energy of the satellite?

The mass of the Earth is 5.98 x 10^{24} kg. The radius of the Earth is 6370 km.

S3 Lowe and Rounce, 3rd Edition, Pg. 86, Ex. 12.3, Q2. The moon has mass 7.7 x 10^{22} kg and radius 1.7 x 10^{6} m. Calculate

(a) the gravitational potential at its surface and

(b) the work needed to completely remove a 1.5 x 10^{3} kg space craft from its surface into outer space.

Neglect the effect of earth, other planets, sun etc.

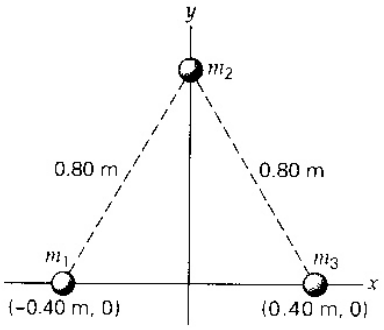
S4 Lowe and Rounce, 3rd Edition, Pg. 86, Ex. 12.3, Q4. A neutron star has radius 10 km and mass 2.5 x 10^{29} kg. A meteorite is drawn into its gravitational field. Calculate the speed with which it will strike the surface of the star. Neglect the initial speed of the meteorite.

S5 N09/1/17 (modified) A satellite orbits a planet at a distance *r* from its centre. Its gravitational potential energy is -3.2 MJ. Another identical satellite orbits the planet at a distance 2*r* from its centre. What is the sum of the kinetic energy and the gravitational potential energy of this second satellite?

Discussion Questions

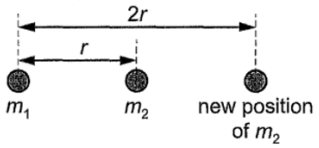
D1 **(a)** What is the total gravitational potential energy of the configuration shown in the figure on the right if all the masses are 1.0 kg?

(b) What is the work done (by an external agent) if we were to add in an additional 1.0 kg mass into the centre of the system of masses? Explain why this answer is negative.



D2 **2020/1/12**

Two objects are separated by a distance *r*. The objects can be considered to be point masses of mass m_1 and m_2. The object of mass m_1 is fixed in position. The object of mass m_2 is moved so that the separation of the two objects is 2*r*.



What is the increase in gravitational potential energy of the object of mass m_2?

A $\frac{Gm_1m_2}{2r}$ **B** $\frac{-Gm_1m_2}{2r}$ **C** $\frac{3Gm_1m_2}{4r^2}$ **D** $\frac{-3Gm_1m_2}{4r^2}$

D3 CIE J85/II/9 (part)

The curve in Fig. 5 shows the way in which the gravitational potential energy of a body of mass *m* in the field of the Earth depends on *r*, the distance from the centre of the Earth, for values of *r* greater than the Earth's radius R_E.

(a) What does the gradient of the tangent to the curve at $r = R_E$ represent?

The body referred to above is a rocket which is projected vertically upwards from the Earth. At a certain distance *R* from the centre of the Earth, the total energy of the rocket (i.e. its gravitational potential energy plus its kinetic energy) may be represented by a point on the line PQ. Five points A, B, C, D, E have been marked on this line. Which point (or points) could represent the total energy of the rocket

(b) if it were momentarily at rest at the top of its trajectory,

(c) if it were falling towards the Earth,

(d) if it were moving away from the Earth, with sufficient energy to reach an infinite distance?

In each case, explain briefly how you arrive at your answer.

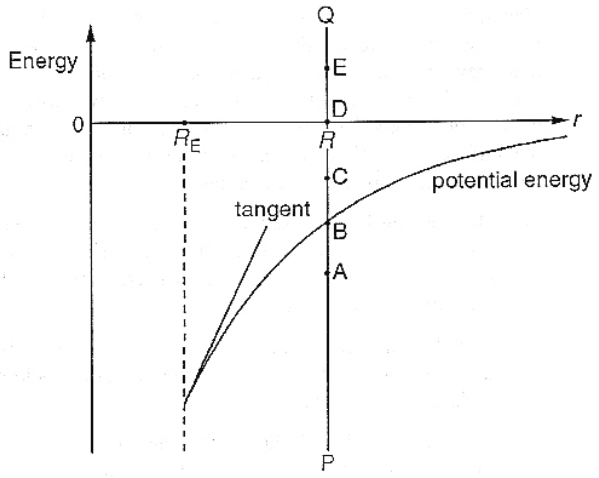


Fig. 5

D4 2021/3/2

(a) The gravitational potential Φ at a distance *x* from a point mass *M* is given by the expression $\phi = -\frac{GM}{x}$, where G is the gravitational constant.

Explain why gravitational potential is a negative quantity.

(b) A planet of diameter 6.8×10^3 km has a mass of 6.2×10^{23} kg. The planet has no atmosphere and it may be assumed to be isolated in space. The mass of the planet may be considered to be a point mass at its centre.

A meteorite collides with the planet. This causes a rock of mass 2.8 kg to be thrown up from the surface of the planet with a speed of 3.8×10^3 m s^{-1}.

(i) Calculate the gravitational potential at the surface of the planet.

(ii) Use energy considerations, and your answer in **(b)(i)** to determine whether the rock returns to the surface of the planet or travels out into space.

D5 A space station of mass 450,000 kg is in a low Earth orbit (LEO) at an altitude of 415 km. Calculate the energy required to bring it to a medium Earth orbit (MEO) of altitude 20,200 km. You are given that the mass of the Earth is 6.0×10^{24} kg and its radius is 6,378 km. You may assume that, as the space station is in LEO or MEO, it is in uniform circular motion about the centre of the Earth.

D6 Adapted from UCLES N08/II/3 and N04/II/3

A satellite of mass *m* orbits a planet of mass *M* and radius R_p with a speed of *v*. The radius of the orbit is *R*. The satellite and the planet may be considered to be point masses with their masses concentrated at their centres. They may be assumed to be isolated in space.

(a) (i) Show that speed *v* is given by the expression

$$v^2 = \frac{GM}{R}.$$

(ii) The mass of the satellite is *m*. Determine an expression for the kinetic energy E_k of the satellite in terms of *G*, *M*, *m* and *R*.

(b) (i) State an expression, in terms of *G*, *M*, *m* and *R*, for the gravitational potential energy E_p of the satellite. Explain why it is negative.

(ii) Hence, show that the satellite in orbit, the ratio of its *gravitational potential energy* to the *kinetic energy* is equal to -2.

(iii) Also show that the total energy E_t of the satellite is given by $E_t = -\frac{GMm}{2R}$.

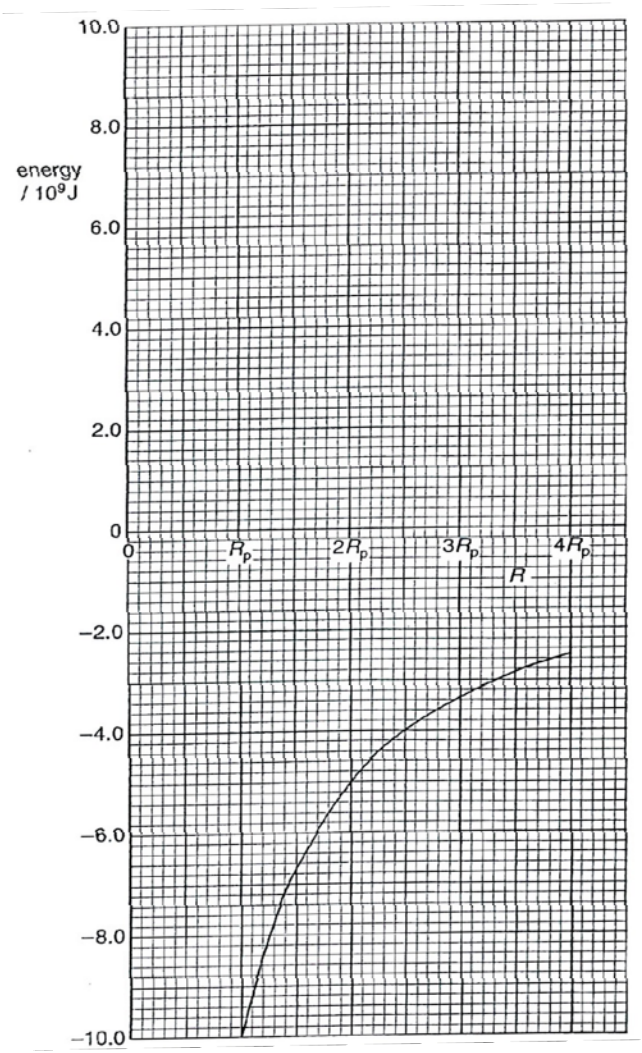
(c) The variation with orbital radius *R* of the gravitational potential energy of the satellite is shown in the figure on the right.

On the figure, draw the variation with orbital radius of

(i) the kinetic energy of the satellite,

(ii) the total energy of the satellite.

Your line should extend from $R = 1.5\ R_p$ to $R = 4\ R_p$.



(d) As the satellite orbits the planet, it gradually loses energy because of air resistance by the atmosphere of the planet.

(i) State whether the total energy E_t becomes more or less negative.

(ii) Hence, state and explain the effect of this change on

1. the radius of the orbit,

2. the speed of the satellite.

(iii) The mass *m* of the satellite is 1600 kg. The radius of the orbit of the satellite is changed from $R = 4\ R_p$ to $R = 2R_p$. Use the figure to determine the change in orbital speed of the satellite.

D7 CIE J87/II/8 (part)

A point mass m is at a distance r from the centre of the Earth.

(a) Write down an expression, in terms of m, r, the Earth's mass m_E and the gravitational constant G, for the gravitational potential energy V of the mass. (Consider only values of r greater than the Earth's radius. Mass of Moon = 7.4 x 10^{22} kg; mass of Earth = 6.0 x 10^{24} kg.)

Certain meteorites (tektites) found on Earth have a composition identical with that of lunar granite. It is thought that they may be debris from a volcanic eruption on the Moon. Fig. 6, which is not to scale, shows how the gravitational potential between the surface of the Moon and the surface of the Earth varies along the line connecting their centres. At the point P the gravitational potential is at a maximum.

(b) State how the resultant gravitational force on the tektite at any point between the Moon and the Earth could be deduced from Fig. 6.

(c) When a tektite is at P the gravitational forces on it due to Moon and Earth are F_M and F_E respectively. State the relation which applies between F_M and F_E. Hence find the values of X/Y, where X and Y are the distances of P from the centre of the Moon and the centre of the Earth respectively.

(d) If a tektite is to reach the Earth, it must be ejected from the volcano on the Moon with a certain minimum speed V_0. Making use of appropriate values from Fig. 6, find this speed. Explain your reasoning.

(e) Discuss very briefly whether a tektite will reach the Earth's surface with a speed less than, equal to or greater than the speed of projection. (Neglect atmospheric resistance.)

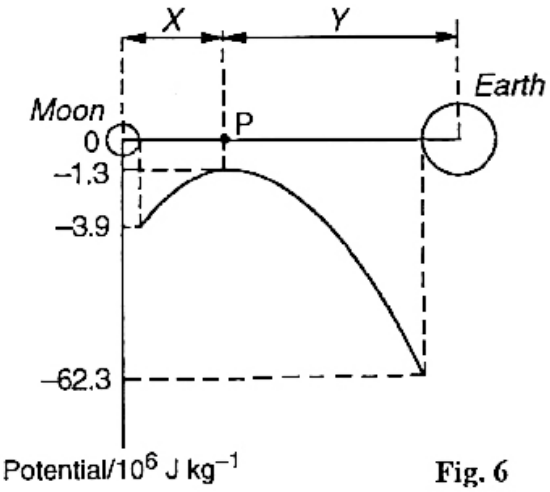
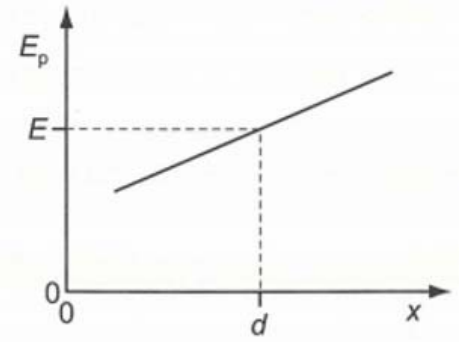


Fig. 6

D8 2011/I/12

A stone of mass *m* moves radially away from Earth.

For a small distance *x* above the Earth's surface, the variation with *x* of the stone's gravitational potential energy E_P is shown.



At point P, a distance *d* from Earth, the potential energy of the stone is *E* and the rate of change of potential energy with distance is *R*.

What is the force acting on the stone?

A *E*/*d* **B** *mE*/*d* **C** *mR* **D** *R*

D9 2021/I/10

Near the surface of a planet, the gravitational field strength is uniform. The gravitational potential difference between two points that are 10 m apart vertically is 6.0 J kg^{-1}.

How much work is done in raising a mass of 2.0 kg vertically through 2.5 m?

A 1.5 J **B** 3.0 J **C** 12 J **D** 15 J

Numerical solutions

S1) D, S2) 7.22 × 10^8 J, S3) (a) -3.0 x 10^6 J kg^{-1}, (b) 4.5 x 10^9 J, S4) 5.8 x 10^7 m s^{-1}, S5) - 0.8 MJ

D1) (a) -2.5 x 10^{-10} J (b) -4.3 x 10^{-10} J
D4) (b)(i) – 1.2 × 10^7 J kg^{-1}
D5) 9.9 x 10^{12} J
D6) (d)(iii) 518 m s^{-1}
D7) (c) *X*/*Y* = 0.11; (d) 2.3 x 10^3 m s^{-1}

Chapter 8

Oscillations

A metronome, which is basically an upside down pendulum, is a device that produces regular ticks (beats). It dates back to the early 19th century. A metronome is used by some performing musicians for practice in maintaining a consistent tempo; it gives the composer an approximate way of specifying the tempo. From its inception, however, the metronome has been a highly controversial tool, and there are musicians who reject its use altogether.

- http://en.wikipedia.org/wiki/Metronome

Oscillations

Candidates should be able to:

(a) describe simple examples of free oscillations.

(b) investigate the motion of an oscillator, using experimental and graphical methods.

(c) show an understanding of and use the terms amplitude, period, frequency, angular frequency and phase difference and express the period in terms of both frequency and angular frequency.

(d) recall and use the equation $a = -\omega^2 x$ as the defining equation of simple harmonic motion.

(e) recognise and use $x = x_0 \sin \omega t$ as a solution to the equation $a = -\omega^2 x$.

(f) recognise and use the equations $v = v_0 \cos \omega t$ and $v = \pm\omega\sqrt{x_0^2 - x^2}$.

(g) describe, with graphical illustrations, the changes in displacement, velocity and acceleration during simple harmonic motion.

(h) describe the interchange between kinetic energy and potential energy during simple harmonic motion.

(i) describe practical examples of damped oscillations with particular reference to the effects of the degree of damping and to the importance of critical damping in applications such as a car suspension system.

(j) describe practical examples of forced oscillations and resonance.

(k) describe graphically how the amplitude of a forced oscillation changes with driving frequency near to the natural frequency of the system, and understand qualitatively the factors which determine the frequency response and sharpness of the resonance.

(l) show an appreciation that there are some circumstances in which resonance is useful and other circumstances in which resonance should be avoided.

RELEVANT E-LEARNING WEBSITES

Applets

1) Spring-Mass SHM http://ngsir.netfirms.com/j/Eng/springSHM/springSHM_js.htm
2) Mass Spring Lab https://phet.colorado.edu/en/simulation/mass-spring-lab

Online lessons

3) University of Salford Online Lessons: http://www.acoustics.salford.ac.uk/feschools/waves/shm.htm#motion
4) PHYSCLIP Online Lessons
   a. Mechanics: SHM (http://www.animations.physics.unsw.edu.au/mechanics/chapter4_simpleharmonicmotion.html)
   b. Waves and Sound: Oscillations (http://www.animations.physics.unsw.edu.au/waves-sound/oscillations/)
5) Schoolphysics: http://www.schoolphysics.co.uk/age16-19/Mechanics/Simple%20harmonic%20motion/
6) How Radio works http://electronics.howstuffworks.com/radio8.htm

Table of Contents

8.0 Introduction 4
8.1 Simple Harmonic Motion 4
8.1.1 Terms used to describe SHM 5
8.1.2 SHM - Displacement-time (*x*-*t*) relationship 6
8.1.3 SHM - Velocity-time (*v*-*t*) relationship and acceleration-time (*a*-*t*) relationship 6
8.1.4 Definition of SHM – relationship between acceleration and displacement 9
8.1.5 Simple Harmonic Motion: velocity-displacement (*v*-*x*) relationship 12
8.2 Relationship between Simple Harmonic Motion and Uniform Circular Motion 14
8.3 Energy in SHM 17
8.3.1 Energy-displacement Graphs of SHM 17
8.3.2 Energy-time Graphs of SHM 18
8.4 Damped Oscillations, Forced Oscillations and Resonance 19
8.4.1 Damped oscillations 19
8.4.2 Forced oscillations and resonance 21
8.4.3 Examples of Destructive and Useful Resonance 26
Appendix 1 Case study: Oscillations of A Simple Pendulum 27
Appendix 2 Case study: Oscillations of a Mass on Vertical Spring 28
Appendix 3 Equations of motion for damped simple harmonic oscillations 29
Appendix 4 Demonstration of Forced Oscillation and Resonance - Barton's pendulums 30
Tutorial 8 Oscillations 31

Playlist of teaching videos and lecture examples:

https://youtube.com/playlist?list=PL_b5cjrUKDlYadbhfFX4Z8AwqHX1vdl9l

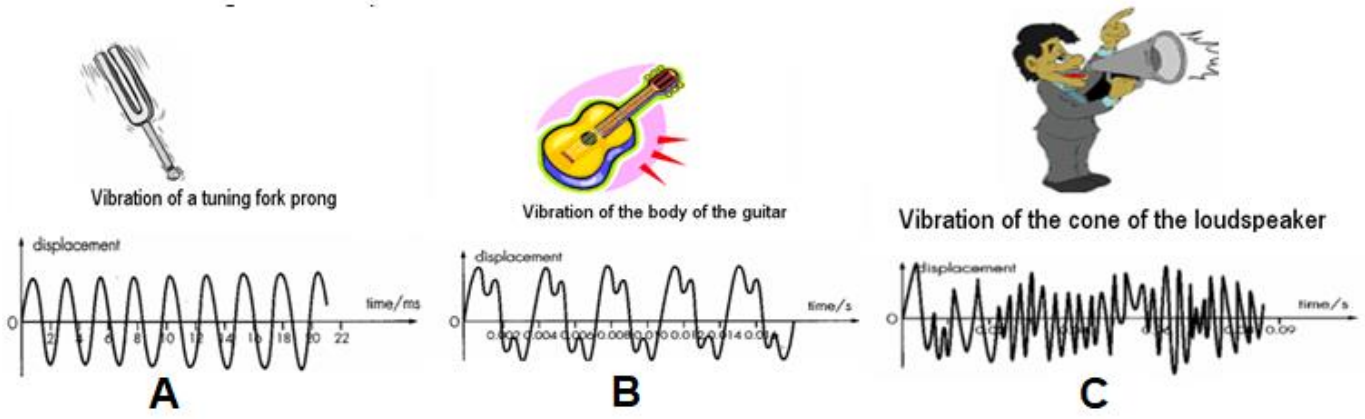
8.0 Introduction

Oscillations are very common in everyday life such as the motion of a clock pendulum, the vibrations of strings in musical instruments, the vertical oscillations of a car after passing over a bump on the road or even the vibrations of atoms in a lattice.

Oscillations can occur when displacing an object that is in a state of stable equilibrium from its equilibrium position results in a restoring force acting on the object directed towards the equilibrium position. The motion is said to be *periodic* if the displacement of a body in oscillatory motion repeats itself at equal intervals. The full range of its motion that is repeated is referred to as a cycle and the time taken for the body to go through a cycle is the period of the oscillation.

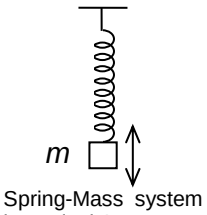
Example 8.1: Periodic Motion

Which of the following cases are periodic oscillations?

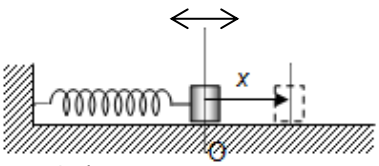


8.1 Simple Harmonic Motion

The free oscillation of a mass attached to a spring is a very special example of a periodic oscillation that we call simple harmonic motion (SHM)**.** It is special because the displacement of the mass about its equilibrium position varies sinusoidally. We will prove this analytically but first, let us look at an experiment that shows the sinusoidal variation of its displacement with time.



Spring-Mass system in vertical SHM



Spring-Mass system in horizontal SHM

Spring-Mass System

$$T = 2\pi\sqrt{\frac{m}{k}} \quad (*)$$

where T = period of oscillation
m = mass of the oscillating object
k = spring constant

(*) See Appendix 2 for the derivation of this relationship.

We can obtain a graph of the variation in the displacement of the mass hanging from a spring via the following experiment. We attach a pen to the mass such that its motion is traced onto a vertical sheet of graph paper that is pulled to the left at a constant rate.

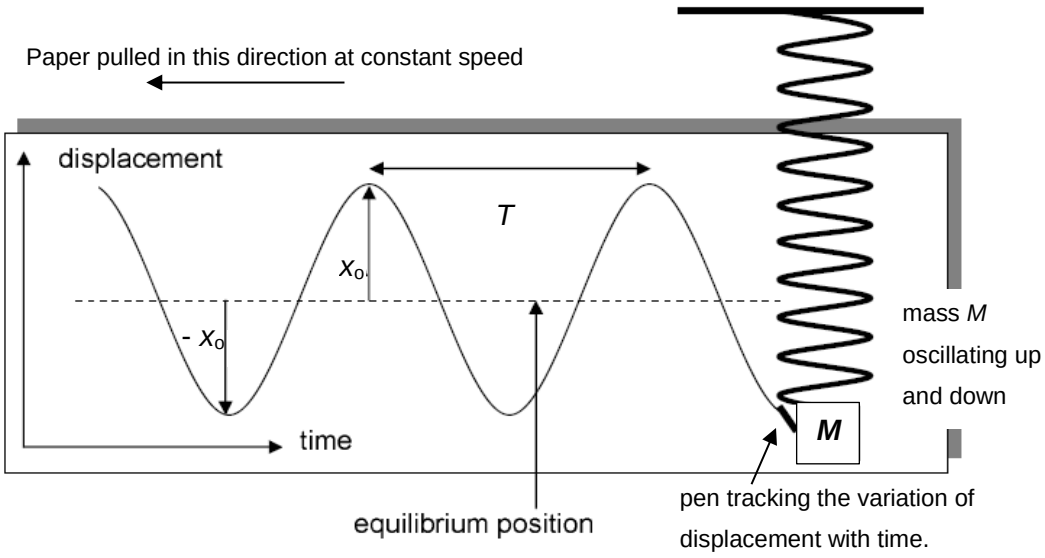


Fig. 8.1. Tracking the oscillation of a vertical spring-mass system

watch MIT Physics Demo -- Spray Paint Oscillator https://youtu.be/P-Umre5Np_0

8.1.1 Terms used to describe SHM (with reference to Fig 8.1)

Terms	Definitions
Displacement, x	The displacement *x* is the linear distance of the mass from its equilibrium position (*x* = 0) in a specified direction. It is a **vector**; the + / – sign tells the direction of displacement from equilibrium. If you set the direction for displacement above equilibrium to be positive, then displacement below the equilibrium would be negative and *vice versa*.
Equilibrium Position (x = 0)	When the oscillating mass is at equilibrium position, the **net force acting on the mass is zero.**
Amplitude, x_o	It is the **magnitude** of the maximum displacement of the mass from the equilibrium position. It is a scalar.
Period, T	It is the time taken for the mass to complete one cycle of oscillation. S.I. unit: second.
Frequency, f	It is the number of cycles per unit time: $f = \frac{1}{T}$. S.I. unit: hertz (Hz). 1 Hz = 1 s^{-1}.
Angular frequency, ω	It is the rate of change of the ***phase angle*** (introduced in Section 8.2) with respect to time: $\omega = \frac{d\theta}{dt}$. S.I. unit: rad s^{-1}. A cycle of oscillation is associated with a phase angle of 2π rad, $\omega = \frac{2\pi}{T} = 2\pi f$.

8.1.2 SHM - Displacement-time (*x*-*t*) relationship

From Fig. 8.1, we see that the time variation of the displacement of the mass obtained in the experiment is remarkably similar to a sine graph as shown in Fig 8.2.

The equation of such a graph can be written as follows,

$$x = x_0 \sin \omega t$$

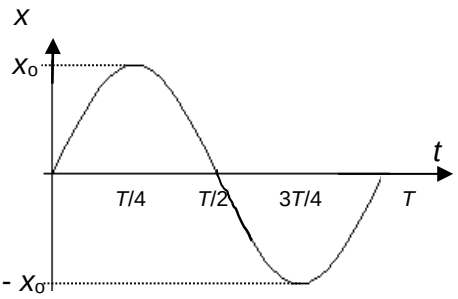
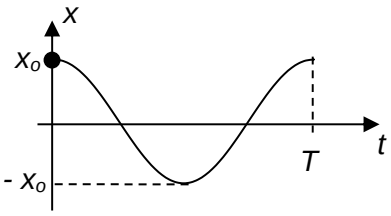
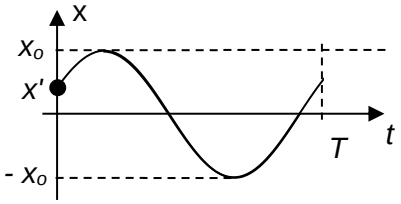


Fig. 8.2. Displacement-time graph of mass

where $\omega = \dfrac{2\pi}{T}$ is known as the ***angular frequency*** and x_o is the ***amplitude*** of the oscillation.

We can see from this graph that when we start timing, the displacement of the mass is zero, which means that the mass is at its equilibrium position at that instant. If the mass is not at the equilibrium position but at some other location at the instant when we start timing, we can adjust the equation easily to fit the graph obtained in the experiment.

Position of mass when timing starts ($t = 0$)	**Displacement-time (*x*- *t*) Graph**	**Displacement-time Equation**
Mass is *released from rest* at *maximum displacement* from equilibrium position.		$x = x_0 \cos \omega t$ where x_o is the amplitude.
Mass is moving with *initial velocity v* at *displacement x'*		$x = x_0 \sin(\omega t + \phi_1)$ or $x = x_0 \cos(\omega t + \phi_2)$ where ϕ is the initial phase angle (not in syllabus)

Note: ωt is in radians. Remember to set your calculator to **RADIAN mode** before calculation!

8.1.3 SHM - Velocity-time (*v*-*t*) relationship and acceleration-time (*a*-*t*) relationship

From the topic of kinematics, we know that from the displacement-time relationship of an object, we can obtain its corresponding velocity-time and acceleration-time relationship by differentiating the equations or analysing the gradients of the tangents at the various points on the graphs accordingly (see Fig. 8.3).

We can similarly do so for SHM.

The table illustrates two examples.

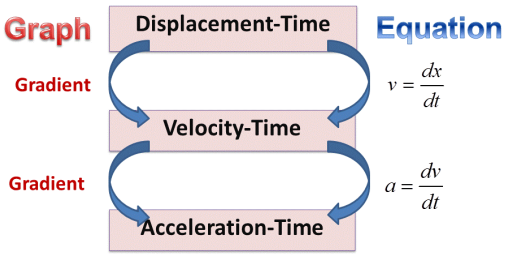
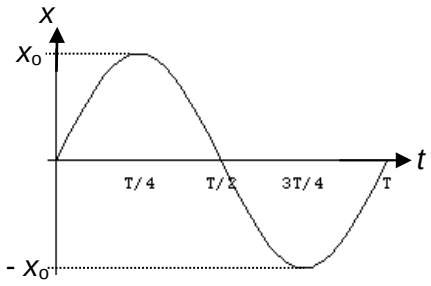
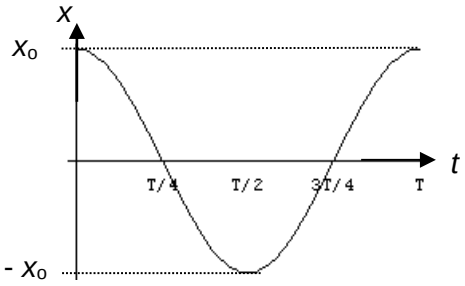
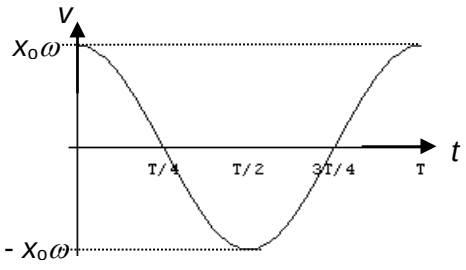
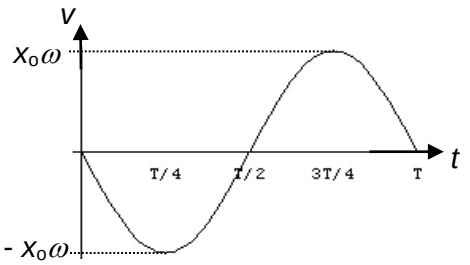
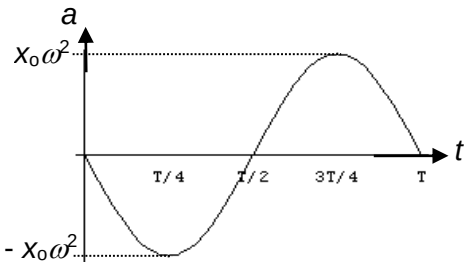
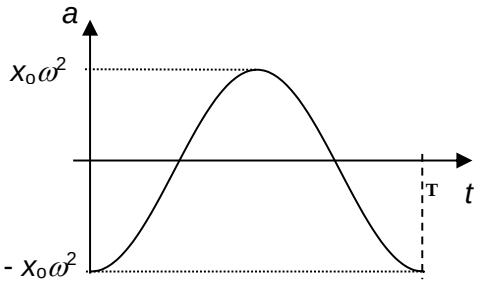


Fig 8.3

***t* = 0 (initial conditions)**	**Example 1: *x* = 0, *v* is at maximum value**	**Example 2: *x* = x_o, *v* = 0**
***x-t* graph**		
Equation	$x = x_0 \sin \omega t$	$x = x_0 \cos \omega t$
***v-t* graph**		
Equation	$v = \frac{dx}{dt} = \omega x_0 \cos \omega t$	$v = \frac{dx}{dt} = -\omega x_0 \sin \omega t$
***a-t* graph**		
Equation	$a = \frac{dv}{dt} = -\omega^2 x_0 \sin \omega t$	$a = \frac{dv}{dt} = -\omega^2 x_0 \cos \omega t$

From the graphs and the equations above, together with the fact that $-1 \le \sin\theta \le 1$ and $-1 \le \cos\theta \le 1$,

we can see that:

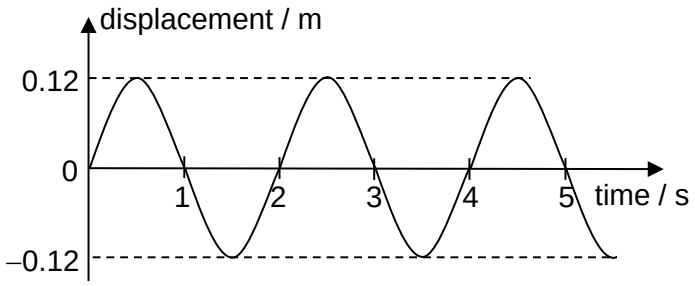
(1) Maximum value of velocity, v_o: $\boxed{v_0 = \omega x_0}$

(2) Maximum value of acceleration, a_o: $\boxed{a_0 = \omega^2 x_0}$

where ω is the angular frequency and x_o is the amplitude of the oscillation.

Example 8.2: Describing SHM in terms of time

The pendulum bob in a particular clock oscillates so that its displacement from a fixed point is as shown.



(a) By taking the necessary readings from the graph, determine for this simple harmonic oscillator

(i) the amplitude x_o

(ii) the angular frequency, ω

(iii) the maximum velocity v_o and

(iv) the maximum acceleration a_o

(b) Hence, write down the equation that describes the displacement-time relationship.

(c) Sketch labelled graphs, illustrating the corresponding time variation of the oscillator's

(i) velocity,

(ii) acceleration.

Example 8.3: Displacement-time Relationship for SHM

The displacement x (in m) of a body moving in SHM is given by the equation

$$x = 3.5 \sin 4.0t$$

a) What is the amplitude of the motion?

b) What is the angular frequency?

c) What will be the displacement 0.20 s after oscillation has begun?

d) Determine the velocity at time t = 0.50 s.

8.1.4 Definition of SHM – relationship between acceleration and displacement

Comparing the *x*-*t* and *a*-*t* relationship for both the examples in the table in section 8.1.3, we see that $a = -\omega^2 x$. This is actually the defining equation for SHM!

Definition of Simple Harmonic Motion

Simple harmonic motion is a periodic motion in which the **acceleration** of the body

- is **directly proportional** to the **displacement from its equilibrium point**, and
- is always in the **opposite direction to the displacement.**

$$a = -\omega^2 x$$

where ω is the angular frequency and *x* is the displacement from the equilibrium point.

Notes:

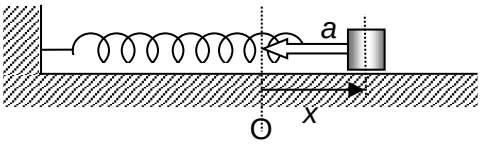
(1) The **negative sign** in the equation indicates that **the acceleration in SHM is always in the direction opposite to the displacement from the equilibrium point**; in other words, the acceleration is always directed towards the equilibrium point.

(2) Since $a = \frac{d^2x}{dt^2}$, we can write $\frac{d^2x}{dt^2} = -\omega^2 x$. This is a 2nd order differential equation and $x = x_0 \sin(\omega t + \phi)$ is a general solution where ϕ depends on the position of the oscillator when observation starts.

You are not required to know how to solve differential equations for the A-level Physics syllabus. However, by differentiating the expression $x = x_0 \sin(\omega t + \phi)$ twice, you can verify that $\frac{d^2x}{dt^2} = -\omega^2 x$.

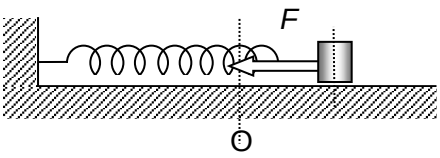
Case Study: SHM of Horizontal Spring-Mass System Explained Using Dynamics

Let us consider a mass oscillating at the end of a horizontal light spring. At the start of section 8.1, we said that the spring-mass system is an example of SHM. We shall now prove that indeed, the mass satisfies the defining equation of SHM. Ideally, the spring obeys Hooke's Law and the mass of the spring is negligible, the mass slides without friction on the horizontal surface. Point O is the equilibrium position of the mass, where the spring is unstretched.

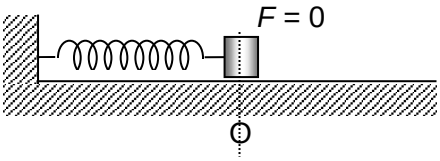


Let us now look at how the restoring force *F* exerted by the spring on the mass varies with displacement *x*:

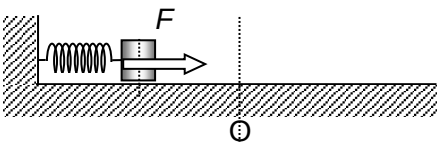
(a) when mass is displaced to the right $(x > 0)$, causing the spring to be stretched, *F* is towards the left.



(b) when the mass is at its equilibrium position $(x = 0)$ (spring is unstretched), *F* is zero.



(c) when the mass is displaced to the left $(x < 0)$, causing the spring to be compressed, *F* is towards the right.



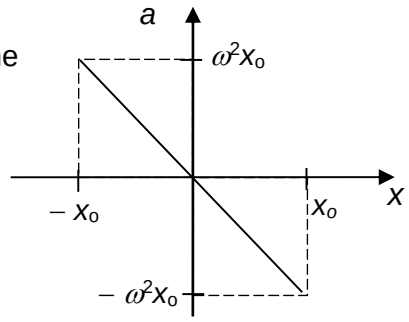
Hence, we see that

1. the **resultant restoring force is always towards the equilibrium position.**
2. the **direction of this resultant force is always opposite to that of the mass's displacement** at all points, except at equilibrium position, where $a = x = 0$.
3. According to Hooke's Law, $F = -kx$, where *k* is the spring constant. The restoring force, *F*, is proportional to the displacement, *x*, but directed opposite to it (as indicated by the negative sign).

   By Newton's 2nd Law, $F_{net} = ma$

   Since F_{net} is provided by the restoring force of the spring acting on the oscillating mass *m*,

   $$ma = -kx$$
   $$a = -\frac{k}{m}x$$

   Hence, the **acceleration is**
   - **opposite in direction to the displacement** and
   - **proportional to the displacement, with proportionality constant -*k* / *m*.**

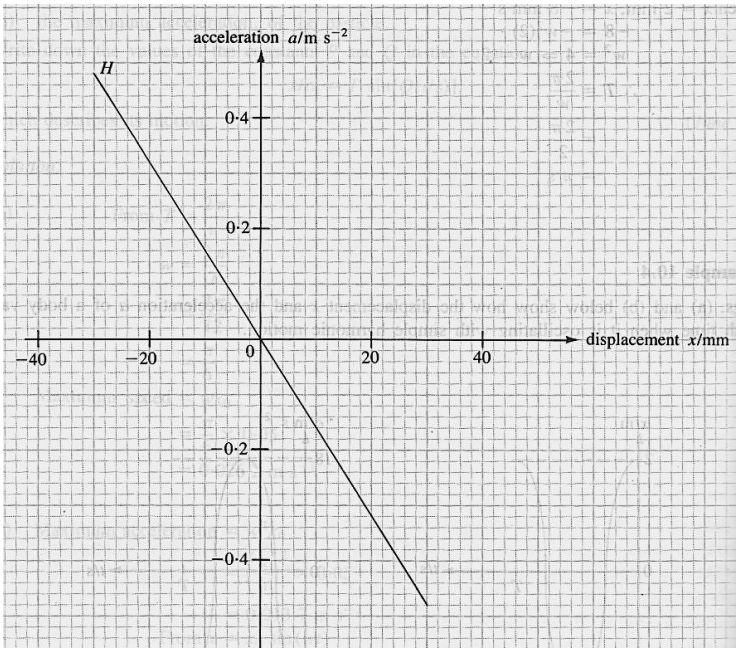
   which is the definition of SHM!

4. Comparing the expression above with the SHM equation $a = -\omega^2 x : \omega^2 = \frac{k}{m} \rightarrow T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}}$

   This is the expression we quoted for the period of the spring-mass system earlier. Note that the period of the oscillation depends on the physical properties of the system.

Example 8.4: Relationship between acceleration and displacement for SHM

The graph represents the motion of a particle.



Find

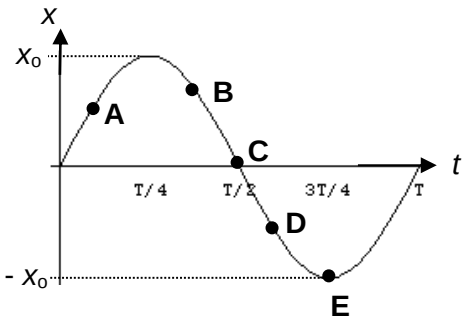
(a) the amplitude of oscillation,

(b) the angular frequency of oscillation,

(c) the period of oscillation.

Example 8.5: Relationship between acceleration and displacement for SHM

For each box in the table below, indicate whether the direction of the displacement *x,* velocity *v* and acceleration *a* of each point on the displacement time graph is "+", "-" or "0".



	A	B	C	D	E
x					
v					
a					

8.1.5 Simple Harmonic Motion: velocity-displacement (*v*-*x*) relationship

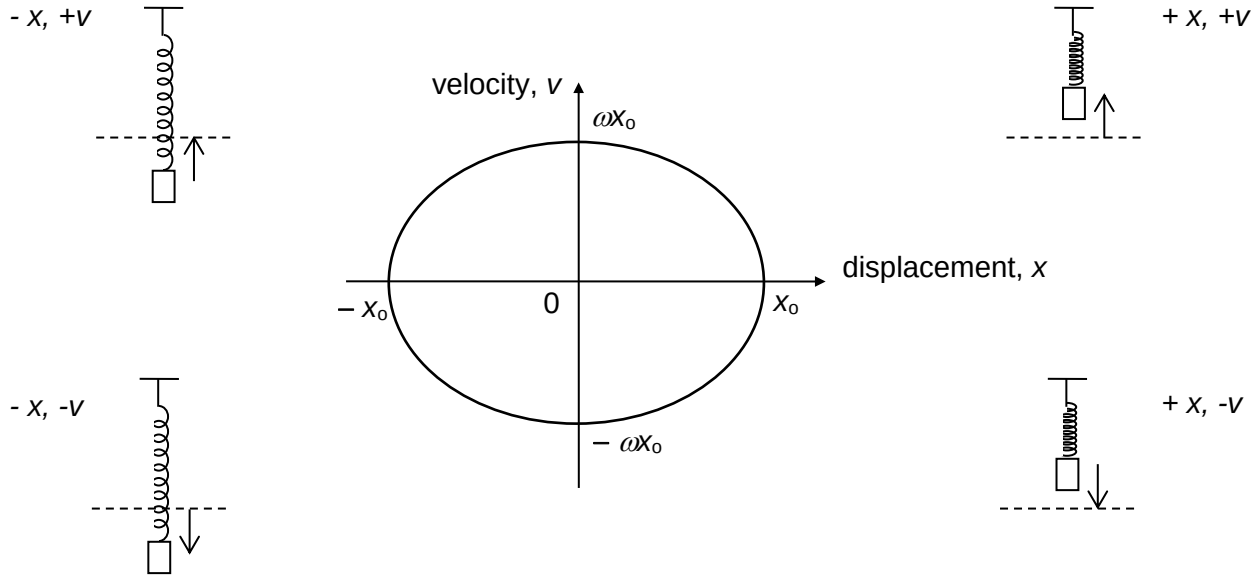
The velocity-displacement (*v*-*x*) relationship for SHM is given by

$$v = \pm\omega\sqrt{\left(x_0^2 - x^2\right)}$$

Note:

1) The equation is provided in the A-level Physics formulae list. It is applicable no matter where the body is when observation starts.
2) The "±" sign implies that the body in SHM always passes through each position back and forth with the same speed.
3) When the body is at maximum displacement $\left(x = x_0\right)$, v = 0.

   At the equilibrium position $\left(x = 0\right)$, the speed is maximum, $v_0 = \omega x_0$.
4) The *v*-*x* graph is an **ellipse** as shown.



Derivation of the relationship between velocity and displacement (not in syllabus)

Consider $x = x_0 \sin\left(\omega t + \phi\right)$.....(1)

Differentiate (1) with respect to t : $v = \dfrac{dx}{dt} = \omega x_0 \cos\left(\omega t + \phi\right)$ …….. (2)

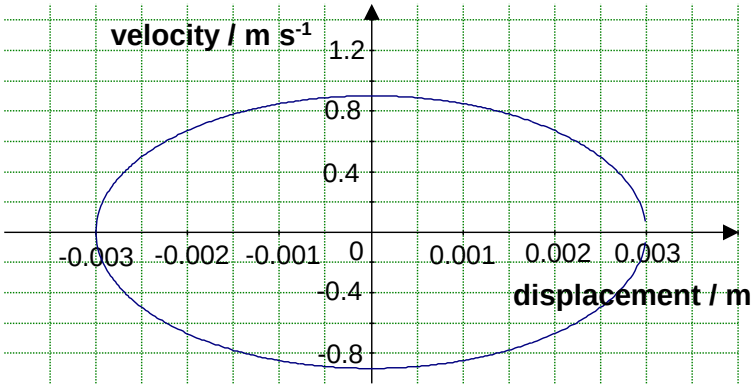
Given that $\sin^2\theta + \cos^2\theta = 1$, $\cos\theta = \pm\sqrt{1 - \sin^2\theta}$

Substitute into equation (2): $v = \omega x_0\left[\pm\sqrt{1 - \sin^2\left(\omega t + \phi\right)}\right]$

$$\Rightarrow v = \pm\omega\sqrt{x_0^2 - x_0^2\sin^2\left(\omega t + \phi\right)}$$

$$\Rightarrow v = \pm\omega\sqrt{x_0^2 - x^2}$$

Example 8.6: Relationship between velocity and displacement for SHM



(a) Explain why there are two values of velocity for zero displacement.
(b) Explain why there are two values of displacement for zero velocity.
(c) Write down the equation, which describes the graph.

Solution:

Example 8.7: Relationship between velocity and displacement for SHM

An object moving with simple harmonic motion has an amplitude of 0.020 m and a frequency of 20 Hz. Calculate

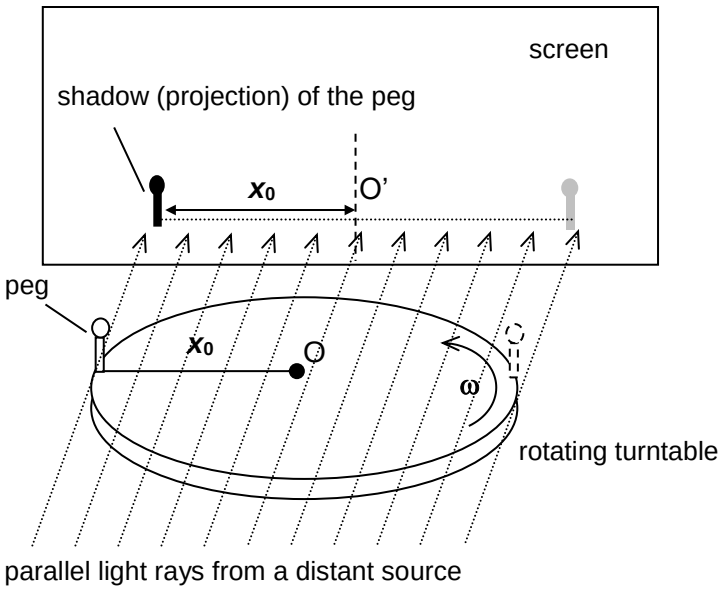
(a) the period of oscillation.
(b) the velocities at the equilibrium point and the position of maximum negative displacement.

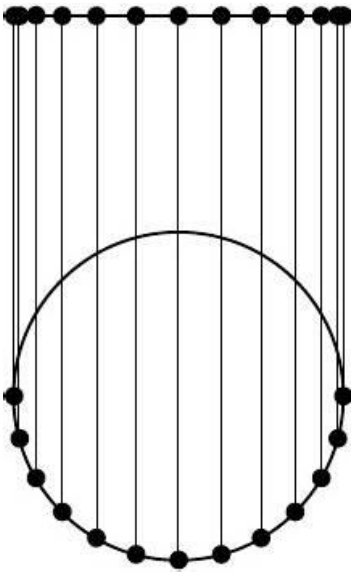
Solution:

8.2 Relationship between Simple Harmonic Motion and Uniform Circular Motion

We have thus far established that for SHM, $x = x_0 \sin \omega t$ where ω is a constant known as the angular frequency, it is also known as the rate of change in phase angle θ. Let us look at an experiment to gain a better understanding of the concept of phase angle.

Consider a peg on a turntable. Parallel light rays from a distant source are shone on the turntable. As the turntable rotates, the peg moves in uniform circular motion and the shadow of the peg against a vertical screen will move back and forth. As it turns out, we can prove that the motion of its shadow is SHM.





Some observations

1) The motion is symmetrical about the fixed point, O' on the screen.

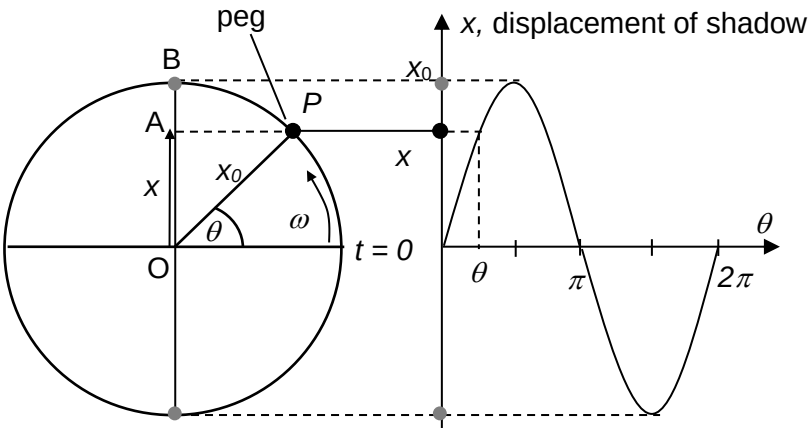
2) The time taken to complete one oscillation (period) is the same as the time taken to complete one cycle. Thus frequency of SHM = frequency of the circular motion.

3) The distance from the fixed point O' to the extremity (amplitude), is equal to the radius of the circle.

4) When the peg is shifted nearer to the centre of the turntable, the amplitude of the SHM decreases, but the period or frequency remains unchanged.

5) The shadow reaches maximum speed as it passes O' and is momentarily at rest at the extreme positions.

Consider the following diagram where the OA is the projection of OP along the vertical line OB. P represents the peg and A represents its shadow on the screen OB. As the turntable rotates, P moves in a circle of radius $OP = OB$ with constant angular speed ω. The angle θ it makes with reference to the equilibrium position is thus given by $\theta = \omega t$.



top view of turntable

Considering the right angle triangle OAP, we can easily see that the displacement *x* of the peg's shadow A with respect to O is given by

$$x = x_0 \sin\theta = x_0 \sin\omega t$$

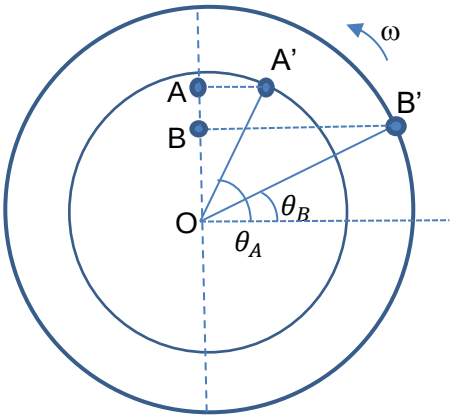
Watch video clip on SHM & uniform CM relationship & phase difference
https://youtu.be/d0p7vDIgqjU

Hence the shadow of the peg does indeed move in SHM.

Phase & Phase Difference

For any object in SHM, we may associate with it the motion of an imaginary peg moving in uniform circular motion such that at any point in time, the angle θ it makes with reference to the equilibrium position is called the **phase (angle)** of oscillation.

Thus the **angular frequency ω** in a simple harmonic motion is also **the rate of change of phase angle** of the imaginary peg.

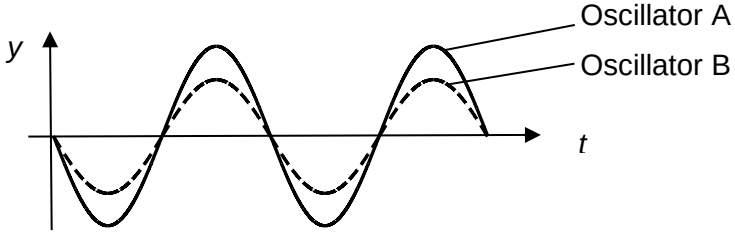


A and B are moving in SHM about O on the vertical dotted line. A is ahead of B by a phase difference of $\Delta\theta = \theta_A - \theta_B$. A and B have different amplitudes but same frequency.

The concept of phase angle is useful for <u>comparing two oscillators of the same frequency</u> (and therefore period), even if they have different amplitudes. We can describe how far ahead one is compared to the other in terms of the difference in their phase angles.

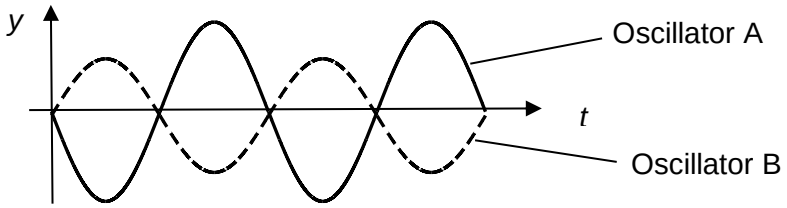
1. The difference in phase angle between two oscillations is called **phase difference $\Delta\phi$**.

2. Two oscillations are said to be ***in phase*** when they are in step, i.e., they reach their respective maximum displacements in the same direction at the same time. It does not matter what their respective amplitudes are.



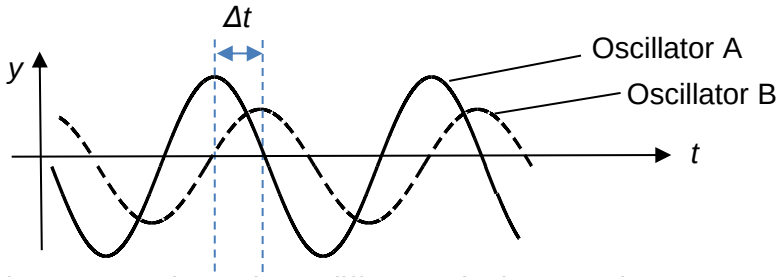
The phase difference $\Delta\phi = 0$ rad or any integral multiple of 2π rad.

3. Two oscillations are said to be ***in anti-phase*** when they are completely out of step, i.e., one oscillation reaches maximum displacement in a given direction at the same instant as the other oscillation reaches maximum displacement in the opposite direction.



The phase difference $\Delta\phi = \pi$ rad or any odd integral multiple of π rad.

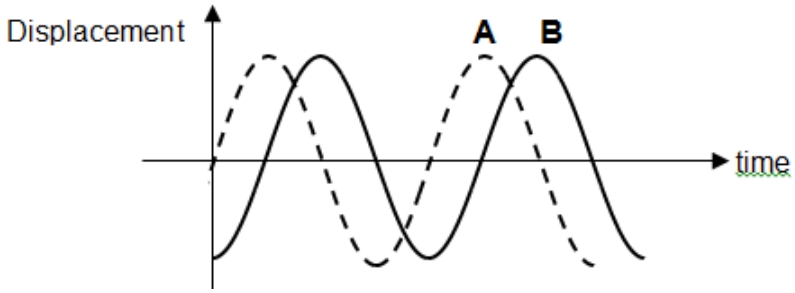
4. Of course the two oscillations may not be in phase or in anti-phase but at some intermediate phase difference. In such a case, the phase difference can be calculated by the following: $\Delta\phi = \frac{\Delta t}{T} \times 2\pi$.



You will learn more about phase difference in the next chapter on Waves.

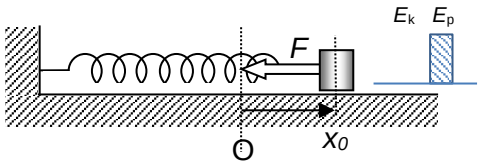
Example 8.8: Phase and Phase Difference

The displacement-time graph of two objects A and B oscillating is shown below. By how much does A lead B in terms of phase angle?

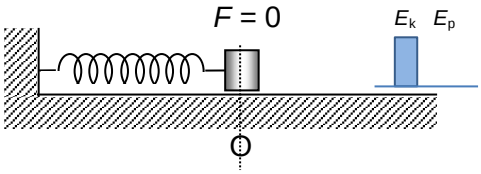


Solution:

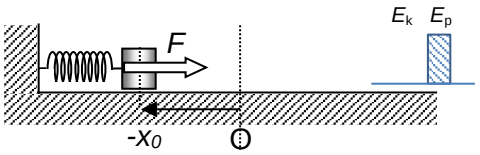
8.3 Energy in SHM



Consider a horizontal spring-mass system on a smooth horizontal plane, where the mass is initially at rest. In order for the mass to oscillate in SHM, we need to displace the mass, say to the right from its equilibrium position O. As we do that, we store elastic potential energy, E_p, in the spring.



When we release the mass, the restoring force, F, in the spring will accelerate the mass towards its equilibrium position. As the restoring force is in the direction of motion, it does *positive work*; the initially stored elastic potential energy, E_p, transforms into kinetic energy, E_k. The kinetic energy reaches a maximum value when the restoring force (that is proportional to the displacement) becomes zero at the equilibrium position.



As the mass passes its equilibrium position and moves to the left, the restoring force now acts in the opposite direction to slow it down. The restoring force now does *negative work* on the mass, converting its kinetic energy back to elastic potential energy. The elastic potential energy reaches a maximum value (which is the same as the amount that was stored initially at the start of SHM) when the mass comes to rest momentarily at its extreme end, where the restoring force is maximum.

This repeats throughout SHM where the work done by the restoring force transfers energy to and fro between potential energy stored in the system and kinetic energy of the mass. The total amount of energy involved in this exchange is the mechanical energy, which is also referred to as the total energy associated with the SHM, and is equal to E_T, the amount that was initially invested when we first displaced the mass from its equilibrium position.

8.3.1 Energy-displacement Graphs of SHM

Let us now look at the energy-displacement graphs of the horizontal spring-mass system. We assume that the spring is light (zero mass) so that it may not possess kinetic energy E_k. Hence the total energy E_T associated with the SHM is at any point in time the sum of elastic potential energy stored in the spring and the kinetic energy of the mass.

For SHM, $v = \pm\omega\sqrt{x_0^2 - x^2}$

hence

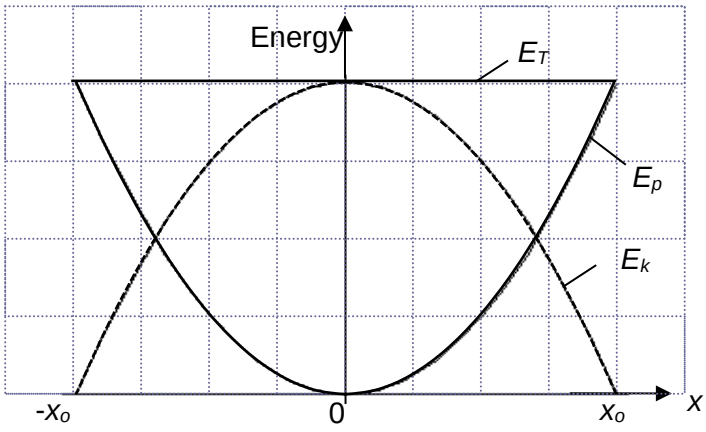
$$E_K = \frac{1}{2}mv^2 = \frac{1}{2}m\omega^2(x_0^2 - x^2)$$

At equilibrium position $(x = 0)$

$$E_T = E_{K,max} = \frac{1}{2}m\omega^2 x_0^2$$

As E_T is constant,

$$E_p = E_T - E_K = \frac{1}{2}m\omega^2 x^2$$



In some systems there are more than one type of potential energy present in the system. For example, in the vertical spring mass system, the potential energy E_p will then refer to the sum of gravitational potential energy and elastic potential energy. Scan the QR code to watch the energy analysis. (https://youtu.be/FKltat9EHqw)

Example 8.9: Relationship between Energies and Displacement

A mass, m = 0.50 kg, connected to a light spring with spring constant k = 20 N m^{-1} oscillates on a frictionless horizontal surface with an amplitude of 3.0 cm. Given that angular frequency $\omega = \sqrt{\frac{k}{m}}$, calculate the

(a) total energy,

(b) (i) kinetic energy of the system when the displacement is 2.0 cm and

(ii) potential energy of the system when the displacement is 2.0 cm.

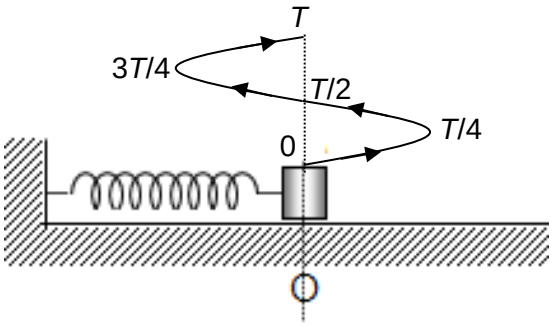
Solution:

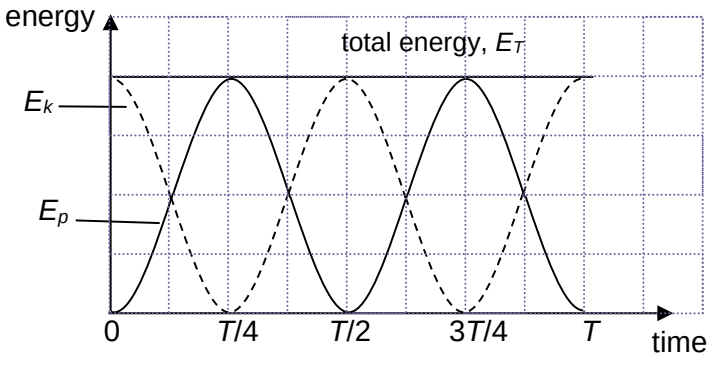
8.3.2 Energy-time Graphs of SHM

The energy-time graphs depend very much on the position and the velocity of the mass at the start of our observation. Suppose the mass is at its equilibrium position when our observation starts ($t = 0$, $x = 0$, $v = v_o$).

$\Rightarrow$ $x = x_0 \sin(\omega t)$ and $v = v_0 \cos(\omega t)$.

$\Rightarrow$ $E_k = \frac{1}{2}mv^2 = \frac{1}{2}mx_0^2\omega^2 \cos^2 \omega t$ $\quad$ $E_p = \frac{1}{2}m\omega^2x^2 = \frac{1}{2}m\omega^2x_0^2 \sin^2 \omega t$





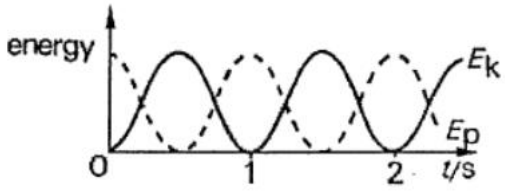
Note that when the oscillatory motion has completed one full cycle in one period, both kinetic and potential energies have gone through two cycles.

Example 8.10: Relationships between Energies and Time

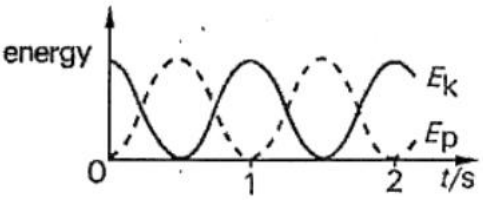
The bob of a simple pendulum of period 2 s is given a small displacement and then released at time $t = 0$.

Which diagram shows the variation with time of the bob's kinetic energy E_k and its potential energy E_p?

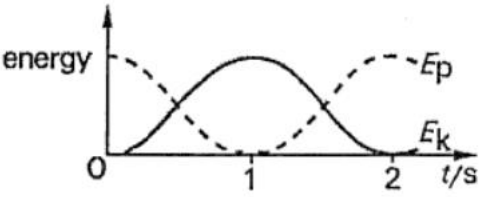
A



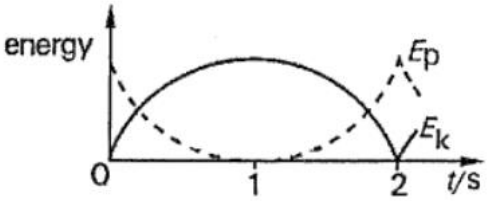
B



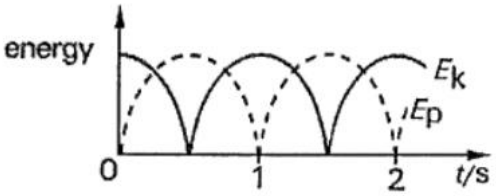
C



D



E



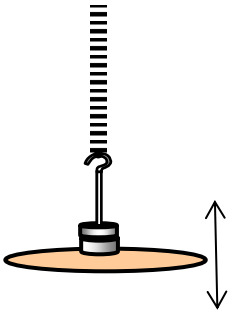
8.4 Damped Oscillations, Forced Oscillations and Resonance

8.4.1 Damped oscillations

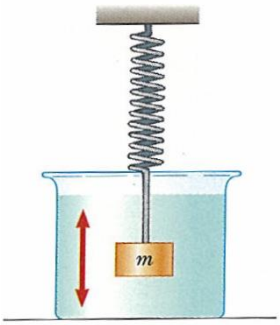
In an ideal oscillating system, there is no friction and the total energy associated with the oscillation is constant. The oscillation continues forever, with no decrease in amplitude.

In real-world systems, there is friction and the total energy associated with the oscillations decreases with time due to the work done against friction. The progressive decrease in amplitude of any oscillatory motion caused by dissipative forces is called ***damping*****.** Such an oscillatory motion is known as a ***damped oscillation*****.**

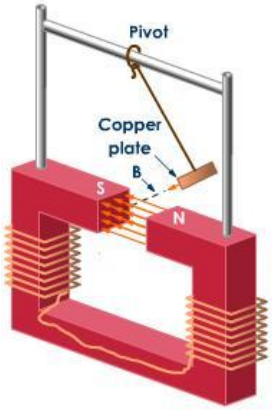
Commonly used damping methods:



A) Attach a piece of cardboard of negligible mass but large surface area to the base of the mass. Air resistance dissipates energy.



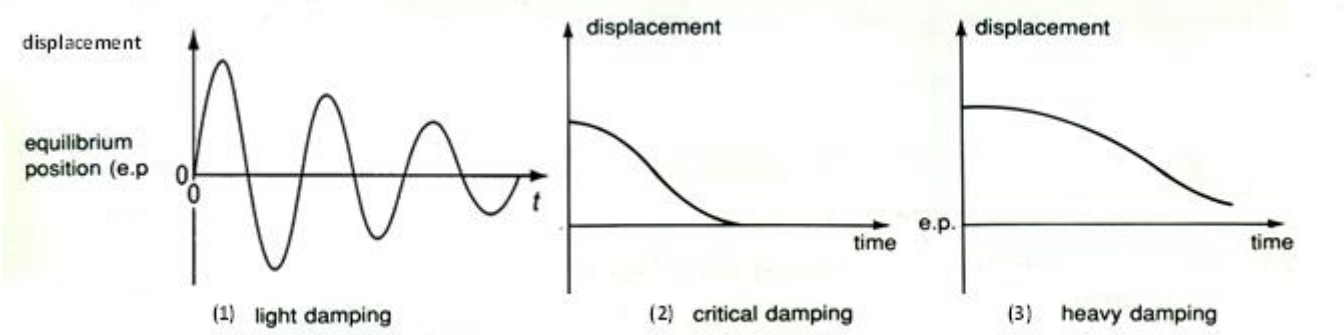
B) Submerge the mass in a viscous fluid. Viscous forces dissipate energy.



C) Swing the copper plate in a magnetic field so that eddy currents are induced in the plate. Eddy currents dissipate energy (will be discussed in Electromagnetic Induction in year 2).

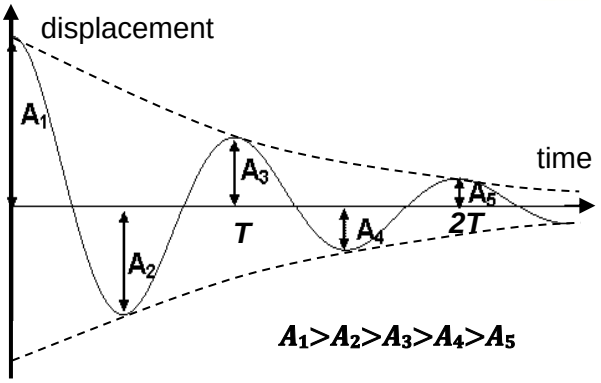
Extent of damping

Damping changes the behaviour of SHM systems. Certain features of the oscillation (amplitude etc.) are dependent on the extent of damping.



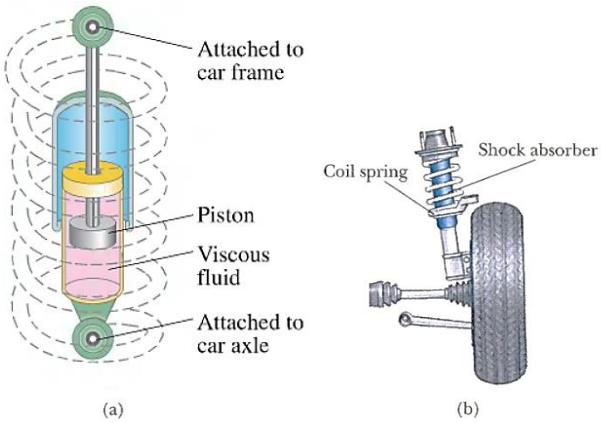
(1) *Light Damping* (*underdamped system*)

- In a lightly damped system, the total energy of the system decreases with time as its energy is dissipated when the system oscillates against resistive forces.
- The period/frequency remains constant, but the amplitude of the system decreases exponentially with time.



(2) *Critical Damping (critically damped system)*

- If we are able to gradually increase the resistive force acting on an oscillatory system, damping increases till a point such that when displaced, the system **returns to the equilibrium position in the shortest possible time without oscillating** at all. The system is said to be **critically damped**.
- Critical damping is deliberately introduced into some systems to prevent continuous oscillations. For example, in analogue instruments such as balances, ammeters and voltmeters, the pointer is critically damped so that it stops in the shortest possible time to indicate the reading, instead of oscillating about the reading (underdamped) or taking an unnecessarily long time to crawl to the reading (overdamped).
- Critical damping is also used in **car suspension systems** to ensure a smooth ride when the car moves on a bumpy road. The shock absorbers in the suspension system have hydraulic pistons to prevent the car from bouncing up and down excessively after hitting a bump.



(a) Spring and shock absorber provides damping so that a car will not bounce up and down endlessly.

(b) A spring and shock absorber attached to one of the wheels in a car suspension system.

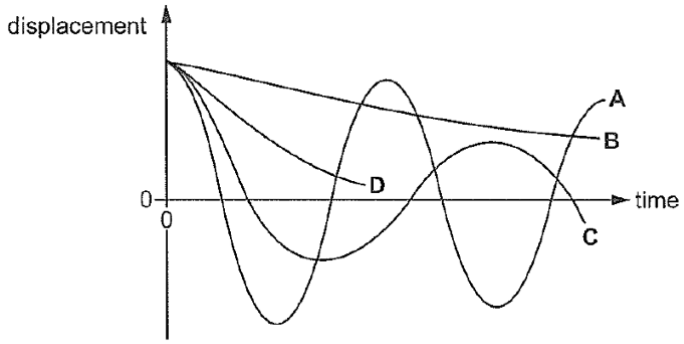
(3) *Heavy Damping* (*overdamped system*)

- When the damping force is increased beyond the point of critical damping, we say that the system is **overdamped** or that it experiences heavy damping. Once displaced from its equilibrium, the oscillator takes a very long time to return to its equilibrium position.

Example 8.11: Car Suspension System (N17/I/13)

The diagram shows four displacement-time curves for oscillations with various degrees of damping.

Which curve would ideally characterize a good car suspension system?



8.4.2 Forced oscillations and resonance

Free oscillations

When a system oscillates about its equilibrium position with its own natural frequency and under no external influence, other than the impulse that initiated the motion, the system is said to be in a ***free oscillation***.

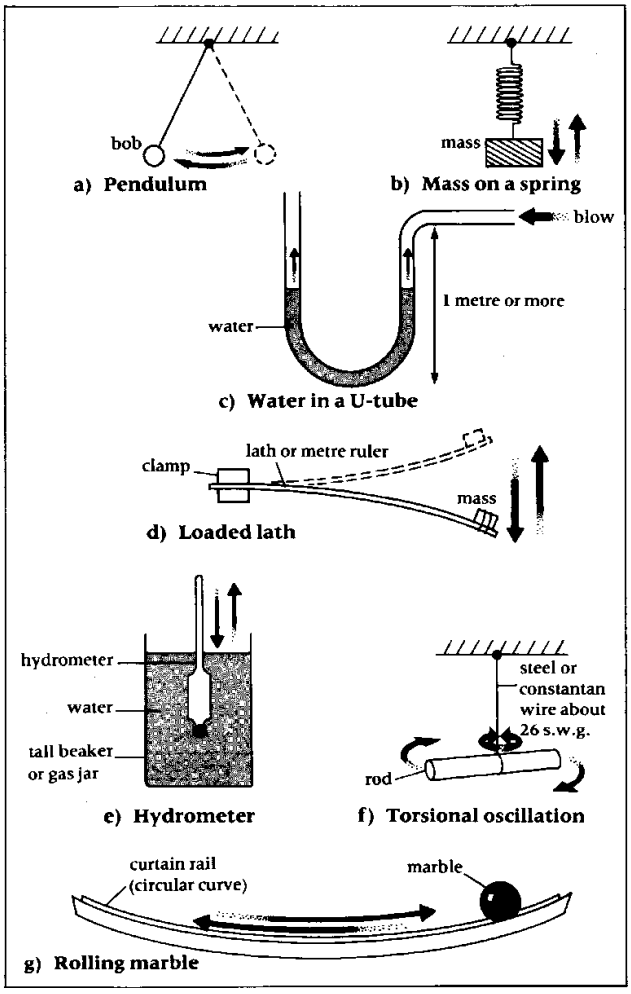
Forced oscillations

A real-world system in a free oscillation always experiences damping and will eventually stop oscillating. To keep it oscillating, one needs to periodically put in energy to the system to make up for the energy lost. In other words, one needs to apply a periodic force to the system.

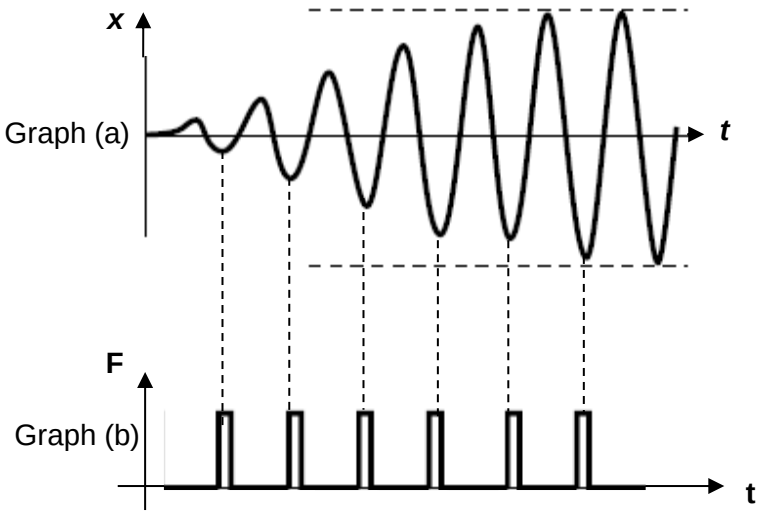
- An oscillation under the influence of an external periodic force is called a ***forced oscillation***.
- The external force that is used to sustain (or drive) the oscillations is known as the **driving force.**
- The frequency with which the periodic force is applied is called the **forced frequency** or **driver frequency**.

Some examples of free oscillations

(the arrows indicate motion, not forces)



One example of forced oscillation is pushing a child on a swing at the playground; one needs to continually give the swing a push at the right moment to sustain its oscillation.



Graph (a) shows the child's *x*-*t* graph as one periodically transfers a fixed amount of energy into the swinging motion.

Graph (b) shows the variation of driving force with time.

Resonance

Consider the above example of a child on a swing. The child will need to be pushed to keep him swinging.

If the frequency of the driving force matches the natural frequency of the oscillatory motion of the child-swing system, an interesting phenomenon happens - the final amplitude of the oscillation would be the greatest compared to the amplitudes of oscillation for other frequencies of driving force. This phenomenon is known as ***resonance***.

Resonance is a phenomenon in which an oscillatory system responds with maximum amplitude to an external periodic driving force, when the frequency of the driving force (driver frequency) equals the natural frequency of the driven system.

Note:

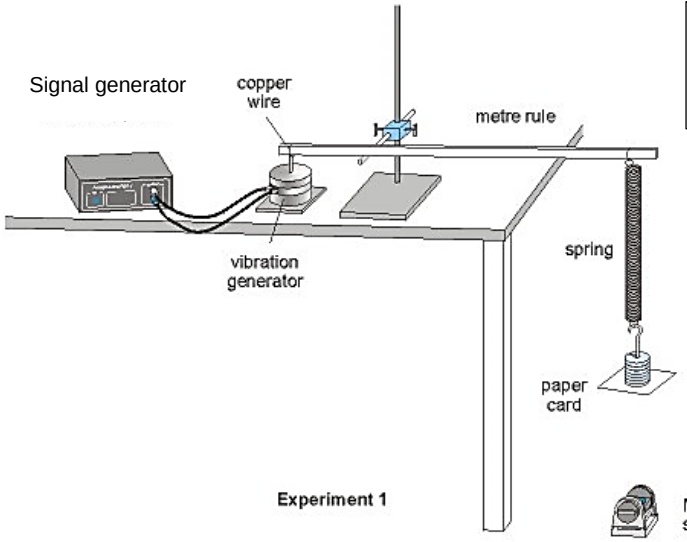
(1) The frequency at which resonance occurs is known as the ***resonant frequency***.

(2) Under ideal conditions, i.e., if no resistive forces are present (there is no damping), the amplitude will build up to an infinitely large value.

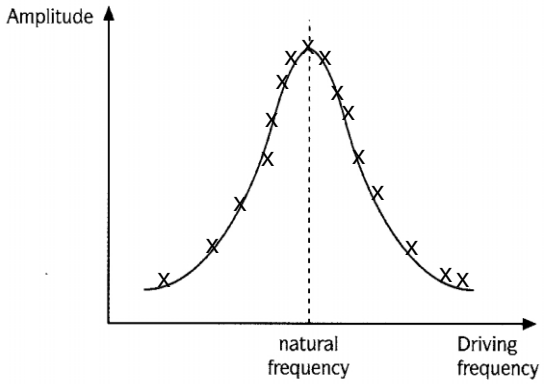
(3) However, in real cases, damping occurs. Whilst putting energy into the system in each cycle, energy is also taken away by damping and the amplitude will stop increasing when the rate of energy supplied is the same as the rate at which it is lost.

(4) Resonance represents the point where (driving) energy is transferred most efficiently into the oscillatory system.

Investigating Resonance in the Lab



Scan the QR code to watch a similar experiment https://youtu.be/jewSVEBkI_s .



The above set-up may be used to investigate the phenomenon of resonance in the lab.

(1) The system under study is a spring-mass system that is driven to oscillate by a metre rule which is in turn driven to vibrate by a vibration generator. To vary the frequency of vibration, the vibration generator is connected to a signal generator.

(2) To vary the amount of damping experienced by the spring-mass system, a light paper card is attached to the mass. By varying the area of the paper card attached, the damping is varied. The larger the paper card, the greater the damping.

(3) To measure the amplitude of vibration, a motion sensor is placed directly below the paper card.

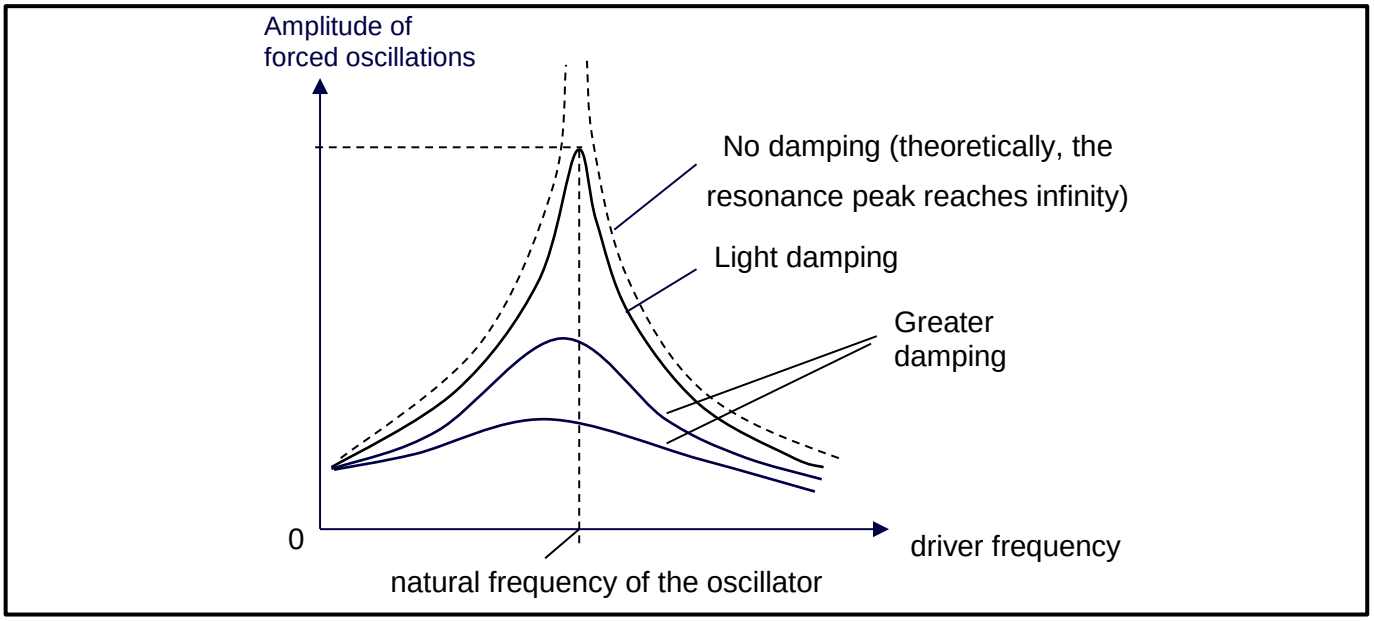
(4) For each frequency of vibration, we can then measure and record the amplitude of vibration of the spring-mass system. We start off with a driving frequency that is much lower than the natural frequency of the spring-mass system and slowly increase this frequency till beyond the natural frequency.

(5) The graph above shows the typical results obtained from the experiment for one area of card.

(6) The experiment is then repeated for different areas of cards in order to produce different amounts of damping.

(7) The data points for each different area of cards / different damping can then be plotted on the same pair of axes for comparison as shown below.

The figure below shows the amplitude vs. driving frequency graphs that will be obtained.



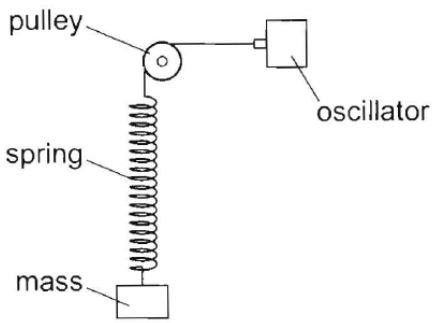
Note:

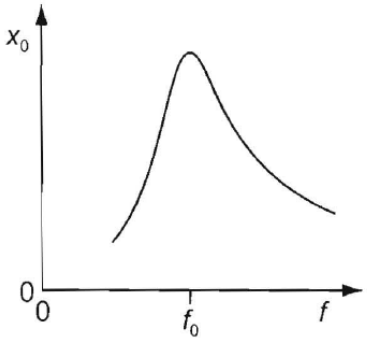
1) The amplitude of a lightly damped system is very large at resonance and the peak is sharp. The amplitude drops off rapidly when the driver frequency differs from the natural frequency of the system.
2) When the driver frequency approaches zero, the amplitude of the forced oscillations will be the same as the amplitude of the driver.
3) As the system experiences greater damping,
   - the amplitude of the oscillation **decreases**.
   - the resonance peak becomes **broader**.
   - the resonance peak shifts slightly to the **left** to a slightly lower value of frequency.

From the graphs, it is evident that damping is a useful way to reduce the effects of resonance. Another way to avoid resonance is to ensure that the frequency of the periodic external force differs from the natural frequency of the driven system.

Example 8.12: Damped Oscillation (N10/1/16)

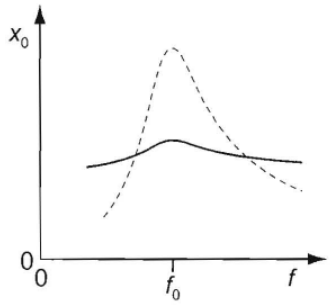
A mass, suspended from a helical spring, is made to oscillate. The graph shows the variation with frequency f of the amplitude x_0 of vibration of the mass.



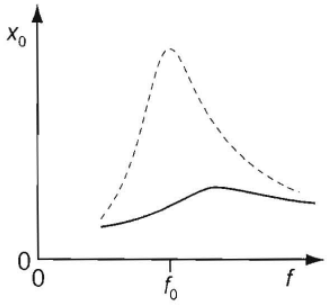


A sheet of cardboard of negligible mass is now fixed to the mass on the spring to cause light damping of the oscillations. Which graph shows how x_0 will vary with f over the same frequency range?

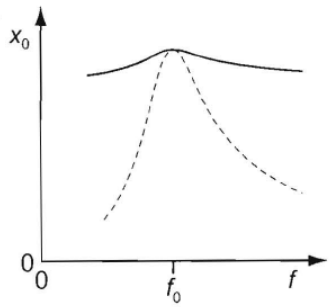
A



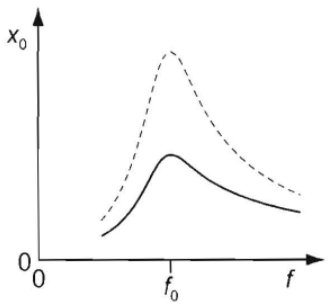
B



C



D



8.4.3 Examples of Destructive and Useful Resonance

Examples of destructive resonance

1) Shattering glass with voice

When an opera singer projects a high-pitched note whose frequency matches the natural frequency of a wine glass, resonance may cause the glass to vibrate at a large amplitude that it breaks.

2) Violent vibrations in vehicles and machines

If the panel in a bus rattles violently when the bus is travelling at a certain speed, it is likely that a resonant vibration is occurring. A washing machine with a load which has natural frequency matching the spinning frequency will vibrate violently as resonance occurs.

3) Collapse of some buildings in earthquakes

An earthquake consists of many low-frequency vibrations ranging from 1 to 10 Hz. During an earthquake, when the frequencies of the vibration match the natural frequencies of buildings, resonance may occur and result in serious damage. That also explains why some buildings collapse while others stand almost unaffected. In regions of the world where earthquakes happen regularly, buildings may be built on foundations that absorb the energy of the shock waves. In this way, the vibrations are damped and the amplitude of the oscillations cannot reach a dangerous level.

Examples of useful resonance

1) Radio receptions

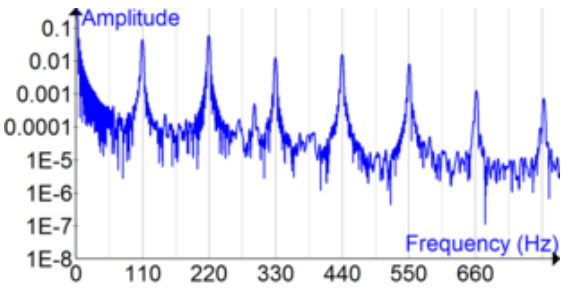
Resonance not only occurs in mechanical oscillations but also in electric circuits. A radio receiver works on the principle of resonance. Our air is filled with radio waves of many different frequencies which the aerial picks up. The tuner can be adjusted so that the frequency of the electrical oscillations in the circuits is the same as that of the radio wave transmitted from a particular station we desire. The effect of resonance amplifies the signals contained in this wave while the radio waves of other frequencies are diminished.

2) Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is increasingly used in medical diagnosis to produce images similar to those produced by X-rays. Strong, electromagnetic fields of varying radio frequencies are used to cause oscillations in atomic nuclei. When resonance occurs, energy is absorbed by the molecules. By analysing the pattern of energy absorption, a computer-generated image can be produced. The advantage of MRI scanners is that no ionising radiation (as in the process of producing X-ray images) is involved.

3) Musical Instruments

Most vibrating objects have more than one resonant frequency. The lowest resonant frequency of a vibrating object is called its fundamental frequency. A harmonic is defined as an integer multiple of the fundamental frequency. (You will learn more about it in the topic Superposition, under Stationary Wave). Strings and air columns used in musical instruments typically vibrate at harmonics of the fundamental frequency. In an acoustic guitar, when you pluck a string, the string initially vibrates at all the frequencies. Those frequencies that are not one of the resonances are quickly filtered out— or attenuated— and all that is left is the harmonic vibrations that are transmitted through air molecules as sound energy into the hollow wooden body of the guitar, making it (and the air inside) resonate and amplifying the musical note (making it considerably louder).



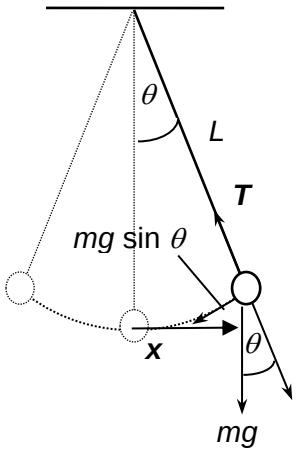
String resonance of a bass guitar A note with fundamental frequency of 110 Hz.

Appendix 1 Case study: Oscillations of A Simple Pendulum

A simple pendulum is in its midst of motion, where its bob is at a horizontal displacement x from its equilibrium position. The string is attached to a bob of mass m and makes a small angle of θ to the vertical as shown.

The restoring force ($mg \sin \theta$) which is pointing towards equilibrium position and is approximately horizontal, is in the opposite direction to the positive direction of the horizontal displacement,

$F_{net} = -mg \sin \theta$



When the angular displacement θ is small,

the horizontal displacement of the bob is approximately equal to

the arc length $x \approx L\theta$

$\Rightarrow \quad \theta \approx x/L$

and $\sin \theta \approx \theta \approx x/L$

$\Rightarrow \quad F_{net} = -mg \sin \theta \approx -mg\dfrac{x}{L}$

From Newton's 2nd law: $ma \approx -mg\dfrac{x}{L}$

$\Rightarrow \quad a \approx -\left(\dfrac{g}{L}\right)x$

i.e., $a = -\omega^2 x$ where $\omega^2 = \dfrac{g}{L}$

Hence, motion of a simple pendulum bob attached to a light inextensible string oscillating with **a small angular displacement of ~ 10° or smaller** is, to a good approximation, a linear SHM with period, T, given by

$$T = 2\pi\sqrt{\frac{L}{g}}$$

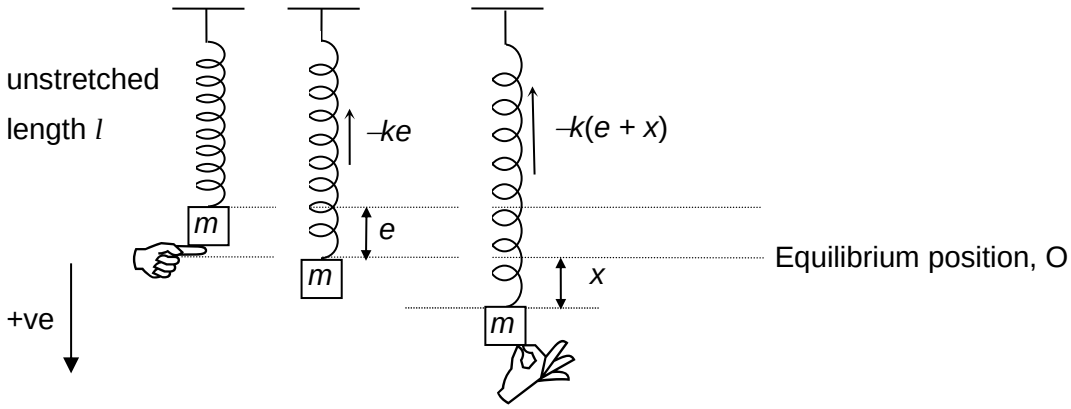
Note:

1) Other assumptions made are that the string is massless, the radius of the bob is small compared with the length of the string and air resistance is negligible.

2) The period T is independent of the amplitude of the oscillation.

Appendix 2 Case study: Oscillations of a Mass on Vertical Spring

A mass *m* attached to the end of a spiral spring exerts a downward force on it and stretches it by an amount *e* as shown in the diagram.



If the spring obeys Hooke's law, then the extension *e* is directly proportional to the extending tension, F_s,

i.e., $F_s = -ke$ where *k* is a constant.

Since the mass is in equilibrium, the net force on it is zero ($\sum F = 0$).

Thus $mg + (-ke) = 0$

$\Rightarrow$ $mg = ke$

Suppose the mass is now pulled down a further distance *x* below its equilibrium position, then the tension in the spring is increased to $-k(e + x)$. Thus the resultant force on the mass is given by

$$F = mg - k(e + x)$$
$$= ke - k(e + x) \quad \text{(since } mg = ke\text{)}$$
$$= -kx$$

i.e., the resultant force is acting upwards, with the direction opposite to that of the displacement.

By Newton's 2nd law, $ma = -kx$

$\Rightarrow$ $a = -\frac{k}{m}x$

Since *k* and *m* are fixed, the oscillation of a mass on spring is simple harmonic.

The angular frequency ω is given by $\omega^2 = \frac{k}{m}$

$\Rightarrow$ $T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}}$

Appendix 3 Equations of motion for damped simple harmonic oscillations

In an oscillatory system, the decrease in amplitude caused by dissipative forces is called damping and the corresponding motion is called damped oscillation. The simplest case to analyse is a simple harmonic oscillator under damping, where its dissipative force is directly proportional to its *velocity*. This behaviour occurs in friction involving the flow of viscous fluid, such as in shock absorbers or sliding between oil-lubricated surfaces. The additional force on the body due to viscous drag is $f = -bv$, where *b* is a constant.

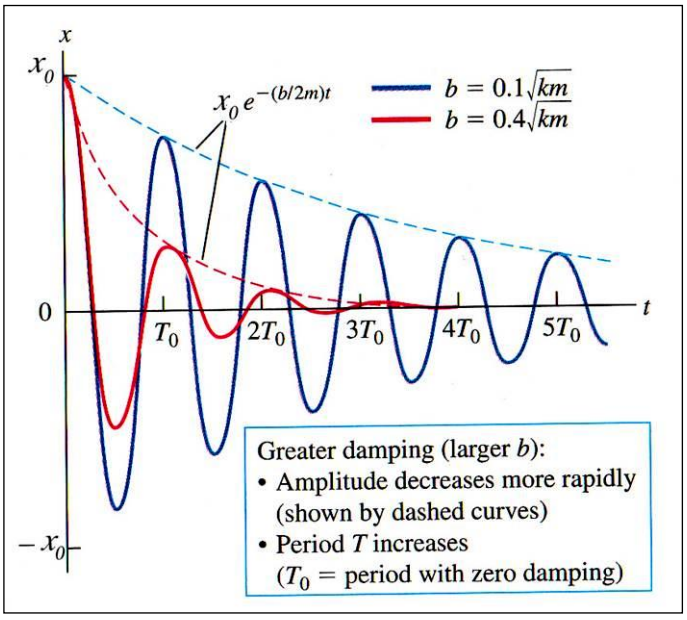
With a viscous drag, the net force on the body is $(-kx - bv)$.

Applying Newton's 2nd second law,

$$m\frac{d^2x}{dt^2} = -kx - bv$$

$$m\frac{d^2x}{dt^2} + b\frac{dx}{dt} + kx = 0$$

This is a second order linear differential equation with constant coefficients.



If $b < 2\sqrt{km}$, the system is under light damping and the equation of motion is

$$x = \left(x_0 e^{-\left(\frac{b}{2m}t\right)}\right)\cos\left(\left(\sqrt{\frac{k}{m} - \frac{b^2}{4m^2}}\right)t + \phi\right)$$

Its amplitude ($x_0 e^{-\left(\frac{b}{2m}t\right)}$) is not a constant, but decreases with time because of the decaying exponential factor. In addition, the larger the value of *b*, the faster the decrease in the amplitude.

Its angular frequency $\omega' = \sqrt{\frac{k}{m} - \frac{b^2}{4m^2}}$, is smaller than that without damping, $\omega = \sqrt{\frac{k}{m}}$, and therefore the period is slightly larger.

If $b = 2\sqrt{km}$, then $x = x_0 e^{-\left(\frac{b}{2m}t\right)}$. The system is in critical damping, where it no longer oscillates but returns to its equilibrium position without overshooting the equilibrium position.

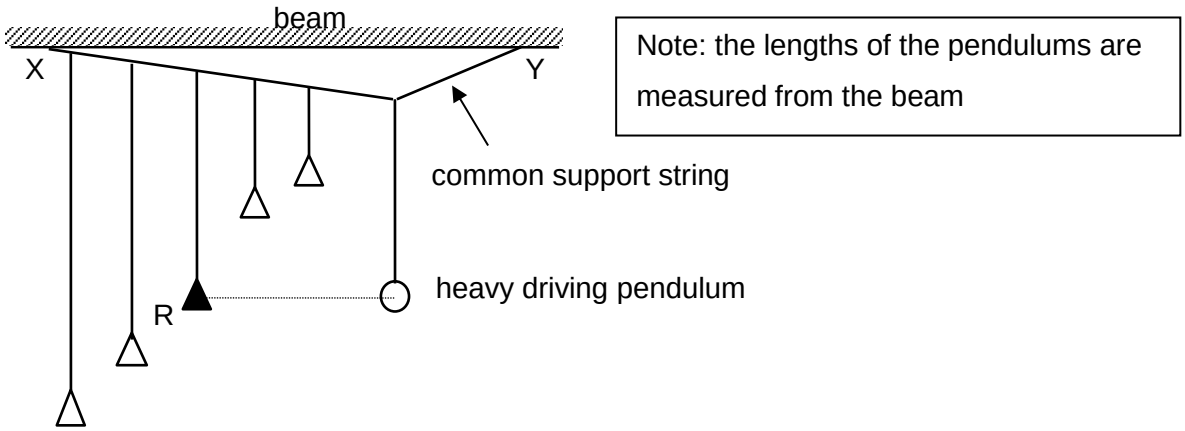
If $b > 2\sqrt{km}$, then the system is heavily damped. Again, there is no oscillation, but the return to its equilibrium takes longer than in the case of critical damping. Its solutions are in the form of $x = C_1 e^{-a_1 t} + C_2 e^{-a_2 t}$, where C_1 and C_2 are constants determined by initial conditions and a_1 and a_2 are constants determined by *m, k* and *b*.

Appendix 4 Demonstration of Forced Oscillation and Resonance - Barton's pendulums

Barton's pendulums consist of several pendulums of different lengths hung from a horizontal string. Each has its own natural frequency of oscillation. The 'driver' pendulum at the end has a large mass and its length is equal to that of pendulum R. When the driver is set swinging, its vibrations are transferred through the common support string to the other pendulums and start the forced oscillations.

Observations show that:

1. All the pendulums vibrate with the same frequency as the driver.
2. Pendulum R whose length matches that of the driver pendulum (therefore has the same natural frequency as the driver) has the largest amplitude. It is said to be resonating with the driver.



Note: the lengths of the pendulums are measured from the beam

Explanation:

All the pendulums are coupled together by the suspension. As the driver swings, it moves the suspension, which in turn moves the other pendulums. The matching pendulum is being pushed slightly once each oscillation, and the amplitude gradually builds up to maximum. The other pendulums are being pushed at a frequency that does not match their natural frequency and they therefore oscillate with smaller amplitudes.

Watch the video clip and further observations on Barton's pendulum

https://bit.ly/3dnr4CQ

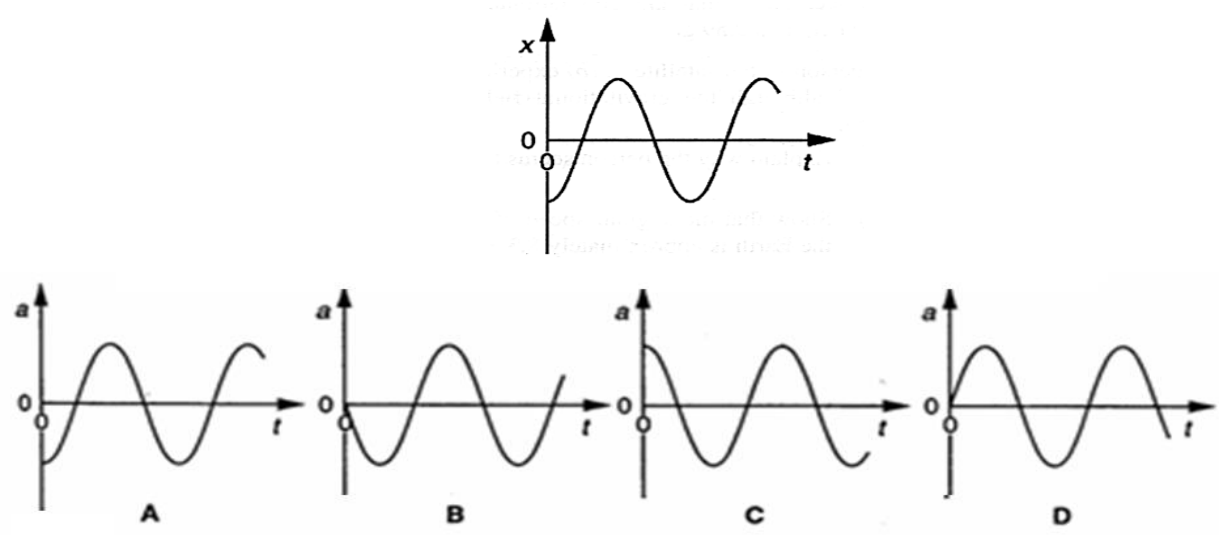
https://bit.ly/3dnk0qb

Tutorial 8 Oscillations

Self-Review Questions

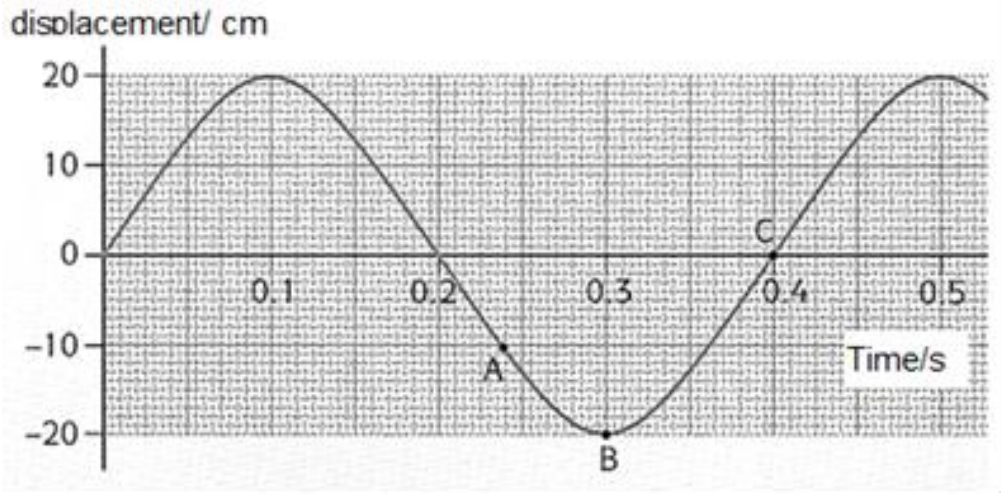
Simple Harmonic Motion

S1 **N03/I/14**: A body moves with simple harmonic motion about point P.
The graph shows the variation with time *t* of its displacement *x* from P.
Which graph shows the variation with time of its acceleration?



S2 **J90/I/10:** A body performing simple harmonic motion has a displacement x given by the equation $x = 30\sin 50t$, where *t* is in seconds. What is the frequency of oscillation?

S3 The figure below shows the displacement-time graph of a simple harmonic oscillator.



(i) Deduce the following quantities:

(a) amplitude,
(b) period,
(c) frequency,
(d) angular frequency,
(e) displacement at A.

(ii) Using your answers to S3(i)(a) and (d), write down an equation that describes the variation of displacement, *x*, with time, *t*, as shown in the graph.

(iii) Substitute the displacement at A in the figure, into the equation to find the corresponding time. Check that it agrees with the value from the graph.

(iv) At which point (A, B, C) is
(a) the magnitude of the velocity maximum?
(b) the magnitude of the acceleration maximum?
(c) State whether the direction of the velocity in (a) and that of the acceleration in (b) is positive or negative. Explain your answers.

(v) Deduce the following quantities:
(a) velocity at B, (b) velocity at C, (c) acceleration at B, (d) acceleration at C.

(vi) Sketch for this oscillator,
(a) the acceleration-time graph
(b) the velocity-time graph
(c) the velocity-displacement graph
(d) the acceleration-displacement graph.

S4 A 0.50 kg body performs simple harmonic motion with a frequency of 2.0 Hz and amplitude of 8.0 mm. Determine the maximum velocity and the maximum acceleration and the corresponding positions of the body.

Energy in simple harmonic motion

S5 Refer to the displacement-time graph in S3. The mass of the oscillator is 1.0 kg.

(i) At which point (A,B,C) is
(a) the potential energy maximum?
(b) the kinetic energy maximum?

(ii) Find the kinetic energy and the potential energy of the oscillator at C and B.

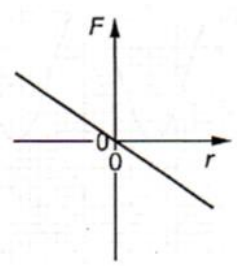
(iii) Sketch for the oscillator, labelled graphs on the same horizontal time axis,
(a) the variation of its kinetic energy,
(b) the variation of its elastic potential energy,
(c) the variation of its total mechanical energy.

(iv) Sketch for the oscillator, labelled graphs on the same horizontal displacement axis,
(a) the variation of its kinetic energy,
(b) the variation of its potential energy,
(c) the variation of its total mechanical energy.

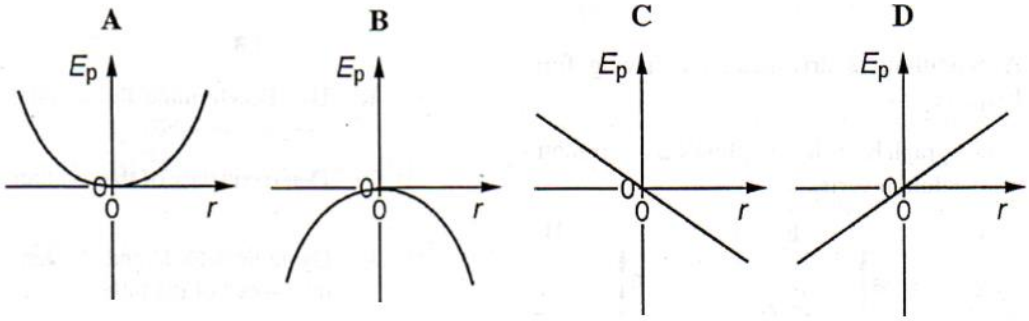
S6 A body oscillates vertically on the end of a light vertical spring with simple harmonic motion of frequency *f* and amplitude *A*. The total energy of the oscillations is proportional to

A $\sqrt{f}$ **B** $\sqrt{A}$ **C** A^2 **D** f

S7 **N04/II/16**: A particle is moving such that the force *F* on it changes with the distance *r* from a fixed point as shown.

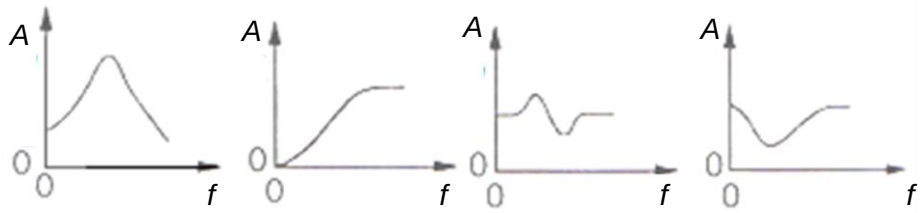


Which graph best shows the relationship between the potential energy *Ep* of the particle and the distance *r*?



Damped Oscillations, Forced Oscillations and Resonance

S8 **N00/I/9:** A pendulum is driven by a sinusoidal driving force of frequency *f*. Which graph best shows how the amplitude *a* of the motion of the pendulum varies with *f* ?



A **B** **C** **D**

Discussion Questions
Simple Harmonic Motion

D1 **J78/II/10**

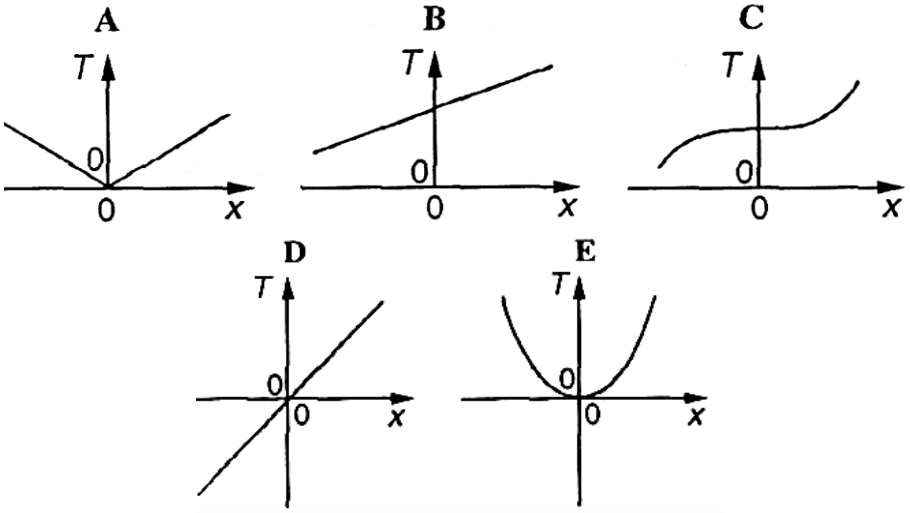
In order to check the speed of a camera shutter, the camera was used to photograph the bob of a simple pendulum moving in front of a horizontal scale.

The extreme positions of the bob were at the 600 mm and 700 mm marks. The photograph showed that while the shutter was open, the bob moved from the 650 mm mark to 675 mm mark.

If the period of the pendulum was 2 s, how long would the shutter remained open?

D2 **N88/I/7**

A mass is hung from the free end of a light helical spring and then given a small displacement vertically downwards. Which graph best represents how *T*, the tension in the spring, varies with *x*, the displacement of the mass from the equilibrium position during subsequent oscillations?

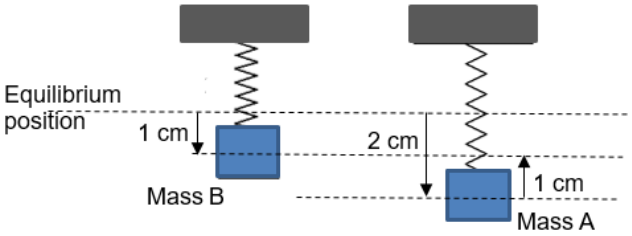


D3 **HClPrelim/15/I/14**

Two identical mass-spring systems are set up side by side. Initially, mass B is displaced downwards 1 cm from its equilibrium position and held there.

Mass A is then displaced downwards 2 cm from its equilibrium position and released.

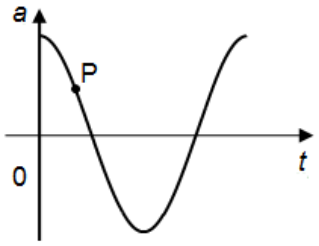
Mass B is released at the instant mass A has moved up a distance of 1 cm.

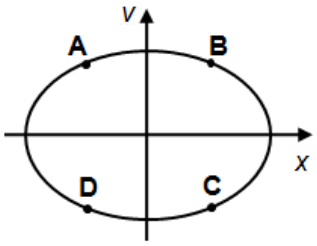


What is the phase difference between mass A and mass B?

A $\frac{\pi}{2}$ rad **B** $\frac{\pi}{3}$ rad **C** $\frac{\pi}{4}$ rad **D** $\frac{\pi}{6}$ rad

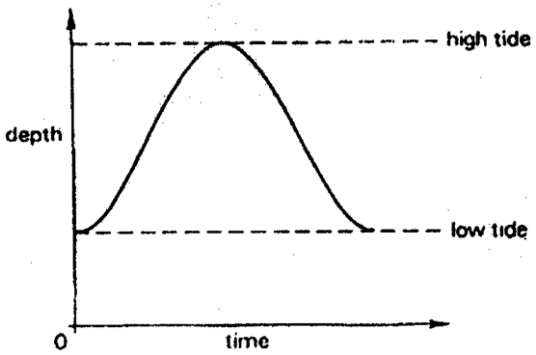
D4 Graph on the left shows how the acceleration of an object undergoing simple harmonic motion varies with time. A graph of velocity against displacement for the same oscillation is shown on the right. Which point in the v-x graph corresponds to point P in the a-t graph?





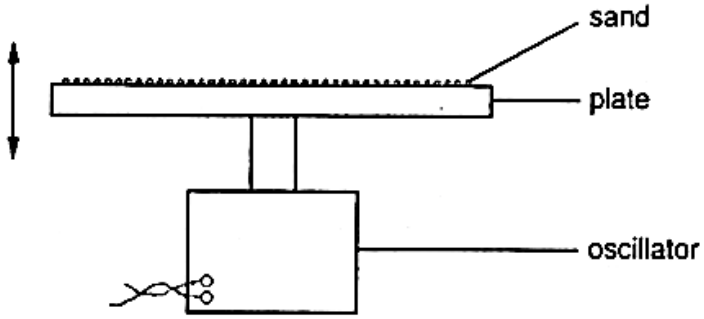
D5 **J91/I/9**

The rise and fall of water in a harbour is simple harmonic. The depth varies between 1.0 m at low tide and 3.0 m at high tide. The time between successive low tides is 12 hours. A boat, which requires a minimum depth of water of 1.5 m, approaches the harbour at low tide. How long will the boat have to wait before entering?



D6 **J99/III/4 (part)**

(b) Some sand is placed on a flat horizontal plate and the plate is made to oscillate with simple harmonic motion in a vertical direction, as shown.



The plate oscillates with a frequency of 13 Hz.

(i) Sketch a graph to show the variation with displacement *x* of the acceleration *a* of the plate.

(ii) The acceleration a is given by the expression

$$a = -\omega^2 x$$

where ω is the angular frequency.

Calculate

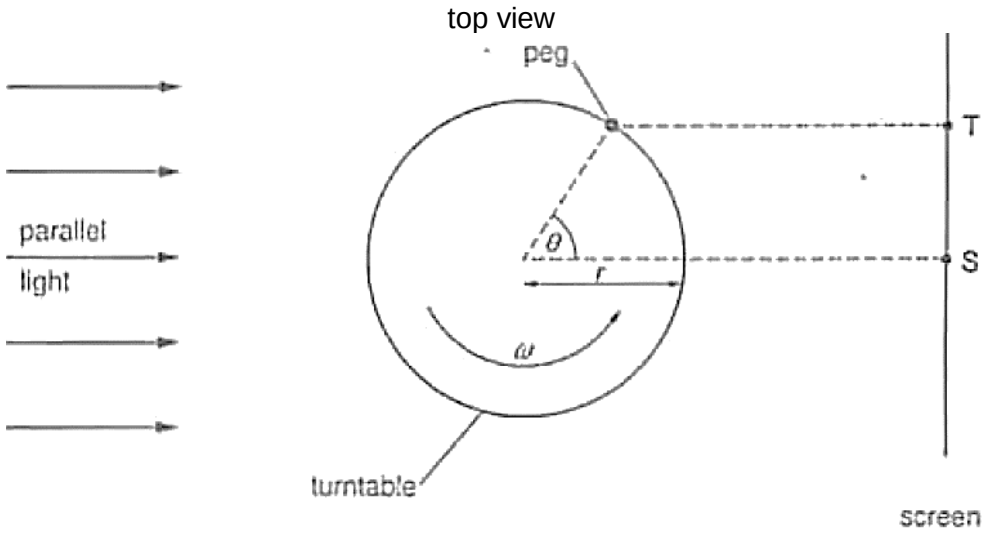
1. the angular frequency

2. the amplitude of oscillation of the plate such that the maximum acceleration is equal to the magnitude of the acceleration of free fall.

(c) Suggest, with a reason, what happens to the sand on the plate in (b) when the amplitude of the oscillations of the plate exceeds the value calculated in (b)(ii)2.

D7 N96/II/2

A vertical peg is fixed to the rim of a horizontal turntable of radius *r*, rotating with a constant angular speed ω, as shown below.



Parallel light is incident on the turntable so that the shadow of the peg is observed on a screen which is normal to the incident light.

At time $t = 0$, $\theta = 0$ and the shadow of the peg is seen at S.
At some later time *t*, the shadow is seen at T.

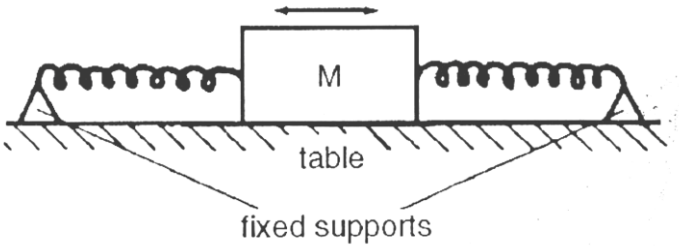
(a) (i) Write down an expression for θ in terms of ω and *t*.
(ii) Derive an expression for the distance ST in terms of *r*, ω and *t*.

(b) By reference to your answer in **(a)(ii)**, describe the motion executed by the shadow on the screen.

(c) The turntable has a radius *r* of 20 cm and an angular speed ω of 3.5 rad s^{-1}. Calculate, for the motion of the shadow on the screen,
(i) the amplitude,
(ii) the period,
(iii) the speed of the shadow as it passes through S,
(iv) the magnitude of the acceleration of the shadow when the shadow is instantaneously at rest.

Energy in Simple Harmonic Motion

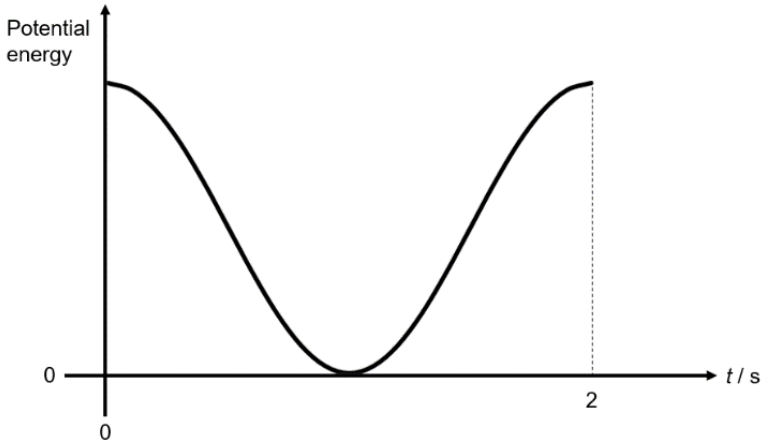
D8 **J81/II/19:** A mass M on a smooth horizontal table is attached by two light springs to two fixed supports as shown below. The mass executes simple harmonic motion of amplitude *a* and period *T*.



Find the energy associated with this simple harmonic motion in terms of *M*, *a* and *T.*

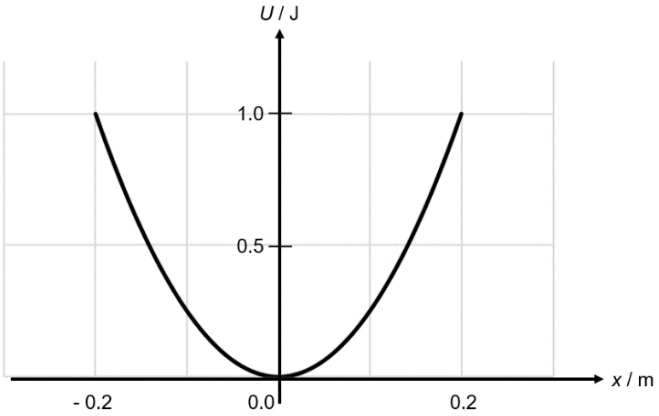
D9 **N21/1/15(modified):** A particle oscillates with simple harmonic motion. The graph shows the variation, with time *t*, of the potential energy of the particle from *t* = 0 to *t* = 2 s.

Sketch the variation of the velocity *v*, with time *t*, of the particle from *t* = 0 to *t* = 2 s.



D10 **N82/II/9:** A particle of mass 4.0 kg moves with simple harmonic motion and its potential energy *U* varies with position *x* as shown below.

What is the period of oscillation of the mass?



D11 One end of a spring is fixed to a support. A mass is attached to the other end of the spring. The arrangement is shown in Fig. 3.1.

This arrangement is used to determine the length l of the spring when mass *M* is attached to the spring. The procedure is repeated for different values of *M.* The variation of mass *M* with length l is shown in Fig. 3.2.

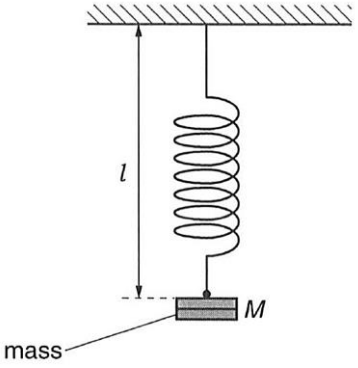


Fig. 3.1

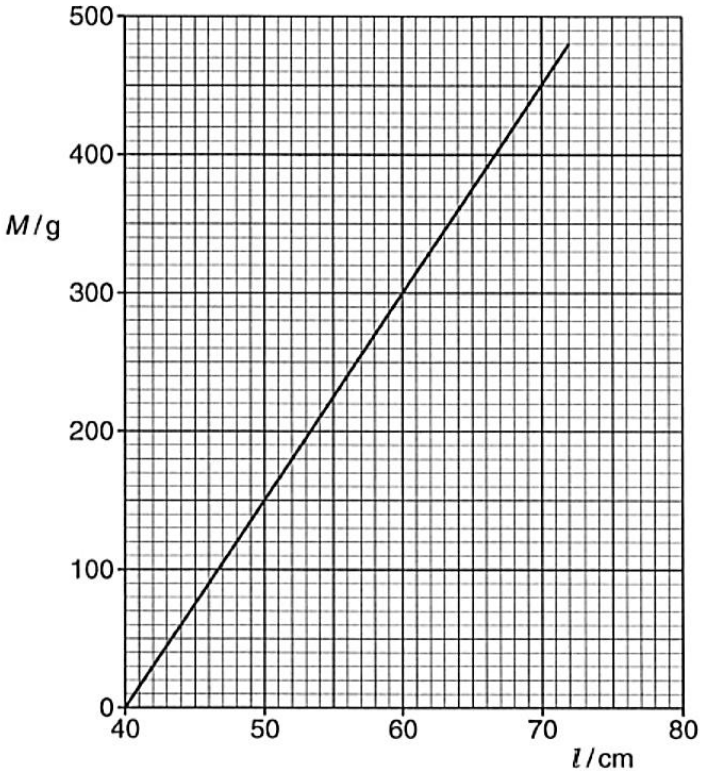


Fig. 3.2

(a) Show that the spring constant *k* of the spring is 14.7 N m^{-1}.

(b) A mass of 450 g is attached to the spring and is held at rest with length l of 50.0 cm. The mass is then released and the mass oscillates freely. The angular frequency of the spring-mass system is given by the formula $\omega = \sqrt{\frac{k}{m}}$.

Calculate the frequency of the system.

(c) Calculate the speed of the mass during its oscillation when the spring is extended to a length l of 80.0 cm.

(d) The spring is assumed to be light. In practice, the spring will have some mass. Assuming that the spring constant *k* is unchanged, suggest and explain the effect on the frequency of oscillation of having a spring with mass.

D12 N08/3/6

(a) Distinguish between *frequency* and *angular frequency* for a body undergoing simple harmonic motion.

(b) A spring that has an unstretched length of 0.650 m is attached to a fixed point. A mass of 0.400 kg is attached to the spring and gently lowered until equilibrium is reached. The spring has then stretched elastically by a distance of 0.200 m.

Calculate, for the stretching of the spring,
(i) the loss in gravitational potential energy of the mass,
(ii) the elastic potential energy gained by the spring.

(c) Explain why the two answers to **(b)** are different.

(d) The load on the spring is now set into simple harmonic motion of amplitude 0.200 m.

Calculate
(i) the resultant force on the load at the lowest point of its movement,
(ii) the angular frequency of the oscillation,
(iii) the maximum speed of the mass.

(e) Fig. 6.1 is a table of the energies of the simple harmonic motion. Complete the table.

	gravitational potential energy/J	elastic potential energy/J	kinetic energy/J	total energy/J
lowest point	0			
equilibrium position				
highest point				

Fig. 6.1

(f) On the axes of Fig. 6.2 below, sketch four graphs to show the shape of the variation with position of the four energies. Label each graph.

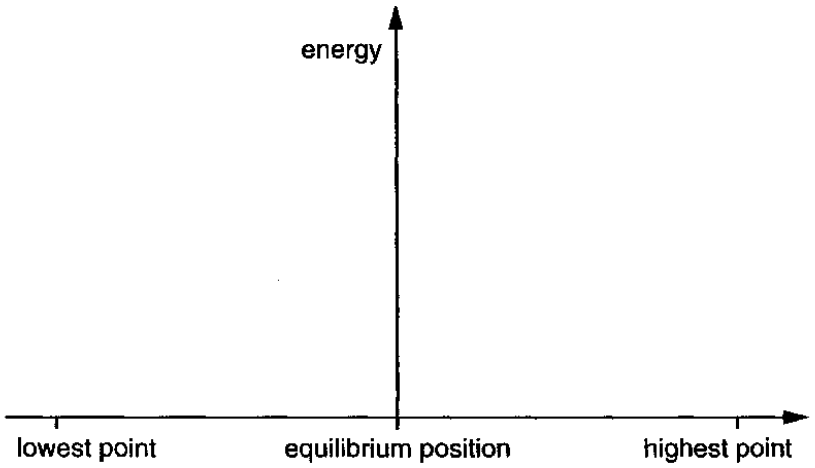


Fig. 6.2

Damped oscillations, forced oscillations and resonance

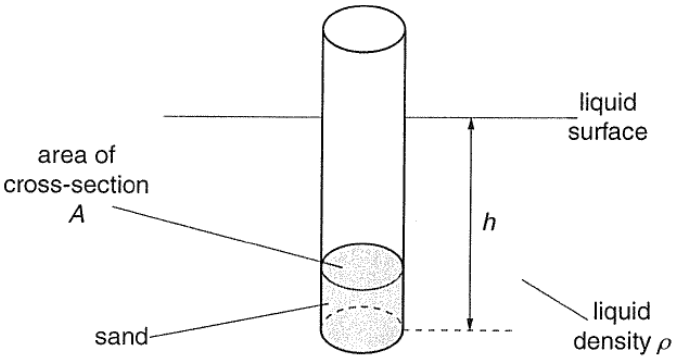
D13 N09/III/1c

In normal use, a loudspeaker produces a range of frequencies of sound. Suggest why it is important that the natural frequency of vibration of the cone of the loudspeaker is not within this range of frequencies.

D14 N13/III/7

(a) State the origin of the upthrust acting on a body in a fluid.

(b) A tube, sealed at one end, has a uniform area of cross-section *A*. Some sand is placed in the tube so that it floats upright in a liquid of density ρ, as shown in the figure.



The total mass of the tube and the sand is *m*. The tube floats with its base a distance *h* below the surface of the liquid. Derive an expression relating *m* to *h, A* and ρ. Explain your working.

(c) The tube in **(b)** is displaced vertically and then released. For a displacement *x*, the acceleration *a* of the tube is given by the expression

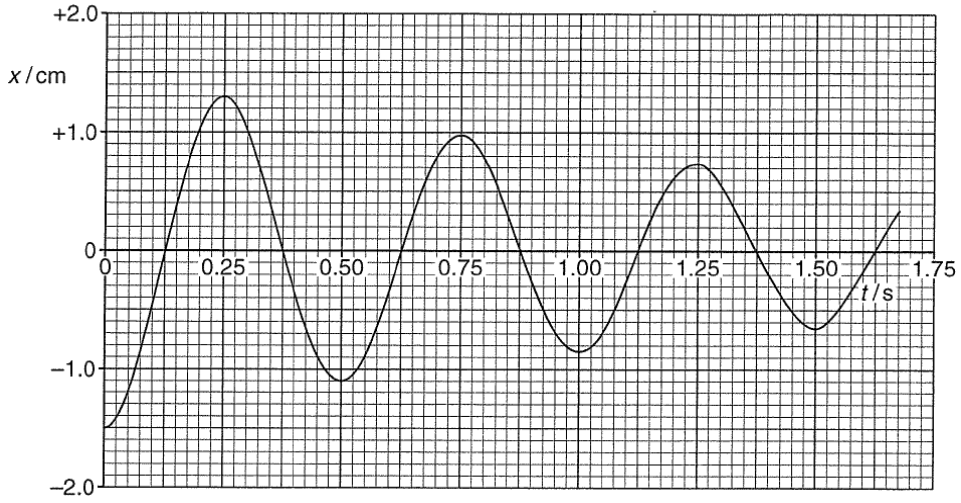
$$a = -\left(\frac{\rho A g}{m}\right) x$$

where *g* is the acceleration of free fall.

(i) Explain why the expression leads to the conclusion that the tube is performing simple harmonic motion.

(ii) The tube has a total mass *m* of 32 g and the area *A* of its cross-section is 4.2 cm^2. It is floating in liquid of density ρ of 1.0×10^3 kg m^{-3}. Show that the frequency of oscillation of the tube is 1.8 Hz.

(d) The tube in **(b)** is now placed in a different liquid. The tube oscillates vertically. The variation with time *t* of the vertical displacement *x* of the tube is shown below.



(i) Use the figure to

1. determine the frequency of oscillation of the tube,

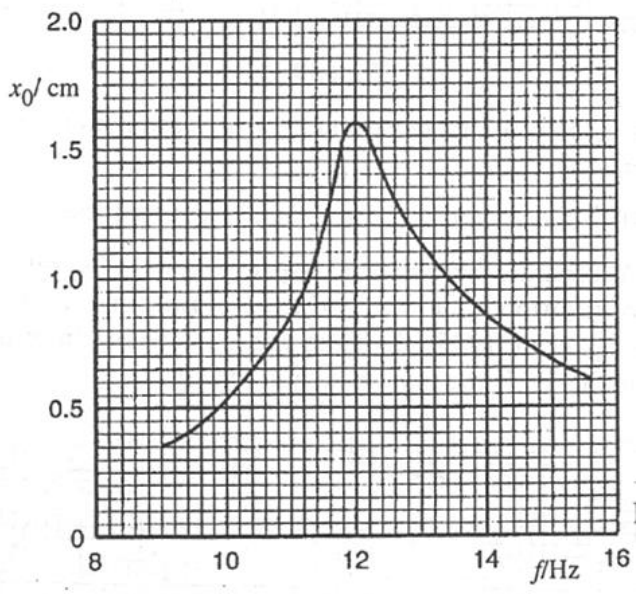
2. calculate the density of the liquid.

(ii) 1. Suggest two reasons why the amplitude of the oscillation decreases with time.

2. Calculate the decrease in energy of the oscillation during the first 1.0 s

D15 N05/II/4 (modified)

The figure shows the variation with frequency f of amplitude x_0 of the forced oscillations of a machine.



(a) What is meant by a forced oscillation?

(b) At any value of frequency, the oscillations of the machine are simple harmonic.

(i) The machine is vibrating at its maximum amplitude,

1. find the maximum magnitude of the linear speed,

2. find the maximum magnitude of the linear acceleration,

3. state the name of this effect.

(ii) Determine the time interval between maximum linear speed and subsequent maximum linear acceleration.

D16 N22/II/1 (modified)

A mass *m* is suspended from a vertical spring attached to a fixed support. The mass is pulled down and held at a vertical displacement of 0.16 m from its equilibrium position as shown in Fig. 1.1.

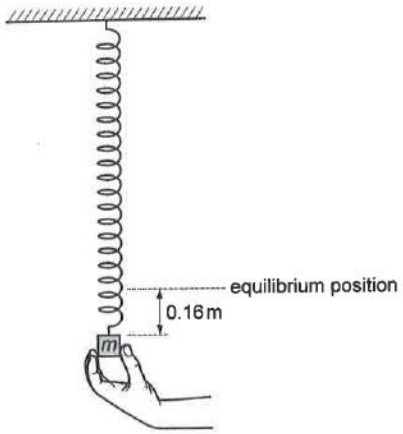


Fig. 1.1

The mass is released.

The variation with vertical displacement *x* of the velocity *v* of the mass on the spring is shown in Fig. 1.2.

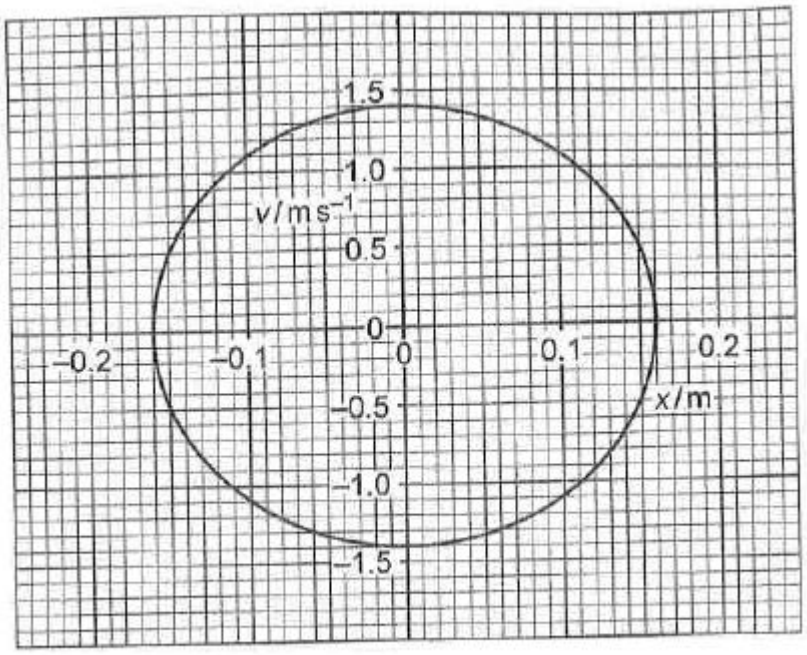


Fig. 1.2

(a) Determine the maximum acceleration a of the mass.

(b) A light piece of card is attached to the bottom of the mass so that the oscillations are lightly damped. The mass is displaced vertically and then released.

On Fig. 1.2, draw the variation of v with x to show the first complete cycle of these damped oscillations.

NUMERICAL ANSWERS

Self-Review Questions

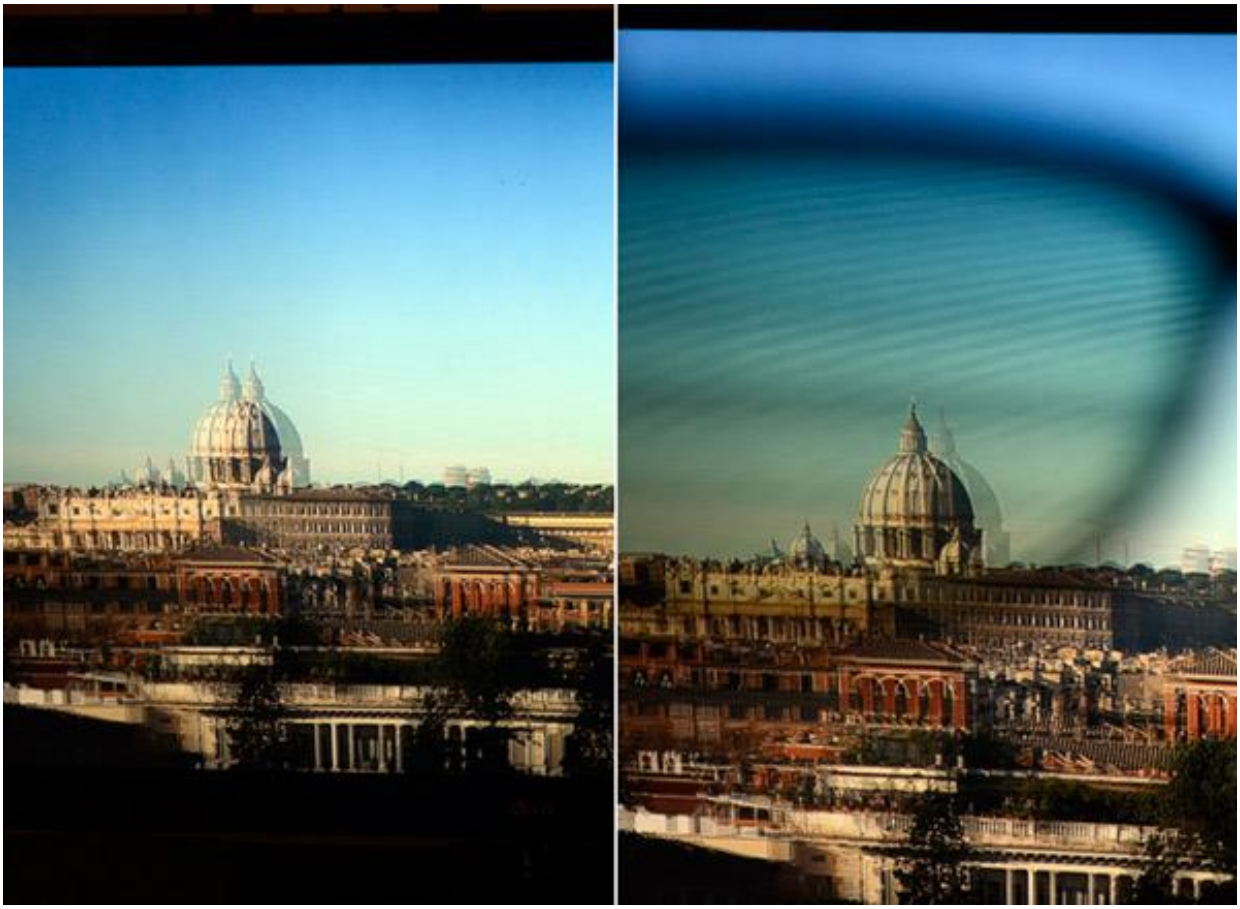
S1 C
S2 8.0 Hz
S3 (i)(a) 20.0 cm; (i)(b) 0.4000 s; (i)(c) 2.500 Hz; (i)(d) 15.71 rad s^{-1}; (i)(e) -10.0 cm
(iii) 0.233 s; (v)(a) 0; (v)(b) 3.14 m s^{-1};
(v)(c) 49.3 m s^{-2}; (v)(d) 0
S4 v_{max} = 0.10 m s^{-1} at equilibrium position; a_{max} = 1.3 m s^{-2} at x = - 8.0 mm
S5 (i)(a) Point B; (i)(b) Point C; (ii) PE_C = 0,
KE_C = 4.93 J, PE_B = 4.93 J, KE_B = 0
S6 C
S7 A
S8 A

Discussion Questions

D1 0.17 s
D5 2.0 h
D6 (b)(ii) 82 rad s^{-1}, 1.47 mm
D7 20 cm, 1.8 s, 70 cm s^{-1}, 2.5 m s^{-2}
D8 $\frac{2m\pi^2a^2}{T^2}$
D10 1.8 s
D11 0.910 Hz; 0.990 m s^{-1};
D12 (b)(i) 0.785 J; (b)(ii) 0.392 J; (d)(i) 3.92 N; (d)(ii) 7.00 rad s^{-1}; (d)(iii) 1.40 m s^{-1}
D14 (d)(i)(1) 2.0 Hz, (2) 1200 kg m^{-3}, (d)(ii)(2) 3.86 x 10^{-4} J
D15 (b)(i)(1) 1.21 m s^{-1}, (2) 91.0 m s^{-2}; (b)(ii) 0.0208 s
D16 (a)12.3 m s^{-2}

Chapter 9

Wave Motion



Left: a 3D image viewed without glasses on the LG set; right: the same image viewed through 3D glasses shows noticeable ghosting and visible horizontal bands. (Photo taken at closer-than-normal viewing distance.)

Source: Consumer Reports.org

Topic 9: Wave Motion

Syllabus 9749

Content

- Progressive waves
- Transverse and longitudinal waves
- Polarisation
- Determination of frequency and wavelength of sound waves

Learning Outcomes

Candidates should be able to:

(a) show an understanding of and use the terms displacement, amplitude, period, frequency, phase difference, wavelength and speed.

(b) deduce, from the definitions of speed, frequency and wavelength, the equation $v = f\lambda$.

(c) recall and use the equation $v = f\lambda$.

(d) show an understanding that energy is transferred due to a progressive wave.

(e) recall and use the relationship, intensity $\propto$ *(amplitude)*^{2}.

(f) show an understanding of and apply the concept that a wave from a point source and travelling without loss of energy obeys an inverse square law to solve problems.

(g) analyse and interpret graphical representations of transverse and longitudinal waves.

(h) show an understanding that polarisation is a phenomenon associated with transverse waves.

(i) recall and use Malus' law (intensity $\propto \cos^2\theta$) to calculate the amplitude and intensity of a plane polarised electromagnetic wave after transmission through a polarising filter.

(j) determine the frequency of sound using a calibrated oscilloscope.

(k) *determine the wavelength of sound using stationary waves.

* To be covered in Topic 10 Superposition.

Table of Contents

Syllabus 9749 I

9.1 Introduction 1

9.2 Classification of Waves 1
- Transverse and Longitudinal Waves 4
- Progressive Waves and Stationary Waves 1

9.3 Basic Terminology 6
- Sinusoidal Waves 6

9.4 Graphical Representation of Waves 12
- Graph of Displacement (*y*) vs. Position (*x*) 12
- Graph of Displacement (*y*) vs. Time (*t*) 14

9.5 Phase and Phase Difference 16

9.6 Energy Transmitted by a Wave 19
- Intensity of a Wave 19
- Intensity of a Wave from Point Source or Spherical Source 20

9.7 Polarisation of Waves 21
  i. Light Passing Through Two Polarisers 23

Appendices 26
- Appendix A: Function Generator and Cathode Ray Oscilloscope (C.R.O) 26
  i. Function Generator 26
  ii. Cathode Ray Oscilloscope 27
- Appendix B: 3-D films and Polarization 28
- Appendix C: Polarisation by Absorption 29

Tutorial 9 Wave Motion 30
- Self-Review Questions 30
- Discussion Questions 34

9.1 Introduction

Consider a rope that is held under tension. If we give the left of the rope a small wiggle, a pulse is formed (called the waveform) which can be seen to travel along the length of the rope (Fig. 9.1.1).

When the waveform travels down the rope, **there is a transfer of energy and momentum** from one end of the rope to the other. However each section of the rope only vibrates up and down about an equilibrium position--- **no section of the rope actually travels from one end to the other**.

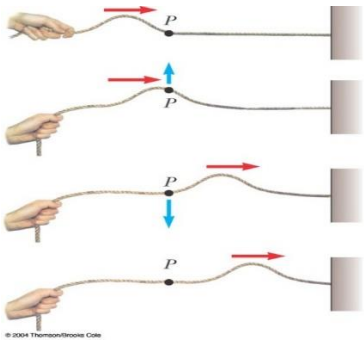


Fig. 9.1.1 Wave pulse travelling down the rope

9.2 Classification of Waves

Waves can be categorised by three broad properties (Fig. 9.2.1):

(i) the type of the wave,

(ii) the mode of vibration of the wave, and

(iii) the motion of the wave.

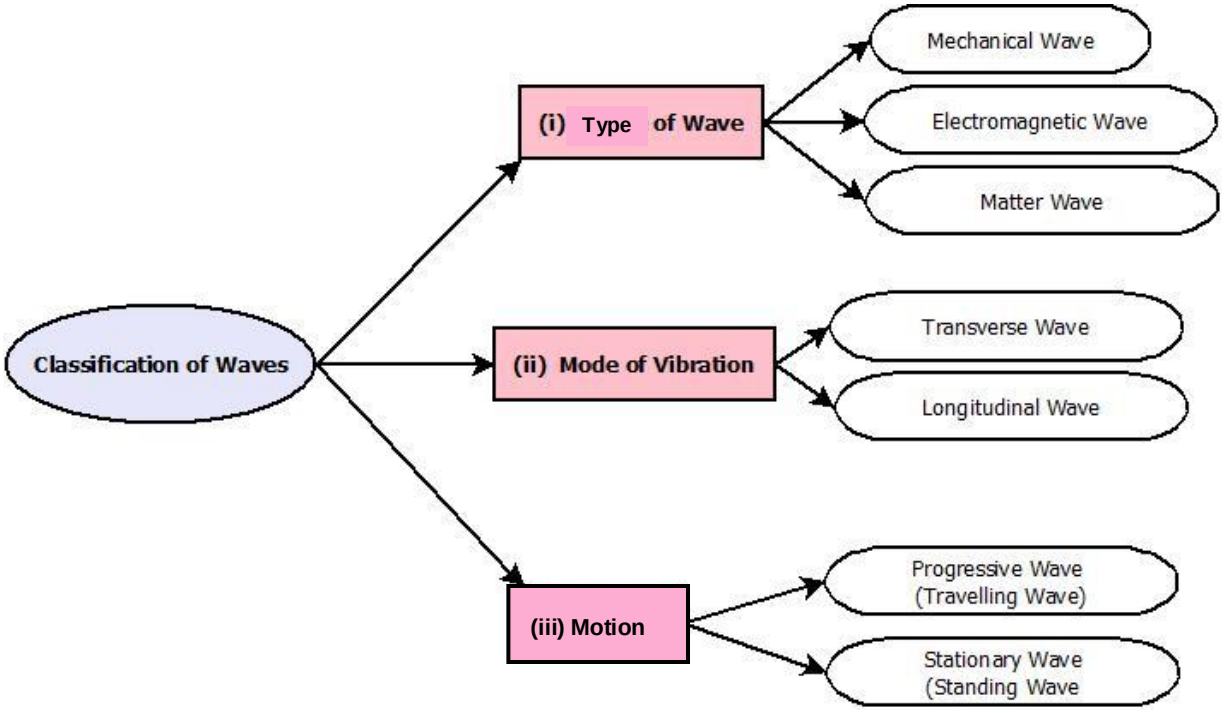


Fig. 9.2.1 Classifications of Waves

For example, a sound wave can be characterised as mechanical in nature, longitudinal in terms of its displacement and can be either progressive or stationary.

Mechanical Waves

Mechanical waves are created when part of a physical medium is disturbed. The disturbance spreads through the medium distributing energy to points far from the original disturbance. All mechanical waves therefore require a medium for propagation.[1] The speed of wave propagation depends on the medium (specifically both its inertial and elastic properties). Sound waves, waves in ropes, water waves and seismic waves are all examples of mechanical waves.

Electromagnetic (EM) waves

Electromagnetic (EM) waves consist of *mutually perpendicular time-varying electric field ($\vec{\mathbf{E}}$) and magnetic field ($\vec{\mathbf{B}}$) oscillating perpendicularly to the direction of wave travel* (Fig. 9.2.2).

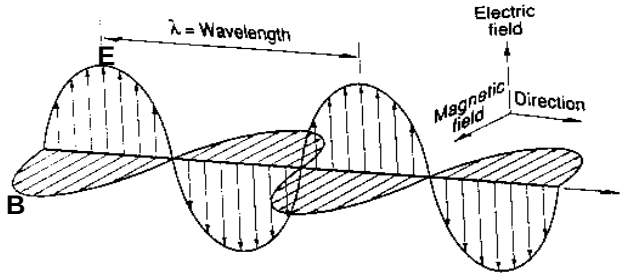


Fig. 9.2.2 The E-field and B-field are mutually perpendicular to each other and both fields are also perpendicular to the direction of wave travel.

Characteristics of EM Waves:

- All EM waves are transverse in nature, i.e. the planes of oscillation of the electric and magnetic fields are perpendicular to the direction of motion.
- All EM waves are self-propagating and hence do not require a medium in which to travel. Unlike mechanical waves they can move through a vacuum but are also able to propagate through materials e.g. light through glass, X-rays through bone and radio waves through walls.
- In vacuum, all EM waves travel at the same speed *c* known as the speed of light:

$$c = 3.00 \times 10^8 \text{ m s}^{-1}$$

The Electromagnetic Spectrum

Common examples of electromagnetic waves include visible light, ultraviolet light, radio wave, microwaves and x-rays. Fig. 9.2.3 shows the electromagnetic spectrum and relative positions of these EM waves on the spectrum.

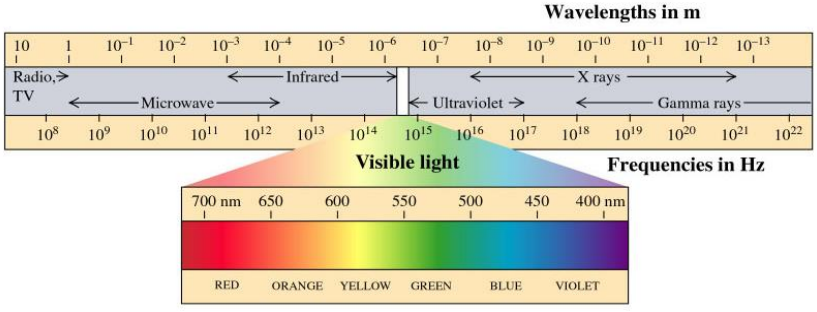


Fig. 9.2.3 The electromagnetic spectrum.

[1] Propagation of mechanical waves is possible due to the restoring forces produced upon deformation; Sound waves, which rely on oscillations of molecules of the medium resulting in pressure variation, obviously cannot be propagated through a vacuum.

The following table shows the names commonly apportioned to the different sections of the EM spectrum and their **corresponding wavelengths and frequencies**.

Type	Wavelength (λ)	Frequency	Typical sources, uses and Effects
Gamma Rays (γ)	~ 10^{-15} m (Size of the nucleus)	~ 10^{20} Hz	From radioactive materials. Used to destroy cancer cells.
X-Rays	~ 10^{-10} m (Size of an atom)	~ 10^{18} Hz	From electrons stopped rapidly in X-ray tube. Causes ionization and fluorescence. Used for X-ray photography.
Ultraviolet (u-v) light	~ 10^{-8} m	~ 10^{16} Hz	From very hot objects, ionizes atoms. Causes fluorescence in some chemicals. Kills germs and causes suntan.
Visible light	~ 400 nm to 700 nm	~ 10^{14} Hz	Only form of radiation visible to the human eyes.
Infrared light	700 nm to ~1 mm	~ 10^{12} Hz	From hot objects. Used for heating. Also used for very short range transmission of information (between laptops, printers, mobile phones.)
Microwaves	~10^{-2} m (Grids of microwave oven)	~ 10^{10} Hz	From electrons oscillating in magnetron. Used for communication (e.g. by cellular mobile phones), radar and cooking.
Radio waves	~ 1 m	~ 10^{8} Hz	From electrons oscillating in aerial. Used for global communication.

- A useful mnemonic to remember for visible light :

Red :	**R**ichard	(700 nm)
Orange :	**O**f	
Yellow:	**Y**ork	
Green:	**G**ave	
Blue:	**B**attle	
Indigo:	**I**n	
Violet:	**V**ain	(400 nm)

The electromagnetic spectrum is a *continuous spectrum* i.e. there are no gaps or *sharp boundaries* between one type of radiation and next. For example, the wavelengths of γ-rays and x-rays overlap. Their key difference is in their mechanism of production.

Transverse and Longitudinal Waves

A **transverse wave** is a wave in which the points of disturbance oscillate about their equilibrium positions ***perpendicular*** **to the direction of wave travel or energy propagation**.

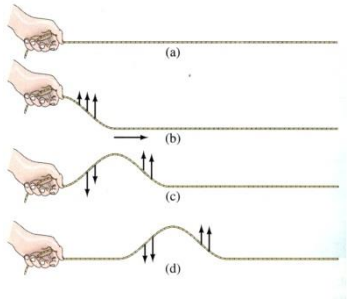


Fig. 9.2.4 Transverse Wave Pulse

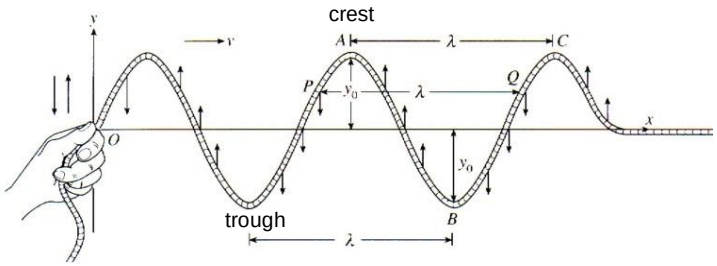


Fig. 9.2.5 Transverse Wave Train

Examples of transverse waves are waves in a rope and electromagnetic waves. Fig. 9.2.4 and Fig. 9.2.5 show a transverse wave pulse and a continuous transverse wave train on a rope respectively.

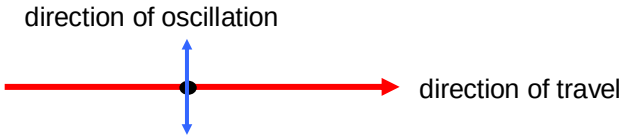


Fig. 9.2.6 Direction of oscillation of a point compared with direction of travel for a transverse wave.

HTML5 simulation: Transverse Traveling Wave

https://ngsir.netfirms.com/j/Eng/Twave/Twave_js.html

A **longitudinal wave** is a wave in which the points of disturbance oscillate about their equilibrium position along **the direction of wave travel or energy propagation**.

- Fig. 9.2.7 shows a longitudinal wave set up on a slinky. Sound wave is an example of a longitudinal wave.
- In a longitudinal wave, the regions in which the points or particles are closest together are known as regions of "**compression**". The regions in which the point or particles are farthest apart are known as regions of "**rarefaction**". Add in diagrams to show compression and rarefaction.

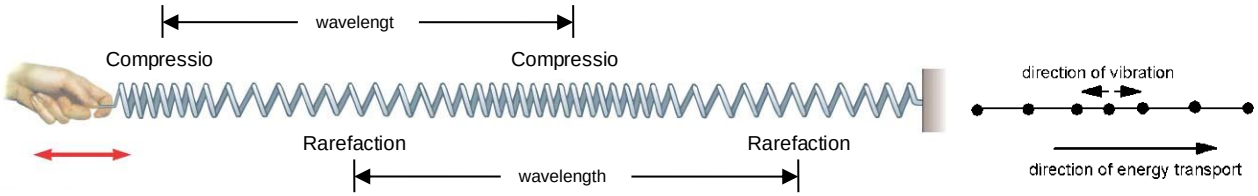


Fig. 9.2.7 The direction of oscillation of a point on longitudinal is parallel to the direction of the wave travel and energy transport

HTML5 simulation: Longitudinal Traveling Wave https://ngsir.netfirms.com/j/Eng/Lwave/Lwave_js.htm

Progressive Waves and Stationary Waves

- Progressive Waves are also known as ***travelling waves***.
  - Progressive waves are waves in which **the wave profile moves away from the source** and **causes energy to be transferred away from the source** to other regions.
  - Fig. 9.2.8 and Fig. 9.2.9 below illustrate a progressive transverse wave pulse and a progressive longitudinal wave pulse respectively.

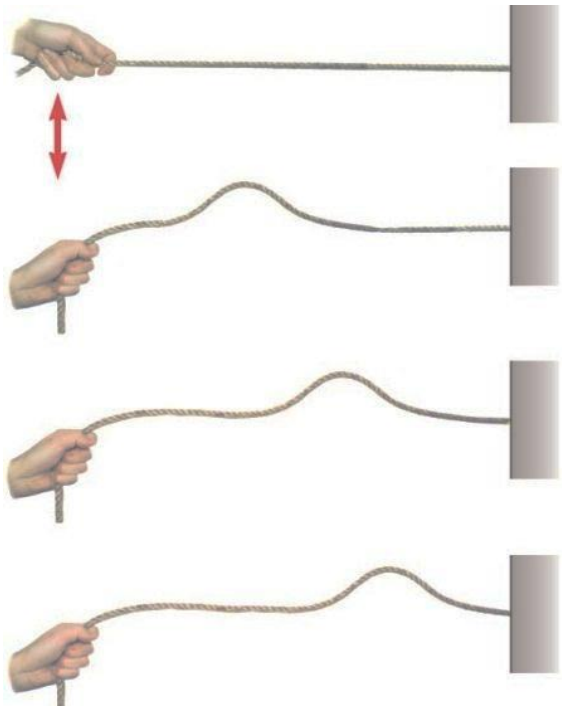


Fig. 9.2.8 Progressive Transverse Wave Pulse

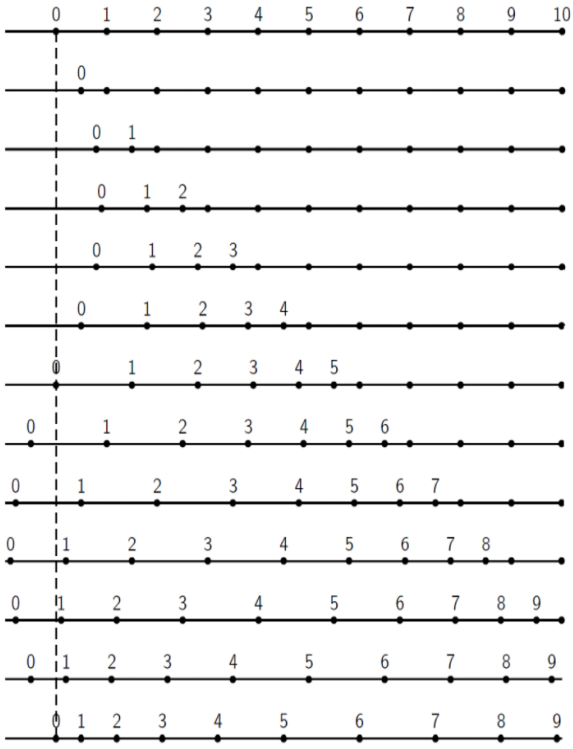


Fig. 9.2.9 Progressive longitudinal wave pulse

- **Stationary waves** are also known as *standing waves*. [2]
  - For a stationary wave, there seems to be no progression of the wave profile in either direction (Fig. 9.2.10) and there are points on the wave which are permanently at rest (known as nodes, N).
  - As there is no progression of the waveform, there is no transfer of energy. Energy is said to be trapped in the waves and does not propagate. The energy is said to be *localised*.

HTML5 simulation: Types of Waves

Transverse vs Longitudinal, Traveling vs Standing

https://ngsir.netfirms.com/j/Eng/waves/waves_js.htm

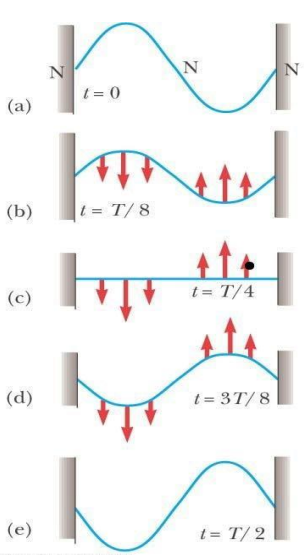
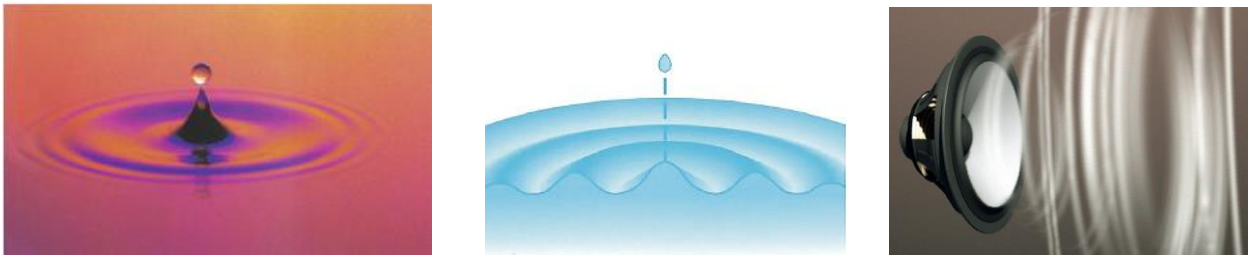


Fig. 9.2.10 A standing wave on a string

[2] We will study stationary waves in greater depth in the next chapter – Superposition.

9.3 Basic Terminology

The waves that we encounter may be one-dimensional, two-dimensional or three-dimensional. Waves on string and slinky are examples of one-dimensional waves, i.e. the waves are travelling only in one dimension. A water wave that is produced by a raindrop that is striking the surface of a lake is an example of a two-dimensional wave. (Fig. 9.3.1). A sound wave in air is an example of a three-dimensional longitudinal wave. It consists of compressions and rarefactions of air molecules spreading out in all directions



(a) (b) (c)

Fig. 9.3.1 (a) and (b) Two-dimensional water wave (c) Three-dimensional sound waves propagating from a speaker

Sinusoidal Waves

If we sound a tuning fork which produces a "pure tone" i.e. a single frequency and use a microphone and oscilloscope (See Appendix A for review of CRO) to capture the wave form we will observe a near perfect sine profile. So, for a medium that undergo a periodic disturbance of a single frequency, we characterise the waves that emanate from the disturbance and pass through the medium as sinusoidal.

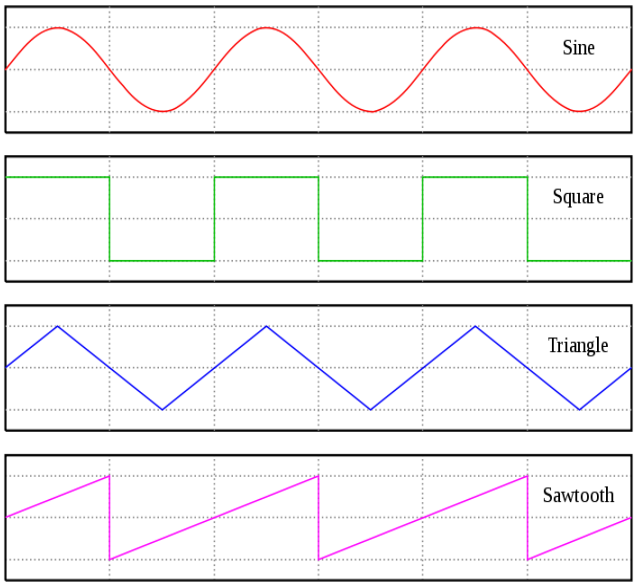


Fig. 9.3.2. Different types of Periodic Waves. All periodic waves can be formed from a combination of sinusoidal waves.

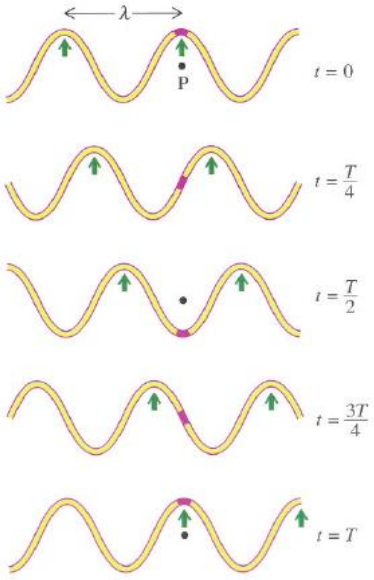


Fig. 9.3.3. As the wave progresses from left to right (see arrow). A point (P) on a wave oscillates simple harmonically.

We will use a transverse sinusoidal wave set up on the string to illustrate some basic terminology used in waves.

Suppose we set the free end of a string in simple harmonic motion. A transverse wave will be set up along the string and propagate down the string.

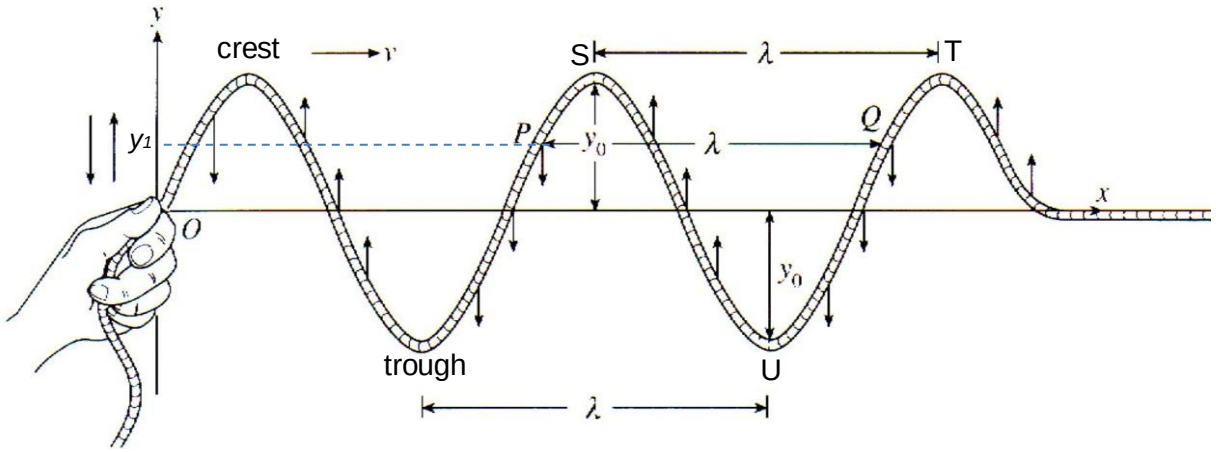


Fig. 9.3.4 The figure shows a snapshot of the wave on the rope at an instant of time.

Displacement (y) :

The displacement of a particular point on the wave refers to the **distance and the direction** of the point from its equilibrium position.

Before the wave is setup, the rope lies along the x-axis. When the wave is setup in the rope, some points on the rope are being displaced from their equilibrium positions.

- Points P and Q are displaced at a distance of y_1 *from* their equilibrium position in the positive direction (as defined by the *y-axis*). Hence, their displacements are both $y = +y_1$.
- Point U is displaced at a distance of y_o from its equilibrium position in the negative direction. Hence its displacement is $y = -y_o$.

Amplitude (A or y_o) :

The amplitude of a wave is the magnitude of the maximum displacement of a point on the wave.

- Fig. 9.3.4: Points S, T and U are at the amplitudes of the wave.

Wavelength (λ) :

The wavelength of a wave is the distance between two adjacent points which are oscillating in phase.[3]

Note:

- For a transverse wave, the distance between two successive crests or two successive troughs corresponds to a single wavelength.
- For a longitudinal wave, the distance between two successive compressions or rarefactions also corresponds to a single wavelength.

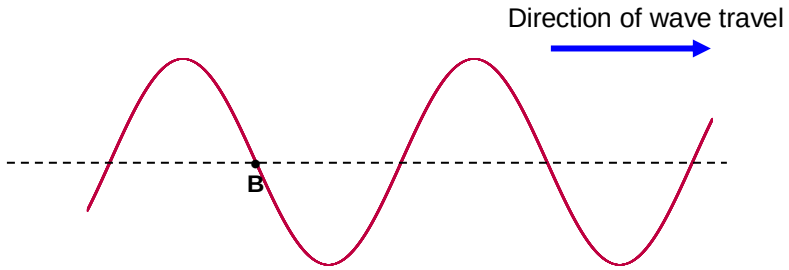
[3] This refers to two points in the same state of oscillation. The concept of phase associated with oscillations will be dealt with further in the chapter.

Example 9.3.1:

Consider the snapshot of the wave shown below. Other than the distance between the 2 successive crests or troughs,

(i) identify another point (B') that pairs with B that will correspond to a single wavelength.

(ii) Choose another pair of points on the wave that correspond to one wavelength.



(iii) What do you notice about the pairs of points that represent one wavelength?

Each pair of points not only have the _____ but also the _____. They are said to be in the same state of oscillation or _____.

Period (*T*) :

Period is the time taken for a point on the wave to *complete one oscillation cycle*.

- Period is also the time taken for the wave to advance by a distance of one wavelength or the time taken for *2 consecutive crests/ compressions or troughs (rarefactions) to pass through a certain point.* (see Fig. 9.3.3)

Frequency (*f*):

Frequency is the *number of oscillations per unit time* made by a point on a wave.

- Frequency is also the number of crests/compressions or troughs/ rarefactions that pass through a point along the path of wave travel in a unit time.
- S.I. unit of frequency is s^{-1} or *Hz (hertz)*.
- The relationship between frequency *f* and period *T* is given by : $f = \frac{1}{T}$

Wavefronts

- A **wavefront** (of a 2-dimensional or 3-dimensional wave) is an imaginary line or curve which joins points on the wave which are oscillating in phase.
  - *All points on a wavefront are at the same part of the cycle* of their vibration.
  - By convention, when we represent waves in a diagram, the *distance between consecutive wavefronts is one wavelength apart*.
- The direction of propagation of the waves perpendicular to the wavefront as indicated by arrows (→) are known as *rays*.

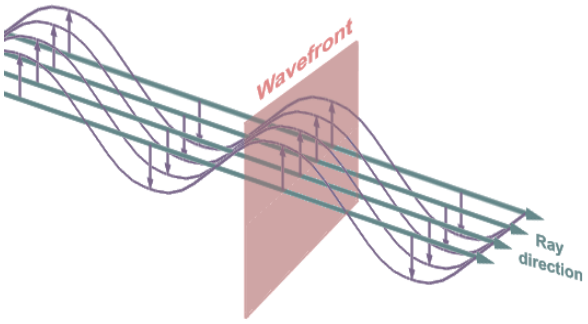
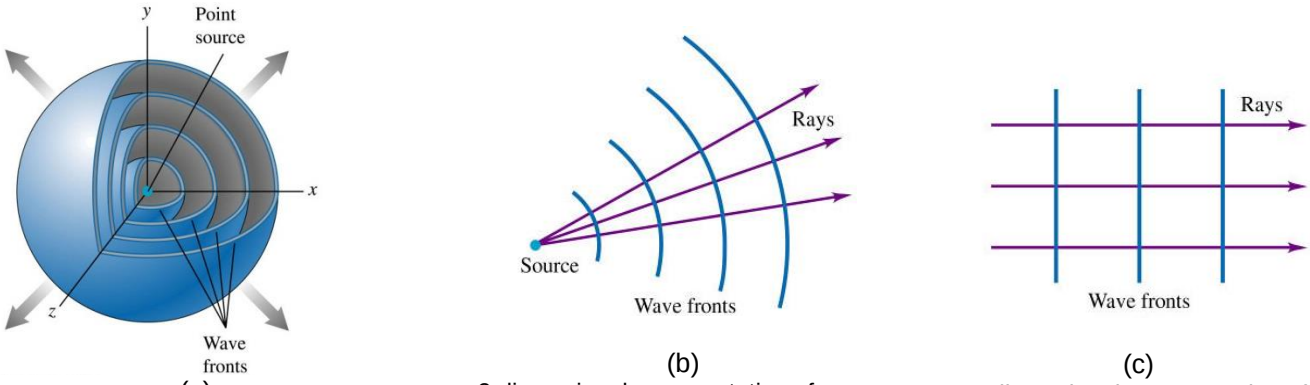


Fig. 9.3.5. Wavefront of a wave.



(a) 3-dimensional waves that spread out uniformly in all directions from a point source.

(b) 2-dimensional representation of waves from a point source.

(c) 2-dimensional representation of waves from a planar source.

Fig. 9.3.6. Wave representations using wave fronts and rays.

Wave Speed (*v*) :

The speed of a wave is the distance that the wave profile travels per unit time.

Note:
There are two sets of motion associated with waves.

- One is that of the motion associated with the oscillation of the individual points on the wave, the points always oscillate about the equilibrium position.
- The other is that of the motion associated with the propagation of waveform of the wave. This is the one that defines the wave speed.

Relationship between Frequency, Speed and Wavelength

The time taken for a wave to move a distance of one wavelength λ is one period T.

From the definition of wave speed: Speed of the wave, $v = \frac{\text{Distance Travelled}}{\text{Time Taken}} = \frac{\lambda}{T}$

Now, $f = \frac{1}{T} \Rightarrow$ $v = f\lambda$ where f is the frequency of the wave.

Example 9.3.2 : Transverse wave

The diagram below shows an instantaneous position of a string as a transverse progressive wave travels along it from left to right.

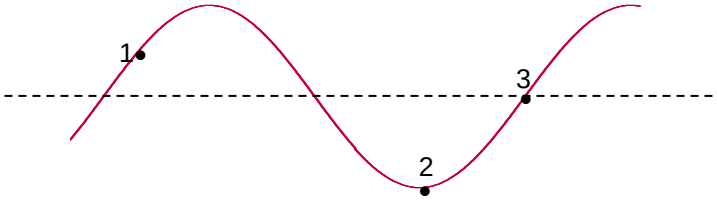


Fig. 9.3.7

(a) Taking upwards to be positive, for each of the points 1, 2, and 3 on the string,
(i) state whether the displacement of each point is positive or negative.

(ii) What are the directions of the velocities of the points?

(iii) Of the three points, which one has the largest speed?

(b) Mark on the string a point A with zero velocity.

(c) What is the direction of acceleration of point A?

Example 9.3.3: Longitudinal Wave

Fig. 9.3.8 shows a 'snapshot' of a horizontal slinky spring when it is (a) at rest and (b) carrying a longitudinal wave moving from left to right. There is no energy dissipation along the slinky spring.

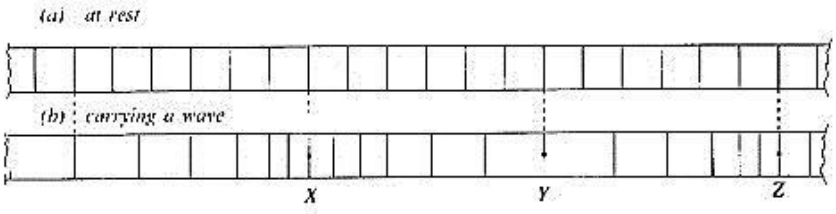


Fig. 9.3.8

State and explain whether each of the following statements is true or false.

Statement A: The distance between X and Y is one wavelength.

False. XY is the distance between the centre of a compression and the centre of a rarefaction. Hence XY corresponds to _____ λ. (_____ corresponds to 1λ)

Statement B: The amplitude of oscillation of Y is greater than the amplitude of oscillation of X.

False. Since no energy is dissipated, the amplitude of oscillation is the same for all points for a 1-D progressive wave. Hence, amplitude of X = amplitude of Y.

Statement C: The displacement of X at this instant is zero.

True. X is at its _____ position. Hence its displacement at this instant is _____ (same for Y, Z).

Example 9.3.4: Speed of a Wave

A fisherman notices that wave crests pass the bow of his anchored boat every 6.0 s. He measures the distance between two adjacent crests to be 15.0 m. How fast are the waves travelling?

Example 9.3.5

Calculate the frequency of green light of wavelength 0.60 μm.

9.4 Graphical Representation of Waves

Consider a transverse water wave travelling from left to right as shown in the diagram below.

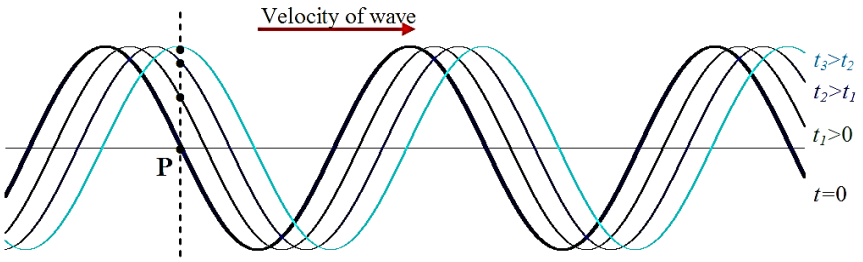


Fig. 9.4.1. A progressive transverse wave moving towards the right.

As the wave moves through this region of space, we note the following:

- At a particular instant of time, the *displacement of each point on the wave varies with the distance it is located from the source*.
- For a particular point on the wave e.g. at **P**, the *displacement of the point also varies with time*.

Therefore, to completely represent a wave, we need two graphs:

1 **Graph of displacement (*y*) varies with position (*x*)** of the points on a wave (at a particular instant of time.)

At a particular instant of time $\Rightarrow$ Graph of y vs. x.

2 **Graph of displacement (*y*) varies with time (*t*) for a particular point on the wave** (at a particular position).

At a particular position $\Rightarrow$ Graph of y vs. t.

Graph of Displacement (*y*) vs. Position (*x*)

Consider a snapshot of a wave on a rope. (Fig. 9.4.2). By superimposing a pair of x and *y* axes on the snapshot, we can describe how the displacement of each point on the rope varies with position.

This graph of displacement vs. position at a particular instant of time (*y vs. x*) gives the **wave profile at an instant of time.** You can imagine this as a snapshot of the wave at a particular instant.

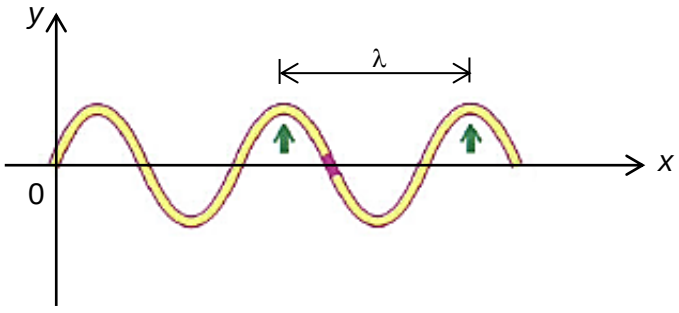


Fig. 9.4.2. Displacement-position graph of a wave on a rope.

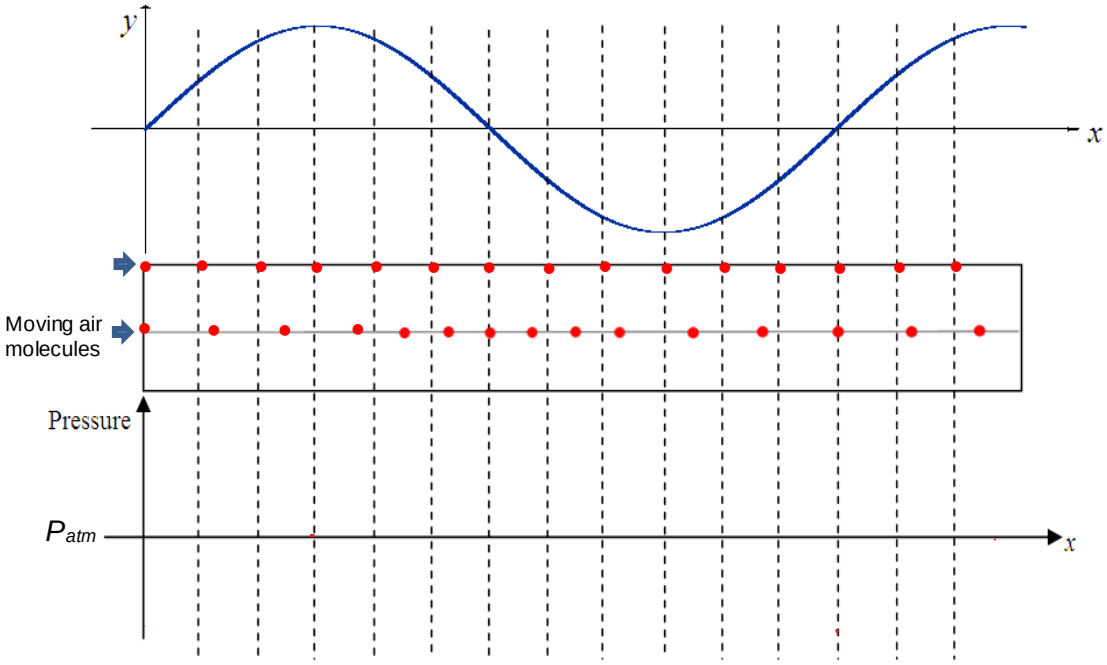
Note :

- +ve y : displacement of a point on the wave when it is above its equilibrium position
- -ve y: displacement of a point on the wave when it is below its equilibrium position
- x axis (y = 0) : equilibrium position of points when there is no disturbance

$\Rightarrow$ ***Wavelength (λ) of the wave can also be obtained from the graph*.**

Example 9.4.1: Displacement-Position Graph for a Longitudinal Wave

A displacement–position graph can also be drawn for a longitudinal wave. Consider the following displacement-position graph for the air molecules along a line when a sound wave passes through a region of space. Sketch the corresponding pressure variation of this region.



Note: Points of **compression and rarefactions** corresponds to points where **displacements of air molecules are zero.**

Compressions are regions of **high air pressure**, while **rarefactions** are regions of **low air pressure**.

Graph of Displacement (*y*) vs. Time (*t*)

Consider a point P on a rope. As a wave propagates through the rope (Fig. 9.4.3), the displacement of this point varies sinusoidally with respect to time (as shown in Fig. 9.4.4).

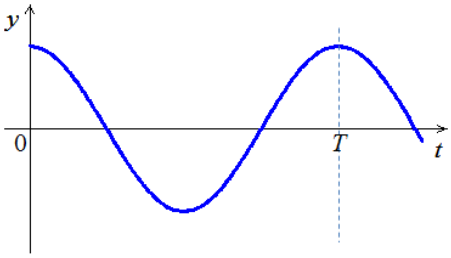


Fig. 9.4.4.

The graph of displacement vs. time for a wave gives us **the variation of the displacement of a particular point on a wave (at a fixed position *x*) with respect to time.** You can imagine it as a motion video of a particle in a wave.

⇒ **The period (*T*) associated with the wave can be obtained from this graph.**

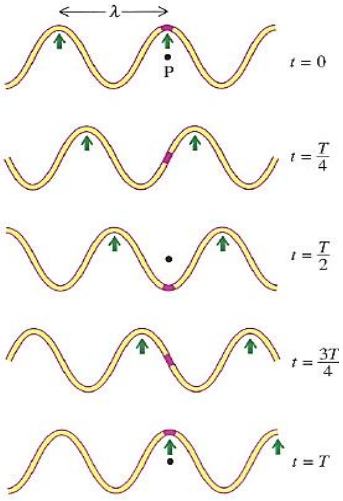


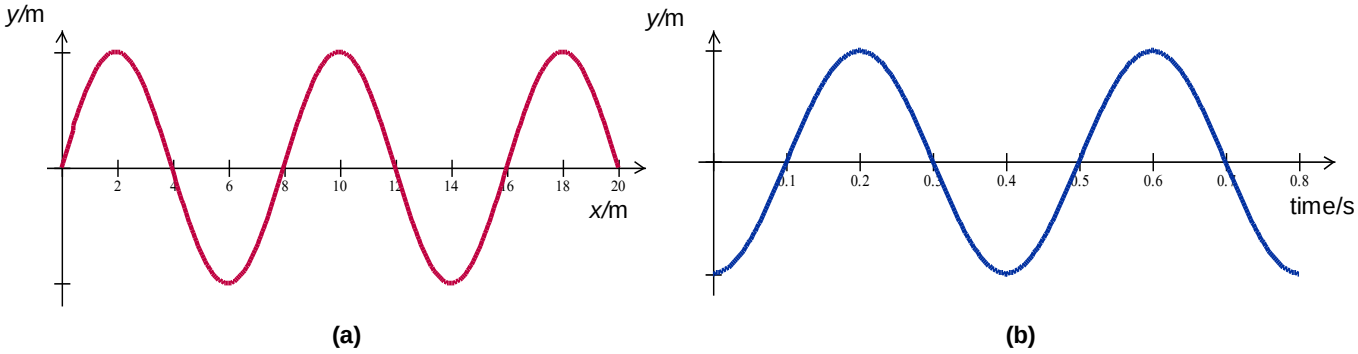
Fig. 9.4.3. A sinusoidal wave travelling on along a rope.

In summary, two graphs are needed to fully depict a wave:

- The graph of **displacement vs. position** (wave profile of a wave at a certain instant of time),
  - ⇒ Wavelength λ may be obtained
- The graph of **displacement vs. time** (Motion of a particular point on a wave w.r.t. time)
  - ⇒ Period *T* of the waves may be obtained.

Example 9.4.2: Determining the speed of a wave

A vibrating bar makes waves in a ripple tank. Fig. 9.4.5(a) shows the displacement of the wave as it travels out from the bar. The position of the floating cork in the tank varies with time as shown in Fig. 9.4.5(b).



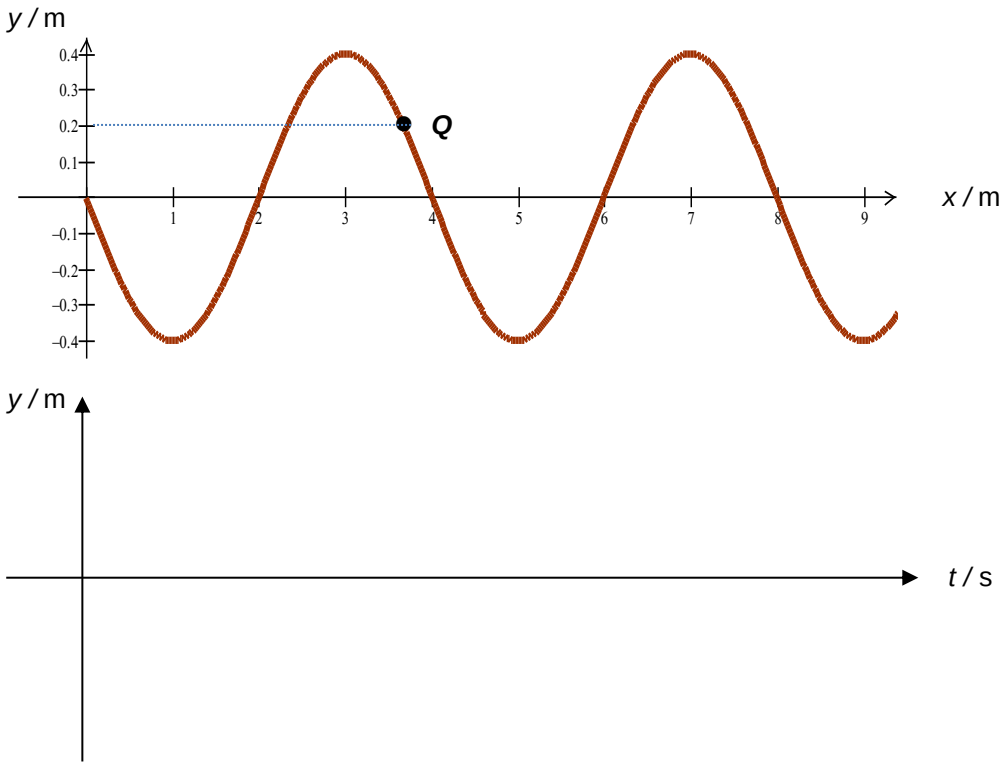
(a) **(b)**

Fig. 9.4.5

From the above two graphs, determine the speed of the wave.

Example 9.4.3

The following graph shows a displacement vs. position graph of a progressive transverse wave at $t = 0$. For the point **Q** on the wave, sketch the corresponding displacement vs. time graph, given that the wave is moving in the positive x direction at a speed of 1.0 m s^{-1}.



9.5 Phase and Phase Difference

Phase of a particle in a wave at any given time indicates the particle's current state of vibration, usually expressed as an angle. (Also refer to Oscillations LN p.15)

Phase difference in waves refers to the difference in phase:
(a) at two different instances of time for a point on a wave
(b) two points on a wave, or
(c) between two waves.

Note the axes for the subsequent explanations of (a) to (c).

(a) Phase difference (Δø) for a point on a wave at two different time instances

For a sinusoidal progressive wave, every point on the wave oscillates simple harmonically. Hence, the phase difference, Δø at two different instances of time for a point on a wave that are at an interval of Δt apart from **displacement vs. time graph** is given by:

$$\frac{\Delta t}{T} = \frac{\Delta \phi}{2\pi}$$

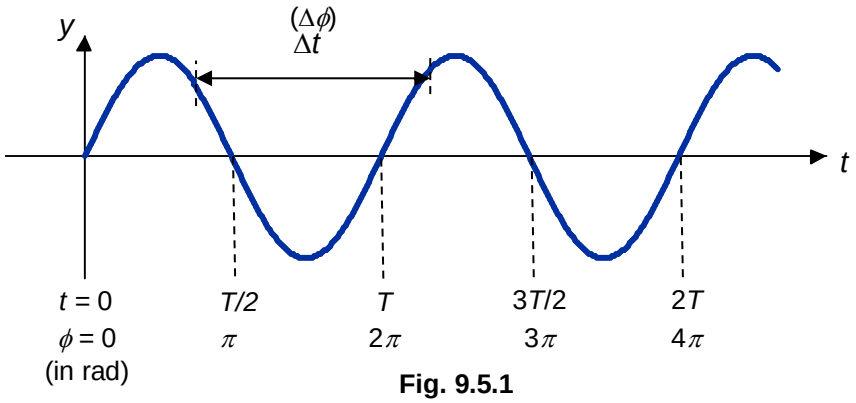


Fig. 9.5.1

(b) Phase difference between two points on a wave

(i) One complete wavelength is represented by a phase angle of 360° or 2π radians.
Therefore, the **phase difference $\Delta\phi$ between two points on a wave separated by a distance Δx from displacement vs. position graph** is given by:

$$\frac{\Delta x}{\lambda} = \frac{\Delta \phi}{2\pi}$$

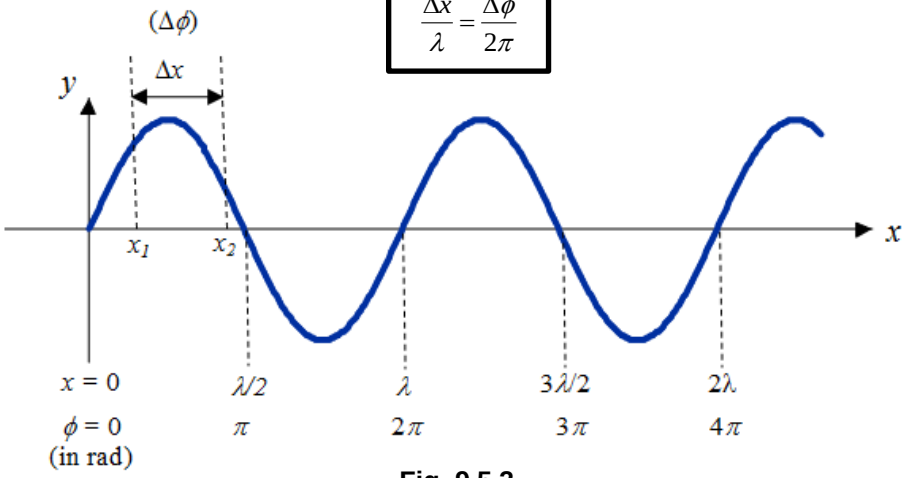


Fig. 9.5.2

(ii) The displacement vs. time graphs of the two points on a wave can also be used to compare the state of oscillation on the wave.

We can plot on the same axes the **displacement vs. time graphs** for the two points on the wave in Fig. 9.5.2 (one graph for each point). If the wave is progressing in the positive x direction, we will get the following graphs for the points at x_1 and x_2 as shown in Fig. 9.5.3:

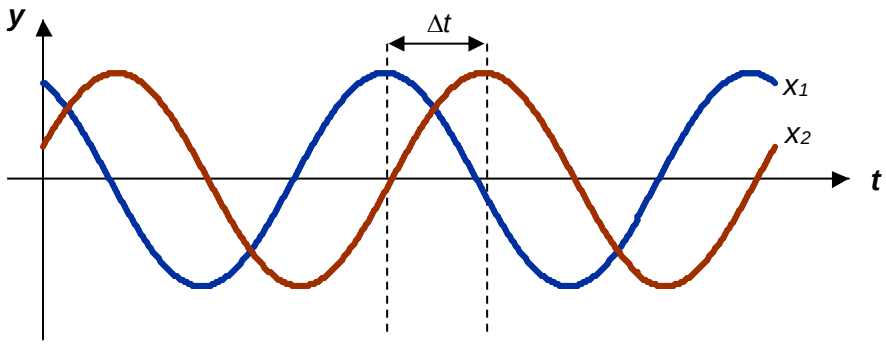


Fig. 9.5.3

From the graphs above, we can clearly see that, the oscillation of the point at x_2 always lags behind the oscillation of the point at x_1 by a time interval of Δt, and hence point at x_2 lags behind point at x_1 by a phase angle of $\Delta\phi$ where

$$\frac{\Delta t}{T} = \frac{\Delta\phi}{2\pi}$$

(c) Phase difference between two waves

We can also compare *the phase (angles) between two waves (usually, having the same frequency)* by comparing the oscillations of every point on a wave at a particular instant.

Consider two waves ***P*** and ***Q*** having the same wavelength and velocity moving towards the right as shown in Fig 9.5.4.

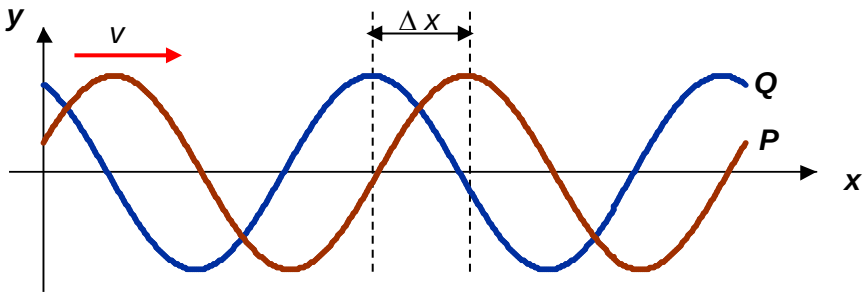


Fig. 9.5.4

Notice that the Wave P always leads Wave Q by a constant distance of Δx. If the wave Q shifts a distance towards the right by Δx then every point on Wave Q will be in the same phase of oscillation on points on wave P at this moment.

Therefore, we say that

- *Wave Q lags Wave P by a phase difference of $\Delta\phi$*, and
- the **phase difference between the two waves is $\Delta\phi$** can be found from: $\frac{\Delta x}{\lambda} = \frac{\Delta\phi}{2\pi}$

Note :

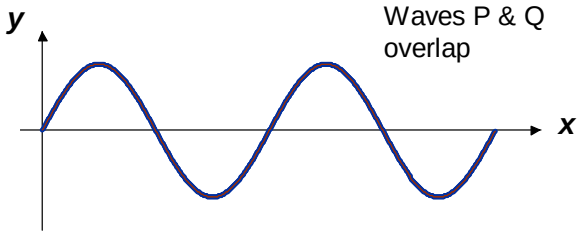
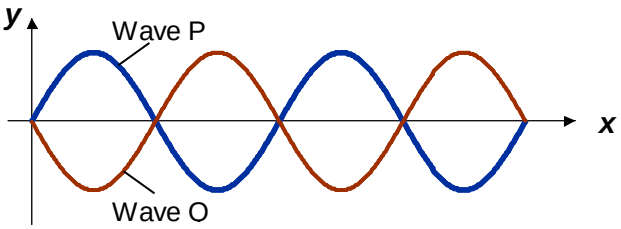


Fig. 9.5.5 (a) Wave P is in phase with wave Q



(b) Wave P is in antiphase with wave Q

1. **The two waves are said to be in phase** if every point on wave **P** is at the same state of oscillation as the corresponding point of the same position in wave **Q** at the same time. The phase difference between the two waves are $\Delta\phi = 0, 2\pi, 4\pi, \ldots$
   - Notice that when the two transverse waves are in phase, their "*crests*" will correspond and their "*troughs*" will also correspond.

2. When the phase difference between the two waves are $\Delta\phi = \pi, 3\pi, 5\pi, \ldots$, we say that the **two waves are in antiphase.**
   - When the two waves are in antiphase, every particle in Wave P is in antiphase with the corresponding point.
   - Note that the waveforms seem like a reflection of one another.

Example 9.5.1 :

(a) The following graph shows the displacement vs. position of two waves **P** and **Q** at a particular instant of time, what is the phase difference between the two waves **P** and **Q**?

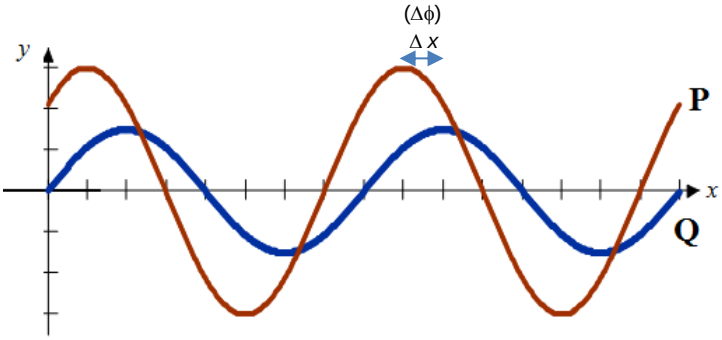


Fig. 9.5.6

(b) Is it possible to compare the phase difference between two waves of different wavelengths? What is the phase difference between the two waves shown below?

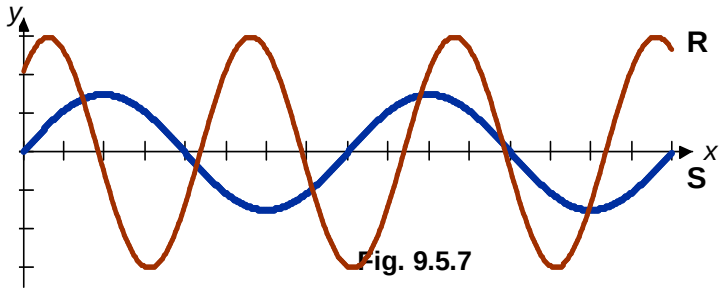


Fig. 9.5.7

Answer:

No. As Δx is not constant for the 2 waves, the phase difference between the two waves keeps changing value.

It is only meaningful to talk about phase difference between 2 waves only if they have the same wavelength or same frequency.

9.6 Energy Transmitted by a Wave

As a wave propagates, it transports energy from one place to another. Consider a sinusoidal wave moving through a string shown in Fig. 9.6.1, the source of energy is some external agent at the end of the string which does work in producing the oscillations.

Fig. 9.6.1

As the external agent performs work on the string, it transfers energy to the string in the form of kinetic energy and (elastic) potential energy to each part of the string.

From Chapter 8 Oscillations, we know that the total energy *E* of a simple harmonic oscillator is given by $E = \frac{1}{2}m\omega^2 A^2$. Since every part of the string is in simple harmonic motion,

Energy transmitted by a wave to a point is proportional to the square of the amplitude of the oscillator at that point.

$$\text{Energy} \propto (\text{Amplitude})^2$$

Intensity of a Wave

Waves on a string carry energy in one dimension of space (along the direction of the string). But most of the waves that we commonly encounter like the sound waves in air and seismic waves in the body of the earth carry energy along all three dimensions of space. As the wave spreads out in space, the power received per unit area along the path of the wave diminishes as the distance from the source of the wave increases.

- **Power of a Wave:**

  The amount of energy transported per unit time by a wave is called the *power* of a wave.

  The SI unit is joules per second (J s^{-1}) or watts (W).

Wave Intensity, *I* :

The **intensity of a wave** is the rate at which energy is transported by the wave, per unit area, across a surface perpendicular to the direction of propagation i.e. it is the power per unit area.

$$Intensity, I = \frac{Power}{Area}$$

- **S.I. unit of intensity:** W m^{-2}

Since the energy transmitted by a wave is proportional to the square of the amplitude of the wave, so similarly, we have

$$\text{Intensity} \propto (\text{Amplitude})^2$$

Intensity of a Wave from Point Source or Spherical Source

In the special case where the source is **point source** or **spherical source**, and the propagation is *isotropic* (i.e. the same in all directions), then the wavefronts of the waves are spherical in shape. The wave itself is said to be a *spherical wave*.

The energy spreads out over the wavefront as the wave propagates, hence, the intensity *I* at a particular point a distance *r* from the source is given by :

$$\text{Intensity, } I = \frac{\text{Power}}{\text{Area}} = \frac{P}{4\pi r^2}$$

where *P* is the time rate of energy delivered by the wave to a distance of *r* away from the source.

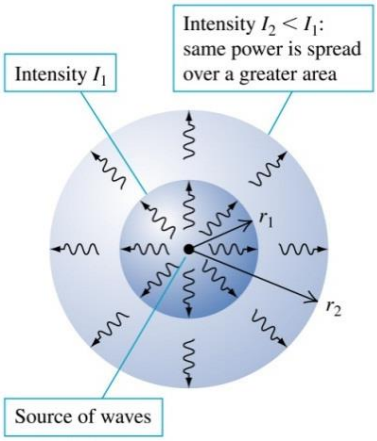


Fig. 9.6.2

Therefore, we see that the **intensity *I* at any distance *r* is therefore inversely proportional to r^2.**

$$I \propto \frac{1}{r^2} \qquad \textit{(For constant P)}$$

Note:

- This relationship is called the *inverse-square law for intensity*.
- This relationship is only valid for *point sources / spherical sources* and is based on the assumption that **the power of the wave does not diminish as it spreads out.**

Example 9.6.1 : (N94/I/10 – modified)

A sound wave of amplitude 0.20 mm has an intensity of 3.0 W m^{-2}.

(a) What will be the intensity of a sound wave of the same frequency which has an amplitude of 0.40 mm?

(b) How will the intensity change if the frequency of the source of wave is doubled but the amplitude remains unchanged? Treat the source as a simple harmonic oscillator.

Example 9.6.2 :

If the intensity of an earthquake *P*-wave is 1.0 x 10^6 W m^{-2} at 100 km from the source, what is the intensity of the wave 400 km from the source?

9.7 Polarisation of Waves

One characteristic that distinguishes *transverse waves* and *longitudinal waves* is *polarisation*[4].

Plane Polarisation

Polarisation is a phenomenon whereby vibrations in a transverse wave are restricted to only one direction in the plane normal to the direction of energy transfer.

To polarise a wave, we often use a *polarising filter*, or *polariser*, that permits only waves with a certain polarisation direction to pass through.

Consider waves set up on a string. A slot as shown in Fig. 9.7.1 can be used to limit the string's plane of propagation to the x-y plane and the oscillation to the y-axis.

Polarising Axis or Transmission Axis

The polarisation direction of the polariser is also known as the ***polarising axis*** or ***transmission axis***.

Only components of the waves parallel to the polarising axis are allowed to pass through.

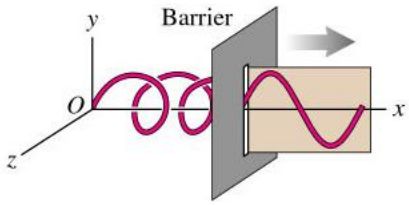
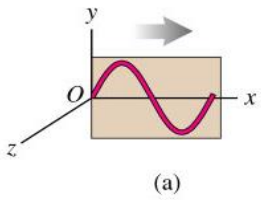


Fig. 9.7.1 A vertical slot is used to polarise the wave on the string so that oscillations are limited to a vertical plane.



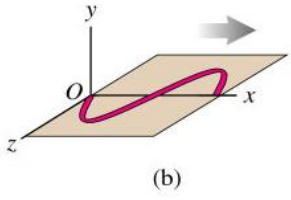


Fig. 9.7.2 A wave on a string polarised to oscillate in the *x-y* plane (a) and *x-z* plane (b). For each case, what is the polarizing axis on the polarizer?

[4] For A-level syllabus, we will consider only plane polarisation, where we limit the oscillations to a plane. If you are interested in knowing more about Circular Polarisation or Elliptical Polarisation, you may refer to Young and Freedman (11th Edition) Chapter 33.5 Pg 1268.

Why Longitudinal Waves cannot be polarised?

Polarisation **does not occur** for longitudinal waves because *oscillations of the points on the wave are along the direction of wave propagation*. Restricting the direction of oscillation will in turn restrict the direction of wave propagation.

Therefore, **longitudinal waves cannot be polarised.**

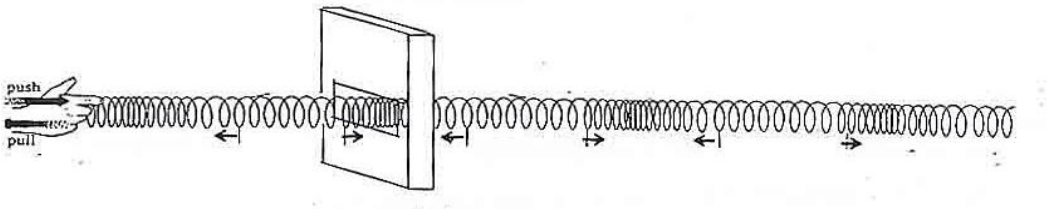


Fig. 9.7.3 Longitudinal wave on a slinky cannot be polarised as the direction of oscillation of the elements on the slinky in parallel to the direction of propagation of the longitudinal wave. Limiting the direction of oscillation will therefore also limit the motion of the wave.

In summary,

1. Only transverse waves can be polarised. Longitudinal waves cannot be polarised.
2. Only the component of the wave that is parallel to the **polarising axis** of the polariser will pass through the polarisers.

9.7.1 Polarisation of Light

Light waves are electromagnetic (EM) waves which consist of *mutually perpendicular time varying electric field ($\vec{\mathbf{E}}$) and magnetic field ($\vec{\mathbf{B}}$) oscillating perpendicularly to the direction of wave travel*. (See Fig 9.2.2 on p.2)

EM wave is commonly represented by *just the electric field* ($\vec{\mathbf{E}}$). The arrows indicate the plane of oscillation of the **E**-field.

- Fig. 9.7.4 shows a plane polarised EM wave travelling out of the paper towards you.
- Fig. 9.7.5 shows an *unpolarised* EM wave travelling out of the paper. Note the multiple planes of oscillation in the unpolarised EM wave.

Fig. 9.7.4 Polarised EM Wave

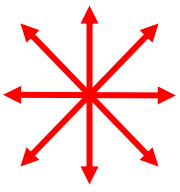


Fig. 9.7.5 Unpolarised EM Wave

When unpolarised light is incident on an ideal polariser, the intensity of the transmitted light is exactly half that of the incident unpolarised light, no matter how the polarising axis is oriented.

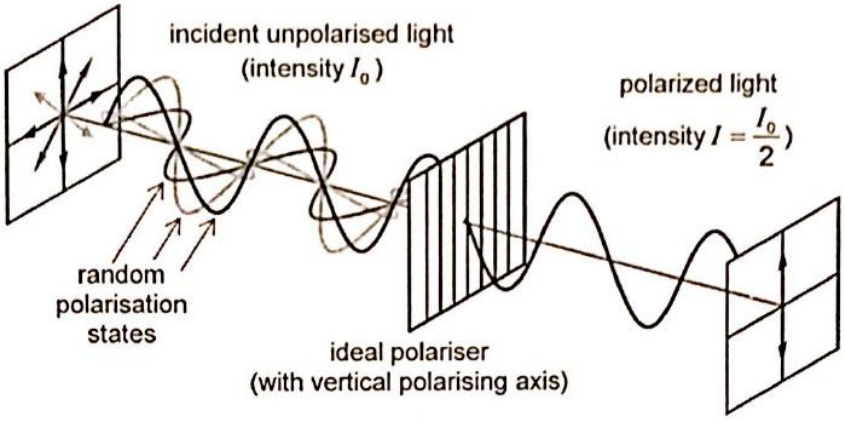


Fig. 9.7.6 Intensity of plane polarised light

Since incident light is a random mixture of all polarisation states, the components of the electric field vectors that are perpendicular or parallel to the polarising axis is roughly equal. By only transmitting the component that is parallel to the polarising axis, **only half the incident intensity** is transmitted.

An ideal polariser passes 100% of the incident light that is plane-polarised along its polarisation direction and 0% of the light that is 90° to its polarisation direction. Since unpolarised light is a mix of all possible polarisations, an ideal polariser transmits only 50% of the incident unpolarised light intensity.

Light Passing Through Two Polarisers

By using two polarisers, we can also control the intensity of light.

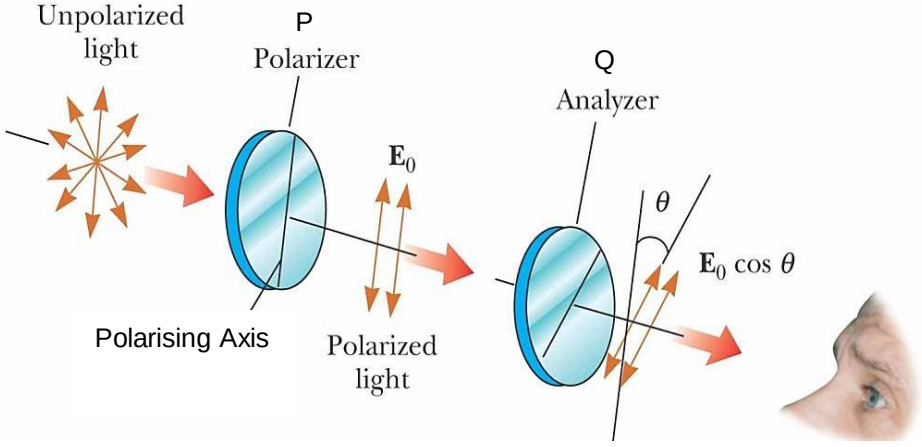


Fig. 9.7.7 Using two polarisers to control intensity of light.

Fig. 9.7.7 shows the typical setup used to control the intensity of light. An unpolarised light beam is incident on two polarisers **P** and **Q** with their polarising axes at an angle of θ with each other.

- In such a setup, the second polariser is often known as an **analyser**.
- After passing through the first polariser **P**, the light is polarised along the polarising axis of polariser P. Suppose the light then has an amplitude of $\mathbf{E_o}$ and an associated maximum intensity of I_o.
- If this polarised light then falls on a second polariser Q, *then the component of* $\boldsymbol{E_o}$ *perpendicular to the analyser axis is completely absorbed, while only the component parallel to polarising axis is allowed to transmit through Q.*
- Since the plane of oscillation of the electric field of this light is oriented at an angle of θ to the polarising axis of Q, the resultant amplitude of vibration of the light after passing through the polariser Q is given by $\mathbf{E = E_o \cos \theta}$.

- Since the intensity of a wave is directly proportional to the square of the amplitude of vibration, therefore,

$$\frac{\text{Intensity of Light after passing through the analyser Q}}{\text{Intensity of Light after passing through the polariser P}} = \frac{I}{I_o} = \left(\frac{E_o \cos\theta}{E_o}\right)^2$$

$$\Rightarrow \qquad I = I_o \cos^2\theta$$

Hence the intensity of light that passes through two polarisers that are aligned with an angle between their polarisation directions is

$$\boxed{I = I_o \cos^2\theta}$$

where I_o is the maximum intensity of light transmitted when the two polarisers are aligned along the same direction of polarisation. The above equation is known as **Malus' law** after its discoverer Etienne Louis Malus, who first observed this effect experimentally in 1809.

Malus's law applies only if the incident light passing through the analyzer is already *linearly polarized.*

Malus' law states that the intensity of a beam of plane-polarised light after passing through a polariser varies with the square of the cosine of the angle through which the polariser is rotated from the position that gives maximum intensity.

Based on Malus' law, when 2 polarisers are aligned at 90° with respect to each other, zero intensity of light passes through the second polariser.

Web Simulation on Polarization: https://javalab.org/en/polarization_en/

Some applications of polarised light are:

1. Polarised sunglasses – Glare is usually due to light reflected off a horizontal ground (e.g. snow surface, car roof). As light gets reflected off a horizontal surface, it will be polarized horizontally. Polarised lenses contain a special filter which block off the reflected light from the ground, which reduces the glare.

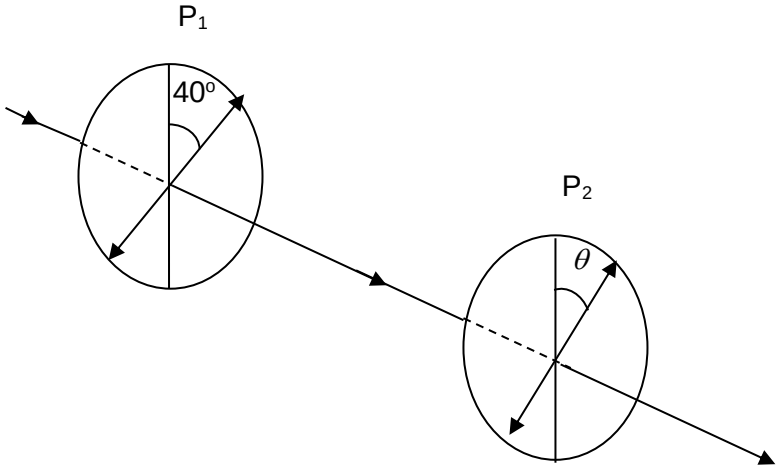
2. Stress Analysis - Certain plastic materials under stress will affect the plane of polarization of light through them. This can be used to study stress in various structures.

3. 3-D films – Two images are projected through polarising filters onto a screen. The viewer will have to wear a pair of glasses which has 2 different polarising filters for each lens, so that the movie appears to be three dimensional. (see Appendix B 3-D films and Polarisation)

Example 9.7

The figure below shows a beam of initially unpolarised light passing through two polarizing filters P_1 and P_2. The polarising axis of P_1 is fixed at 40° with respect to the vertical axis.

The polarizing axis of P_2 is then rotated clockwise from its vertical axis. At what values of θ will intensity minima of the emergent light occur?



For intensity minima to occur (no emergent light), the two polarizing axes must be

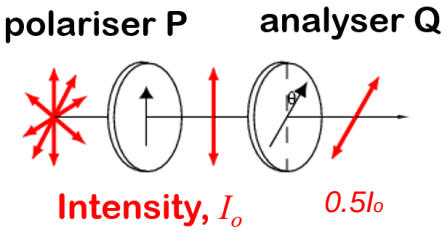
_____ to each other.

Rotate P_2 through _____ with respect to P_1 , therefore θ =

Rotate P_2 through _____ with respect to P_1 , therefore θ =

Example 9.8

Two polarising sheets have their polarizing directions parallel so that the intensity I_o of the transmitted light is a maximum. Through what angle must either sheet be turned if the intensity is to drop by half?



Appendices

Appendix A: Function Generator and Cathode Ray Oscilloscope (C.R.O)

Function Generator

A function generator is an instrument that produces alternating (a.c.) voltage of different frequencies.

The standard voltage signals it can produce are:

- sine waves,
- triangular waves
- rectangular waves

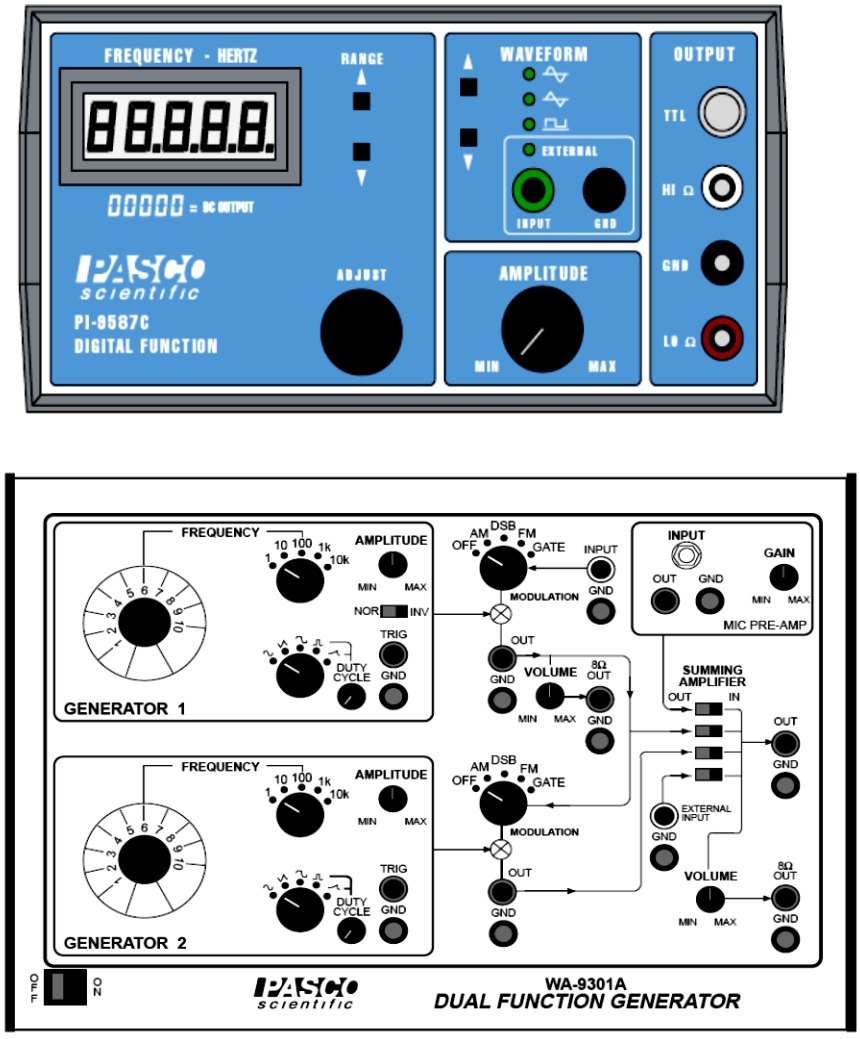


Fig. 1. Examples of function generators. The top one is a single output generator. The bottom generator is able to produce two outputs at the same time.

As not all function generators have a precise digital display of the frequency of the waveform or the amplitude of the voltage wave produced. It will sometimes be necessary to connect the function generator to a CRO in order to determine the precise frequency and amplitude of the voltage wave.

To generate the sound waves, the function generator is often connected to a speaker so that the speaker can convert the voltage signals to vibrations of the membrane of the speakers to produce sound waves.

Interactive Online Virtual Oscilloscope and Signal Generator: https://eleceng.dit.ie/dsp/elab/

Cathode Ray Oscilloscope

The other commonly used equipment in sound experiments is the **cathode ray oscilloscope** (CRO). The oscilloscope is basically a graph displaying device that traces the variation of the input voltage signal with respect to time.

Using the oscilloscope, you can measure the amplitude, period, and frequency of any voltage signal. Therefore, a CRO is actually a more complex voltmeter. Hence, to measure the frequency of a sound wave, we need to first connect a microphone to the scope. The microphone will then convert the vibrational energy of the air molecules to voltage signals which can be measured by the CRO.

Most scopes can display two signals on the screen at one time, enabling you to observe their time relationships. These scopes are known as dual-channel oscilloscopes. A typical front panel view of such a scope is shown in Fig. 2.

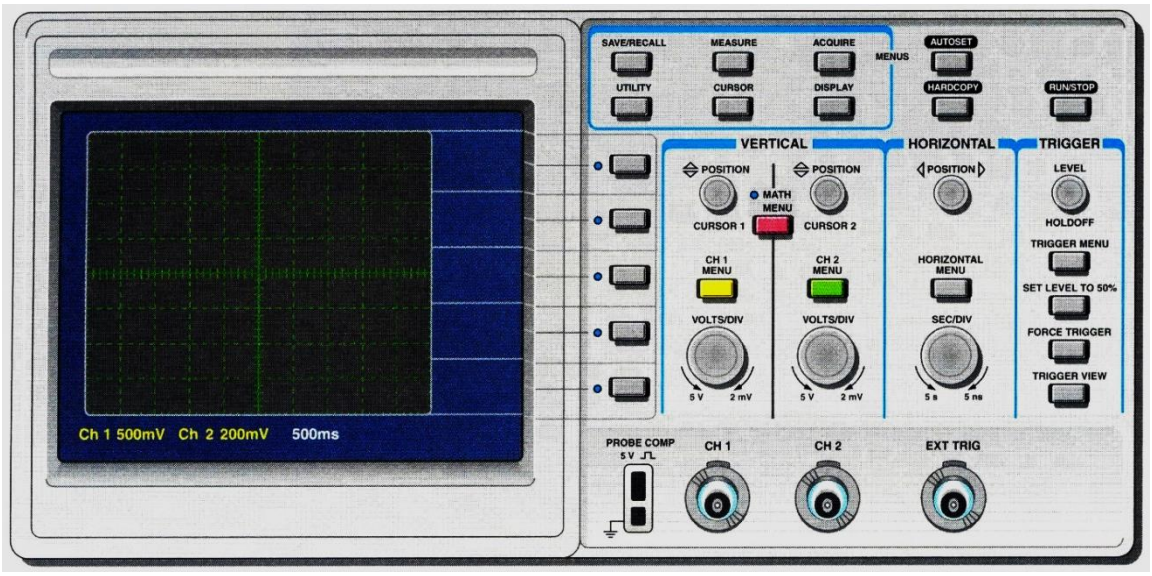


Fig. 2. A typical dual channel oscilloscope. Numbers below screen indicate the values for each division on the vertical (voltage) and horizontal (time) scales and can be varied using the vertical and horizontal controls on the scope. Although this may be a typical dual-channel oscilloscope, it does not look like the ones we have in the lab!

Actual CROs vary depending on model and manufacturer, but most have certain common features. The main controls will be discussed below. You will need to refer to the user manual for complete details of the scope you are using.

Vertical Controls:

In the vertical section of the scope (Fig. 2), there are identical controls for each of the two channels (CH1 and CH2).

- **Position Control**. The position control lets you move a displayed waveform up or down vertically on the screen.
- **Volts/Div control.** The Volts/Div (volts per division) control is also known as the **Y-sensitivity / Y-gain control**. It adjusts the number of volts represented by each vertical division (usually 1 cm) on the screen. For this scope, the Volts/Div setting for each channel is displayed at the bottom of the screen. For CH1, it is 500 mV/div and for CH2, it is 200 mV/div.

Horizontal Controls:

In the horizontal section, the controls apply to both channels.

- **Position Control**. The position control lets you move a displayed waveforms left or right horizontally on the screen.

- **Sec/Div Control.** The Sec/Div (seconds per division) control is also known as the **Timebase** control, it adjusts the time represented by each horizontal division or main time base. Similarly, for this scope, the Sec/Div setting is displayed at the bottom of the screen.

Appendix B: 3-D films and Polarization

3-D films such as Avatar (by 20th Century Fox), G-Force (by Walt Disney Pictures) and Cloudy With A Chance of Meatballs (by Sony Pictures Animation) have been a hit with the cinema these days. In fact, 3-D films have existed in some forms since 1950s. There are various techniques to produce 3-D films. Current methods include interference filter technology, for example Dolby 3D and polarization systems, for example RealD. We will look at the latter in detail.

For 3-D films which are shown using polarization systems, the viewers will have to wear 3-D glasses to view the polarized image on the screen. The image will be blurry when viewed without the special glasses. There are two types of polarization systems: linear polarization and circular polarization.

For the linear polarization system, two images are superimposed onto the same screen through two polarizing filters which are at 90 degree angles of each other (orthogonal). The silver screen is specially constructed such that it is non-depolarizing to compensate for light loss and to preserve the polarization. The viewer wears a pair of glasses which is made of a pair of orthogonal polarizing filters. Each filter only allows light from the image which is polarized in the same plane to pass through and blocks the orthogonally polarized light, each eye will only see one of the images. The 3-D effect is then achieved.

Fig. A1. Two pairs of 3-D glasses showing the polarizing effect.

Similarly, for a circular polarization system, two images polarized differently are superimposed onto the same screen. However, unlike linear polarization system, the two images pass through circular polarizing filter of opposite handedness. The viewer wears a pair of glasses made of a pair of circular polarizer mounted in reverse handedness. Hence, each eye will only see one of the images and the 3-D effect is achieved. RealD Cinema is a digital stereoscopic projection technology which uses similar techniques as a circular polarization system; except that it only needs one projector. A circularly polarizing liquid crystal filter which can switch polarity is placed in front of the projector lens, so that the left and right images are displayed alternately.

Information adapted from:
http://entertainment.howstuffworks.com/digital-3d1.htm
http://en.wikipedia.org/wiki/RealD_Cinema
http://en.wikipedia.org/wiki/3-D_film

Appendix C: Polarisation by Absorption[5]

Most light sources e.g. filament lamp, sun etc. produce light that is unpolarised. However, there are ways to polarise light that is initially unpolarised, or to change the direction of polarisation of polarised light. The most common method to produce polarised light waves is by using a polariser which is known as a ***polaroid***. (E.H. Land, 1928).

When light is incident on a polaroid sheet, ideally,

- if the light is polarised with **its electric field E <u>parallel</u> to the polarising axis of the polaroid, it will be <u>completely transmitted</u>,**
- if the light is polarised with **its electric field E <u>perpendicular</u> to the polarising axis of the polaroid, it will be <u>completely absorbed.</u>**
- if the light is **linearly polarised at some angle θ relative to the polarising axis**, the light will be partially absorbed and partially transmitted - **only the component of the electric field E parallel to the polarising axis will be transmitted.**

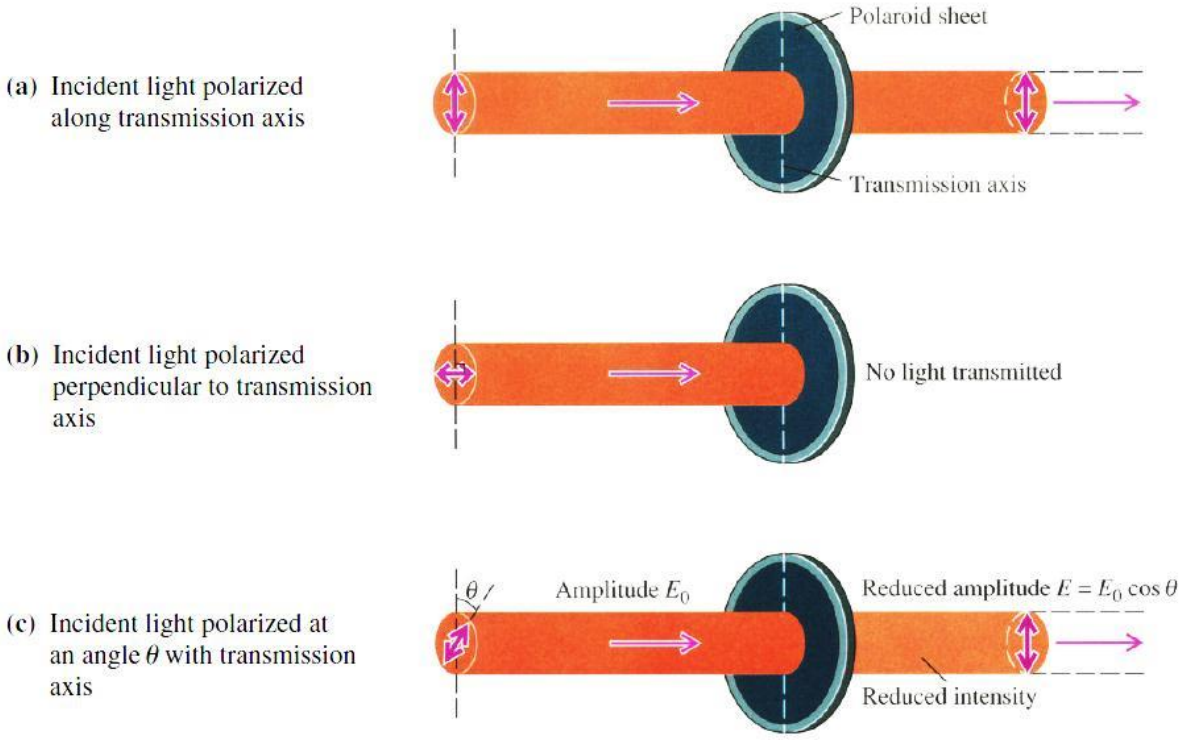


Fig. A2. The effect of a Polaroid sheet on initially polarized light depends on the direction of the initial polarisation relative to the direction of the sheet's transmission axis. (Source: Coletta, p748, Fig. 26.36.)

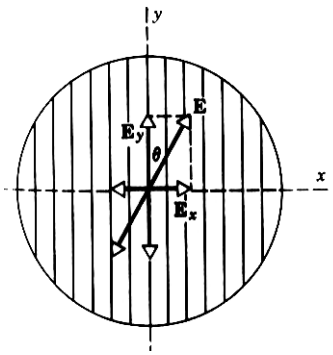


Fig. A3. The figure on the left shows what happens when polarised EM wave falls on a polariser with a vertical polaring axis. The electric field E of the EM wave makes an angle θ with the polarisation axis. The vertical component of E has magnitude $E_y = E\cos\theta$ and is transmitted. The horizontal component $E_x = E\sin\theta$ is absorbed. As a result of polarization, the amplitude of light that passes through has decreased. Hence, the intensity of the light is reduced.

[5] It is also possible to polarise light through reflection. You may read more about it from our e-textbook "Physics Fundamentals" by Coletta, p750.

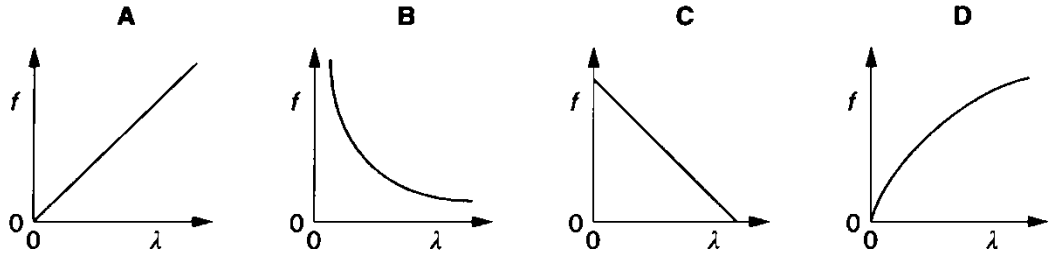
Tutorial 9 Wave Motion

Self-Review Questions

S1 N99/I/10.

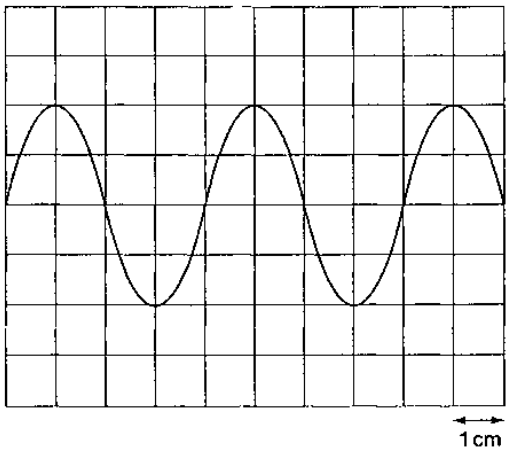
A sound wave of frequency *f* and wavelength λ travels through air. It may be assumed that its speed is independent of the frequency.

Which graph correctly shows the variation of *f* with λ ?



S2 N05/I/17

The diagram shows the trace produced by a sound wave on a cathode-ray oscilloscope. The time base is calibrated at 2.00 ms cm^{-1}.

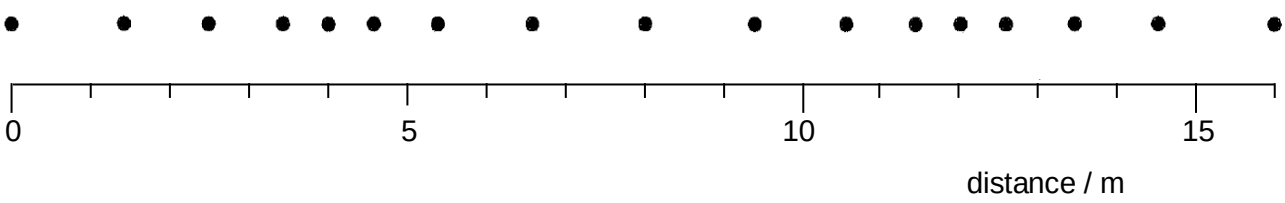


What is the frequency of the sound wave?

A 62.5 Hz **B** 125 Hz **C** 250 Hz **D** 500 Hz

S3 N20/I/18

The diagram illustrates the position of particles along a longitudinal wave at an instant in time.

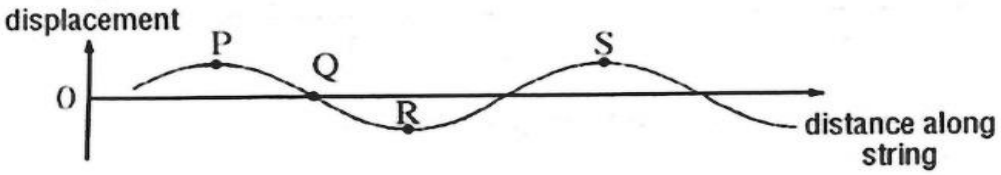


The wave travels at a speed of 12 m s^{-1}.

What is the wave frequency?

A 0.33 Hz **B** 0.67 Hz **C** 1.5 Hz **D** 3.0 Hz

S4 The graph shows a segment of a transverse wave travelling on a string at a particular point in time.



Which statement about the motion of the elements of the string is correct?

A The speed of the element at P is a maximum.

B The displacement of the element at Q is always zero.

C The energy of the element at S is a maximum.

D The acceleration of the element at S is a maximum.

S5 **J87/I/11**

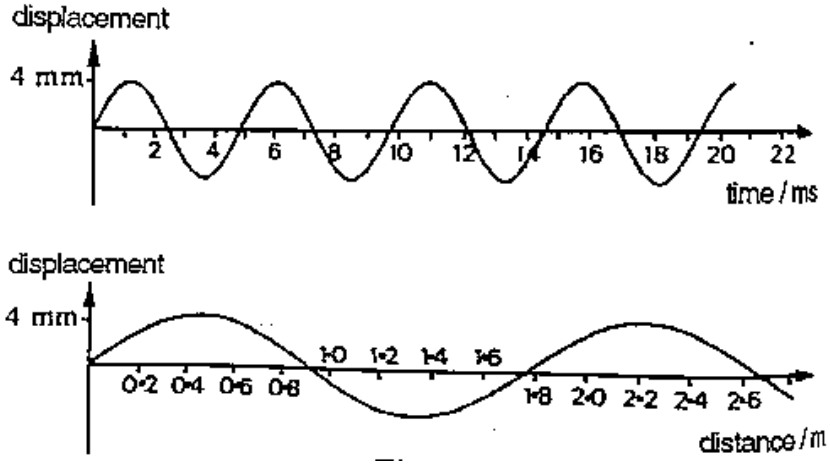
Visible light has wavelengths between 400 nm and 700 nm. What is the maximum frequency of visible light?

S6 Which of the following gives three regions of the electromagnetic spectrum in order of increasing wavelength?

A γ-rays, microwaves, visible radiation
B radio waves, ultraviolet, X-rays
C ultraviolet, infra-red, microwaves
D visible radiation, γ-rays, radio waves

S7 **N90/III/3**

The figure below shows two graphs which refer to the same wave. What is the speed of the wave?



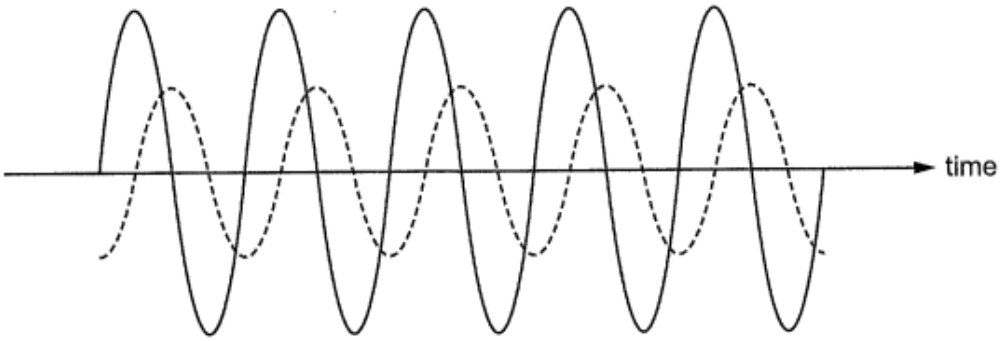
S8 J79/II/13; N82/II/10; N90/I/13; J99/I/11

A sound wave of frequency 400 Hz is travelling in a gas at a speed of 320 m s^{-1}. What is the phase difference between 2 points 0.2 m apart in the direction of travel?

A $\frac{\pi}{4}$ rad **B** $\frac{\pi}{2}$ rad **C** $\frac{2\pi}{5}$ rad **D** $\frac{4\pi}{5}$ rad

S9 N17/I/14, N20/I/17

Two waves of the same frequency are displayed on the screen.



What is the phase difference between the waves?

A $\frac{1}{4}$ rad **B** $\frac{\pi}{4}$ rad **C** $\frac{\pi}{2}$ rad **D** $\frac{3\pi}{4}$ rad

S10 (Young and Freedman)

By measurement, you determine that sound waves are spreading out equally in all directions from a point source and that the intensity is 0.026 W m^{-2} at a distance of 4.3 m from the source.

(a) What is the intensity at a distance of 3.1 m from the source?

(b) How much sound energy does the source emit in one hour if its power output remains constant?

S11 N05/I/16

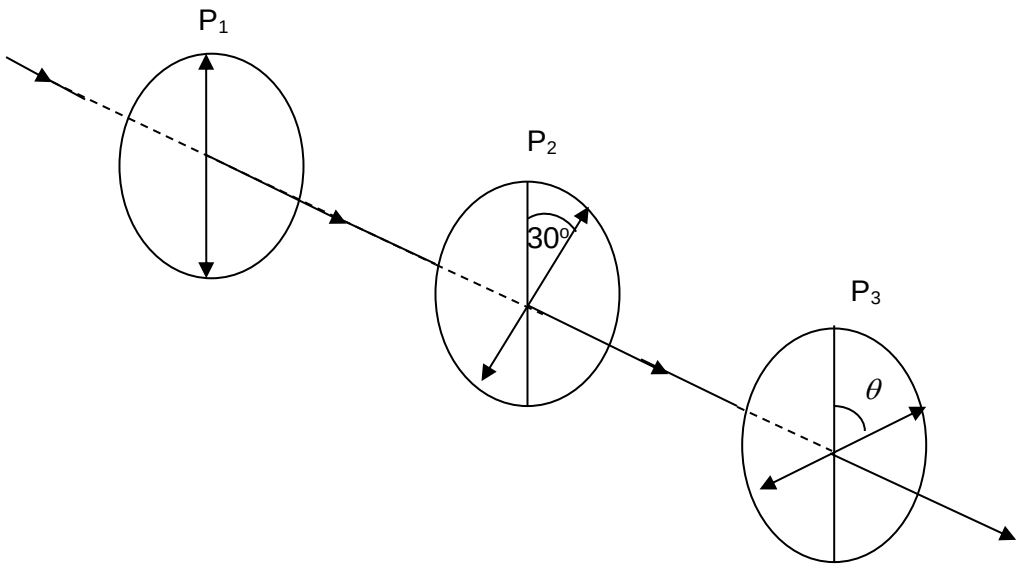
The intensity of a wave depends on the amplitude. The intensity is also proportional to the square of the frequency. A wave has frequency 3.0 Hz, amplitude 1.5 cm and intensity I.

What is the intensity of a similar wave of frequency 6.0 Hz and amplitude 0.5 cm?

A $\frac{4}{9}I$ **B** $\frac{4}{3}I$ **C** $\frac{9}{4}I$ **D** 36 I

S12 J82/II/13

Figure below shows a beam of initially unpolarised light passing through three polaroids P_1, P_2 and P_3. The polarising axis of each polaroid is shown by an arrow. Polaroids P_1 and P_2 are fixed, with their polarising axes at 30° to one another, and P_3 can be set with its polarising axis at a variable angle θ to that of P_1.



For what values of θ do intensity minima of the emergent light occur?

A 30°, 120°, 210°, 300°
B 90°, 120°, 270°, 300°
C 60°, 240°
D 90°, 270°
E 120°, 300°

S13 Two ideal polarisers are aligned with a 30° angle between their polarization directions. If unpolarised light of intensity I_o is incident upon them, what is the intensity of the transmitted light?

Numerical Answers for Self-Review Questions

S1 B **S2** B **S3** C **S4** D **S5** 7.5×10^{14} Hz **S6** C **S7** 372 m s^{-1} **S8** B **S9** C
S10 (a) 0.0500 W m^{-2}, (b) 2.17×10^4 J **S11** A **S12** E **S13** 0.375 I_0

Discussion Questions

D1 N87/I/8

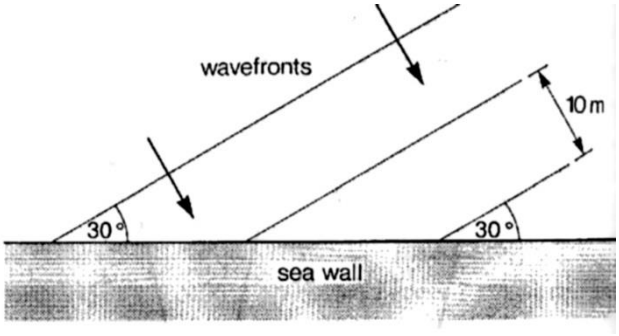
Transverse progressive sinusoidal waves of wavelength λ are passing vertically along a horizontal rope. **P** and **Q** are points on the rope $\frac{5\lambda}{4}$ apart and the waves are travelling from P to Q.

Which one of the following correctly describes Q at the instant when P is displaced upwards but is moving downwards?

	Displacement of Q	Movement of Q
A	upwards	downwards
B	upwards	upwards
C	downwards	upwards
D	downwards	downwards
E	downwards	stationary

D2 J85/I/8; N95/I/10.

Parallel water waves of wavelength 10 m strike a straight wall. The wavefronts make an angle of 30° with the wall as shown.



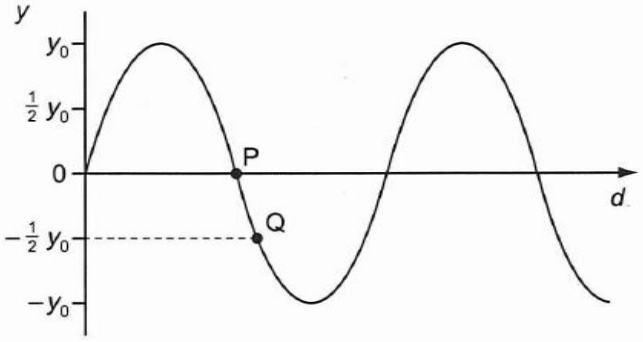
What is the difference in phase at any instant between the waves at two points 5 m apart along the wall?

A 45° **B** 55°

C 90° **D** 180°

D3 N09/I/23

The diagram shows the variation with distance *d* along a sinusoidal wave of displacement *y* of particles in the wave. The amplitude of the wave is y_o.



What is the phase angle between the two particles P and Q in the wave?

A 30° **B** 45° **C** 90° **D** 180°

D4 N13/I/19

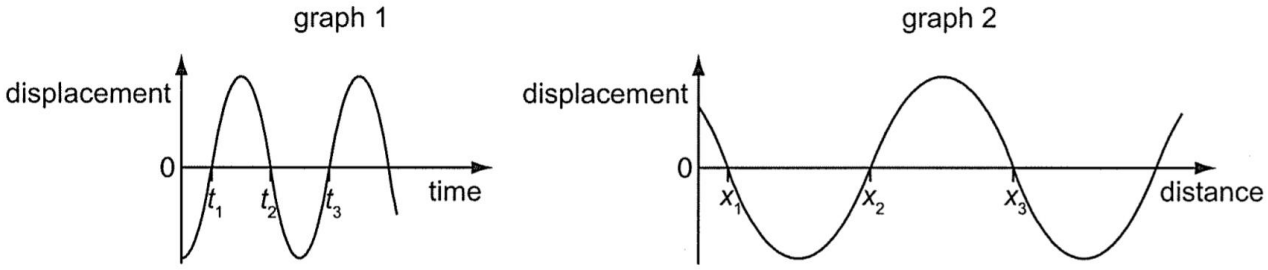
As the intensity of a single frequency sound wave travelling through the air is increased, how do the maximum speed of vibration of the air molecules and the speed of wave travel change?

	Maximum speed of vibration of air molecules	Speed of wave travel
A	increase	increase
B	increase	no change
C	no change	increase
D	no change	no change

D5 N14/I/21

Graph 1 shows how the displacement of one particular point of a wave varies with time.

Graph 2 shows how the displacement of the same wave varies with distance along the wave at one particular time.



Which expression gives the speed of the wave?

A $\frac{x_1}{t_1}$ **B** $\frac{x_2}{t_2 - t_1}$ **C** $\frac{x_2 - x_1}{t_2}$ **D** $\frac{x_3 - x_2}{t_2 - t_1}$

D6(a) The following graphs refer to the same wave. The graph in Fig. D6 (a) shows the displacement vs. position at $t = 0$. The graph in Fig. D6 (b) shows the displacement vs. time graph of the point **Y** labelled on Fig. D6 (a). What direction is the wave progressing?

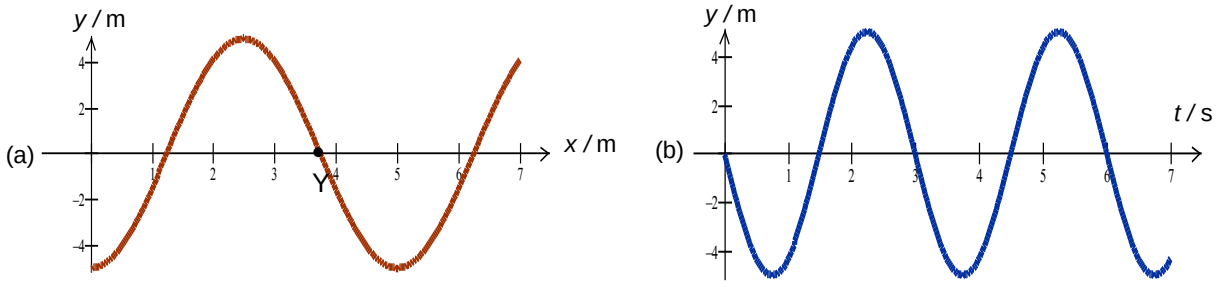


Fig. D6

D6 (b) N21/I/16 Mod

A beam of laser light of frequency 5.0 x 10^{14} Hz is travelling through vacuum.
What is the phase difference between points on the wave that are 1.5 μm apart?

D7 J91/III/2 (part - modified)

Fig. D6 is a full-scale diagram showing the rest position and the actual positions of a series of particles through which a sinusoidal longitudinal wave is passing from left to right.

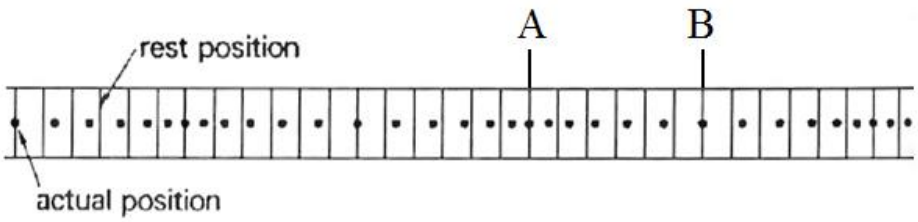


Fig. D6

(a) Describe the movement of a single particle. Describe how the compressions and rarefactions move.

(b) Sketch the displacement vs. position graph for points along A and B for this longitudinal wave at this instant.

(c) Consider the particle at A. Sketch its displacement vs. time graph taking $t = 0$ to be the instant shown. Label clearly on your graph the period T.

D8 Longitudinal Wave - Pressure and Displacement Graph (Hutchings)

The graph shows how particle displacement varies with distance for a sound wave at one instant in time. The speed of the wave is 340 m s^{-1}.

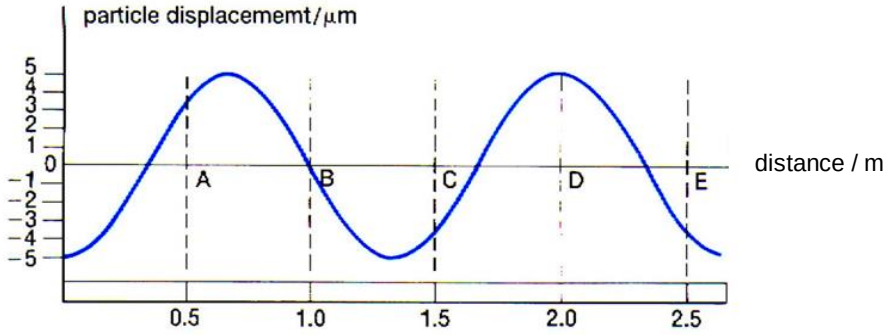


Fig. D8

(a) From the graph, determine

(i) the amplitude,

(ii) the wavelength,

(iii) the frequency,

(iv) the phase difference between the vibration of the particle at A and that of the particle at B.

(b) Which points on the graph correspond to compressions and rarefactions?

Indicate them clearly.

(c) Sketch the corresponding pressure vs. position graph for the sound wave at this instant.

D9 N97/I/10

A plane wave of amplitude *A* is incident on a surface of area *S* placed so that it is perpendicular to the direction of travel of the wave. The energy per unit time intercepted by the surface is *E*.

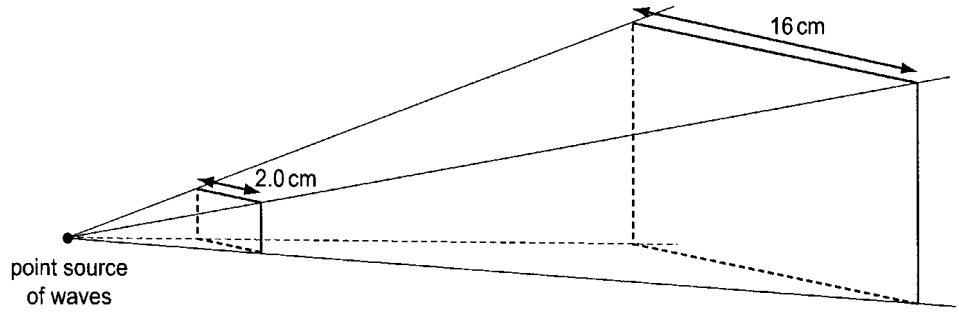
The amplitude of the wave is increased to 2*A* and the area of the surface is reduced to *S/2.*

How much energy per unit time is intercepted by this smaller surface?

A *4E* **B** *2E* **C** *E* **D** *E/2*

D10 N08/I/20

Waves from a point source pass through an area that is 2.0cm wide, as shown.



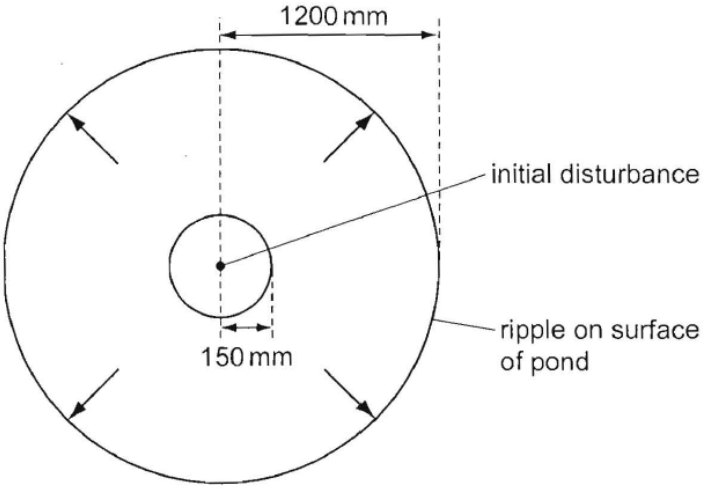
Within this area, the intensity of the waves is *I* and their amplitude is *A*. The waves reach a second area of width 16 cm. What will be the intensity and amplitude of the waves when they reach the second area?

	intensity	amplitude
A	$\frac{I}{8}$	$\frac{A}{4}$
B	$\frac{I}{64}$	$\frac{A}{4}$
C	$\frac{I}{64}$	$\frac{A}{8}$
D	$\frac{I}{256}$	$\frac{A}{16}$

D11 N10/I/20

Ripples on the surface of a pond spread out in circles from the point of an initial disturbance. Assume that the energy of the wave is spread over the entire circumference of the ripple.

For one such ripple, the amplitude of the ripple at a distance of 150 mm from the disturbance is 2.0 mm.



[Please turn over

What will be the amplitude of the ripple at a distance of 1200 mm from the disturbance? (Assume that no energy is lost in the propagation of the ripple.)

A 0.031 mm **B** 0.13 mm **C** 0.25 mm **D** 0.71 mm

D12 N17/I/15

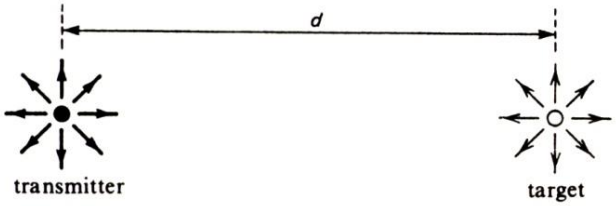
A wave of frequency *f* and wavelength λ has intensity I. The wave travels through a boundary into a new medium where the speed and amplitude are both halved.

What will be the intensity, frequency and wavelength in the new medium?

	intensity	frequency	wavelength
A	$0.25I$	$0.50f$	λ
B	$0.25I$	f	0.50λ
C	$0.50I$	$0.50f$	λ
D	$0.50I$	f	0.50λ

D13 J88/III/8

A radar transmitter produces pulses of microwaves each with a mean power *P* which are emitted uniformly in all directions. A small spherical target of effective area *S* is placed at a distance *d* from the transmitter. The target reflects a small fraction *k* of the energy incident on it uniformly in all directions as shown in the following figure.



(a) Show that the mean intensity I_r if the reflected pulse when it is received back at the transmitter is given by $I_r = \dfrac{PkS}{16\pi^2 d^4}$

(b) Given that the mean power *P* is 2 MW and the pulse duration is 3μs, calculate:

(i) the energy in each emitted pulse,

(ii) the mean intensity of the emitted pulse at a range of 50 km,

(iii) the mean intensity of the reflected pulse when received back at the transmitter if the range of 50 km and the product kS = 1 m^2.

(c) Briefly discuss the effect on your answer to (b)(iii) if the pulses were emitted in an almost parallel beam.

D14 N19/II/2

(a) **(i)** State what is meant by a *transverse* wave. Give an example. [2]

(ii) State what is meant by *plane polarisation* of a wave. [1]

(b) Two polarising filters are held in line with one another and at different angles to an incoming polarised beam of light of intensity I_0, as shown in Fig. 2.1.

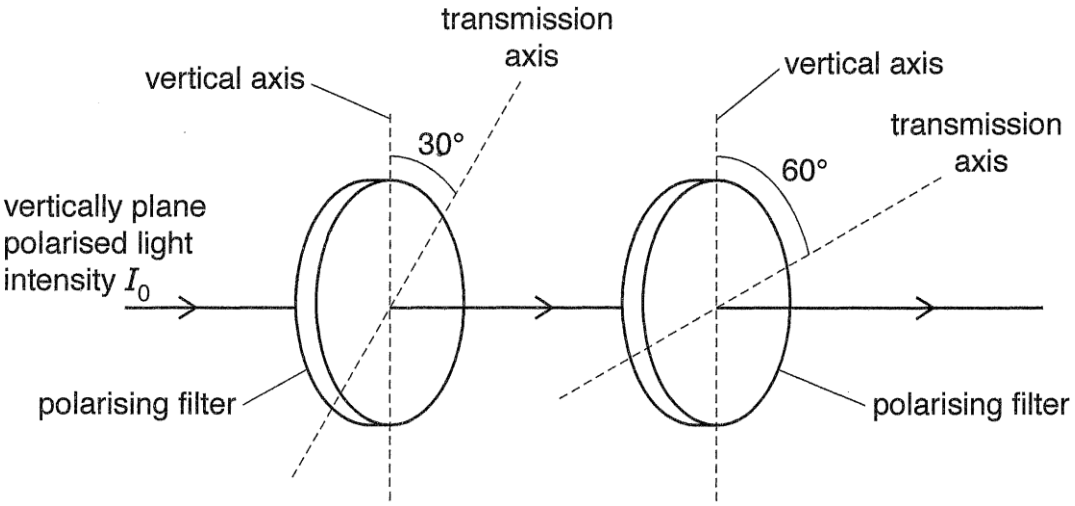


Fig. 2.1

The incoming light is vertically plane polarised. The transmission axis of the first filter is at an angle of 30° to the vertical. The transmission axis of the second filter is at an angle of 60° to the vertical.

(i) Calculate, in terms of I_0, the intensity I_T of the beam emerging from the second filter. [3]

(ii) Use your answer in **(b)(i)** to calculate the ratio

$$\frac{\text{amplitude of beam emerging from second filter}}{\text{amplitude of beam incident on first filter}}$$

[2]

(iii) The second polarising filter is now removed. The first filter is then placed so its transmission axis is parallel to the plane of polarisation of the incoming light. The transmission axis is then rotated through 360°.

Complete Fig. 2.2 to show the variation with the angle of rotation of the intensity of the transmitted light.

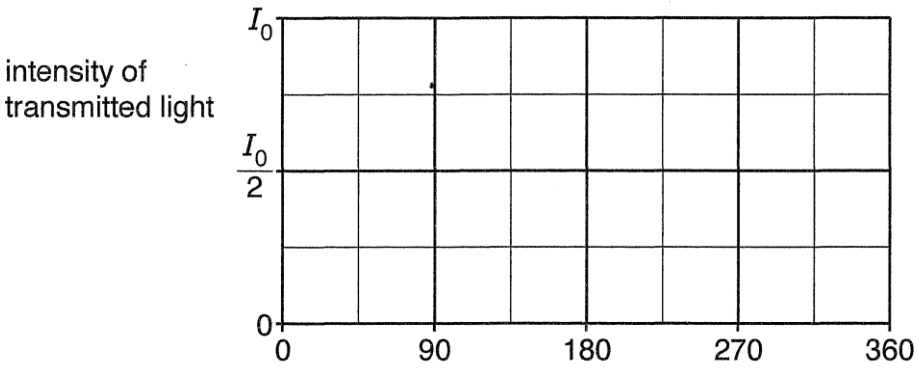


Fig. 2.2 [2]

(c) **(i)** State why a sound wave **cannot** be polarised. [1]

(ii) A microphone and an oscilloscope are used to measure the period of a sound wave of frequency 740Hz.

The distance between two adjacent peaks of the waveform displayed on the oscilloscope screen is 6.8cm.

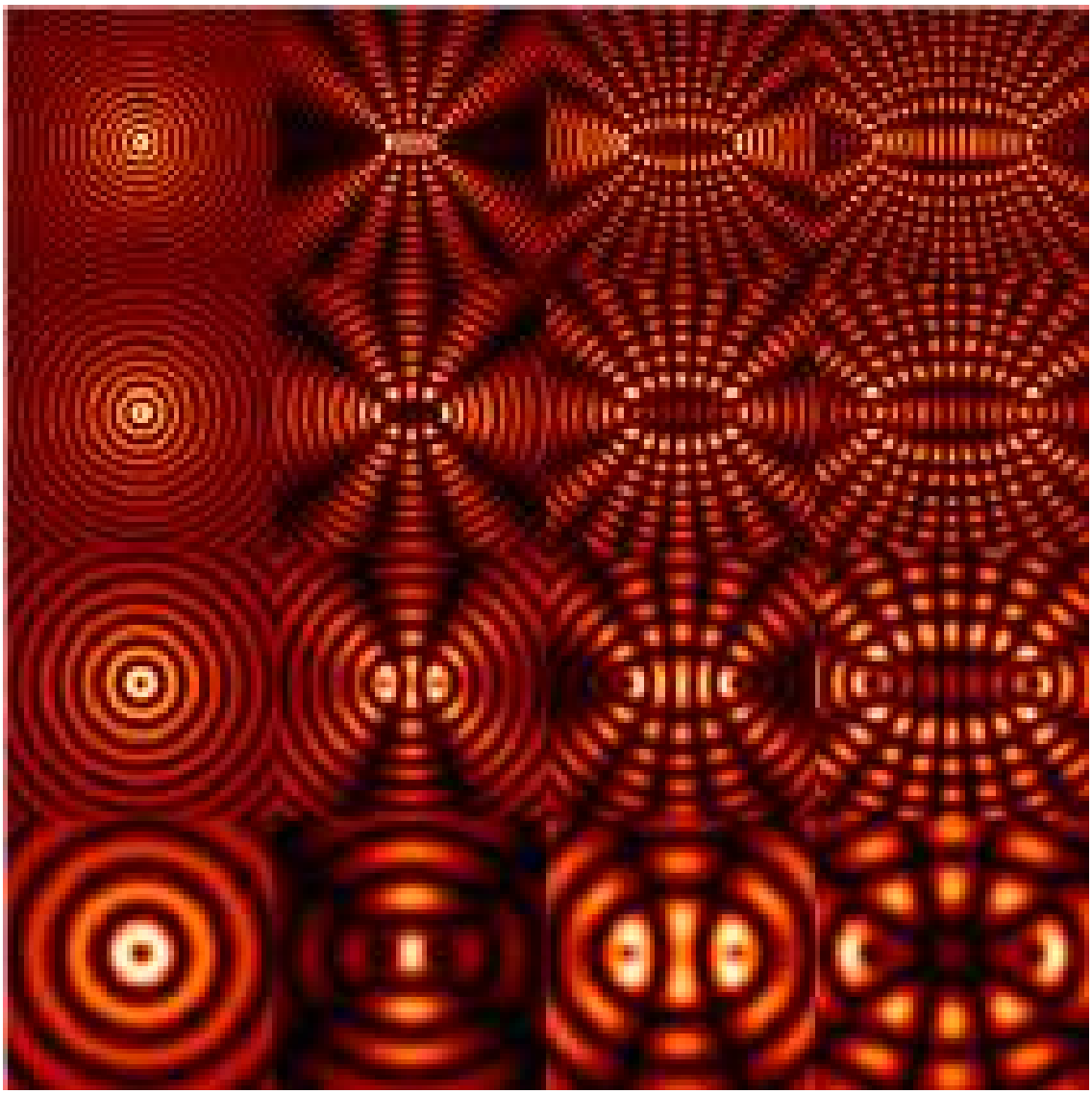
Calculate the time-base setting on the oscilloscope in ms cm^{-1}. [2]

Numerical Answers for Discussion Questions

D6 (b) π rad
D8 (a)(i) 5 μm (ii) 1.33 m (iii) 255 Hz (iv) 2.36 rad
D13 (b)(i) 6 J, (ii) 6.37 x 10^{-5} W m^{-2}, (iii) 2.03 x 10^{-15} W m^{-2}
D14 (b)(i) 0.75 I_o ; 0.563 I_o , (b)(ii) 0.75, (c)(ii) 0.20 ms cm^{-1}

Chapter 10

SUPERPOSITION



Interference pattern produced by two sources placed at different distances apart and of different wavelengths. Taken from www.arthistoryclub.com

Playlist of Lecture Example Solutions and Others available at
https://youtube.com/playlist?list=PL_b5cjrUKDlbeRspxU7s_UswuIC_U1VMJ

Syllabus 9749

Content

- Principle of Superposition
- Stationary Waves
- Diffraction
- Two-source interference
- Single slit and multiple slit diffraction

Learning Outcomes

Candidates should be able to:

(a) explain and use the principle of superposition in simple applications.

(b) show an understanding of the terms: interference, coherence, phase difference and path difference.

(c) show an understanding of experiments which demonstrate stationary waves using microwaves, stretched strings and air columns.

(d) explain the formation of a stationary wave using a graphical method, and identify nodes and antinodes.

(e) explain the meaning of the term diffraction.

(f) show an understanding of experiments which demonstrate diffraction including the diffraction of water waves in a ripple tank with both a wide gap and a narrow gap.

(g) show an understanding of experiments which demonstrate two-source interference using water waves, sound waves, light waves and microwaves.

(h) show an understanding of the conditions required for two-source interference fringes to be observed.

(i) recall and solve problems using the equations $\lambda = a\,x / D$ for double slit interference.

(j) recall and use the equation $\sin\theta = \lambda/b$ to locate the position of the first minima for single slit diffraction.

(k) recall and use the Rayleigh criterion $\theta \approx \lambda/b$ for the resolving power of a single aperture.

(l) recall and use the equation $d \sin\theta = n\lambda$ to locate the positions of the principal maxima produced by a diffraction grating.

(m) describe the use of a diffraction grating to determine the wavelength of light. (the structure and use of a spectrometer are not required.)

Contents

Syllabus 9749 2
Content 2
Learning Outcomes 2

10.1 – Introduction 4

10.2 – Principle of Superposition 4

10.3 – Interference 5

10.4 – Interference of Two Wave Sources 7
10.4.1 – Interference of Waves from Two Sources that are in Phase 7
10.4.2 – Accounting for Phase Difference at the Sources 11

Tutorial 10A: Superposition and Interference 13

10.5 – Interference of Light Waves: Young's Double Slit Experiment 17

10.6 – Diffraction Grating 23

Tutorial 10B: Young's Double Slits and Diffraction Grating 28

10.7 – Diffraction of Waves 36
10.7.1 – Diffraction of water waves 36
10.7.2 – Single Slit Diffraction Pattern 37
10.7.3 – Resolving Power and Rayleigh's Criterion 40

Tutorial 10C: Single Slit and Rayleigh's Criterion 43

10.8 – Stationary Waves 47
10.8.1 – Characteristics of Stationary Wave 48
10.8.2 – Stationary Waves in Strings 50
10.8.3 – Standing Waves in Air Column 53
10.8.4 – What about the pressure? 58
10.8.5 – Investigating the speed of sound using stationary waves 60

Tutorial 10D: Stationary Waves 61

Appendices
Appendix A: Huygen's Principle 66
Appendix B: Single Slit Diffraction Pattern 68
Appendix C: Ripple Tank 69
Appendix D: Spectrometer 70

10.1 Introduction

In your secondary schools, you have seen that all waves exhibit *reflection* and *refraction.*

For the A-levels, we extend our understanding of waves by looking at two other phenomena that are exhibited by waves: *interference* and *diffraction*.

We will learn about *Young's double slit experiment*, *the measurement of wavelengths* of light and sound, and the formation of musical notes in strings and woodwind instruments through the formation of *stationary waves*.

10.2 Principle of Superposition

Imagine throwing two pebbles into a still pond. Each pebble makes its own ripples and as the ripples spread out, they kind of overlap. What determines the resultant "ripple" that we see?

The **displacement** of the resultant wave is determined by the **Principle of Superposition**:

Principle of Superposition:

The principle of superposition states that **when two or more waves of the same kind overlap**, the **resultant displacement** at any point at any instant is given by the **vector sum of the individual displacements** that each individual wave would cause at that point at that instant.

$$y_{resultant} = y_1 + y_2 + y_3 + ...+ y_N$$

where $y_{resultant}$ is the displacement of the resultant wave and $y_1, y_2, ..., y_N$ are the individual displacements of the waves at the point of consideration.

The statement of the Principle of Superposition is a popular exam question. Be careful to use the word "**displacement**" and not "amplitude" in your statement.

Fig. 10.2.1 and 10.2.2 are pictorial representations of the superposition of two wave pulses, y_1 and y_2, travelling in opposite directions. When the waves begin to overlap, the resultant displacement of the waveform on the string at every point at any time is given by $y_1 + y_2$.

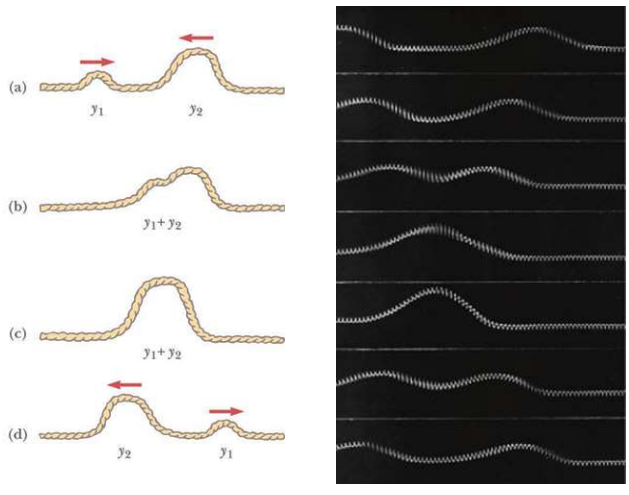


Fig. 10.2.1. Constructive interference of two wave pulses.

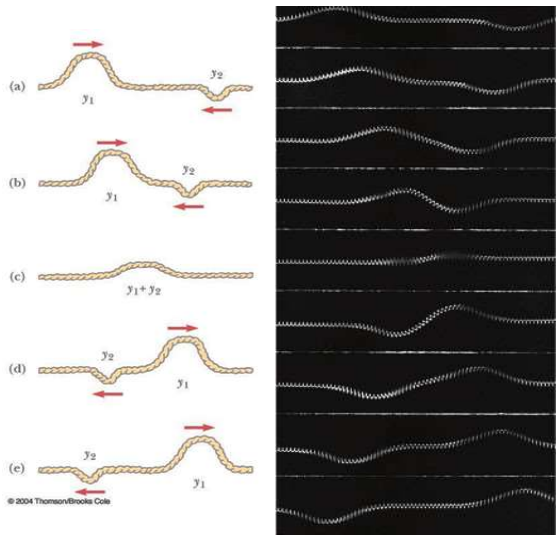


Fig. 10.2.2. Destructive interference of two wave pulses.

Two pulses travelling on a stretched string in opposite directions pass through each other. When the pulses overlap the net displacement of the string is equal to the sum of the displacements produced by each pulse.

Note that when the two pulses separate, they will continue to move in their original directions as if nothing has happened. From this illustration, we also see that:

After two travelling waves have passed through each other, **the pulse shapes of each wave are unchanged, as if the two pulses had never met.**

10.3 Interference

This combination of separate waves in the same region of space at the same time to produce a resultant wave is called **interference**:

Interference

Interference is the **superposing** or overlapping of two or more waves to give a resultant wave whose **displacement** is given by the **Principle of Superposition,** which states that **displacement** of the resultant wave at any point is the **vector sum** of the **displacements** of the individual waves at that point.

Constructive interference

Constructive interference **occurs at a point** when **two waves** meet **in phase** at that point. The resultant **amplitude** of the oscillation at that point is therefore a **maximum**.

- Phase difference between the two waves at that point = 0, 2π, 4π, …
- Resultant amplitude $A_R = A_1 + A_2$

Constructive Interference of Transverse Waves

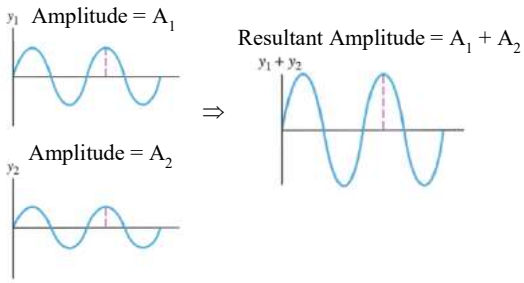


Fig. 10.3.1. Two transverse waves in phase undergo constructive interference at every point to produce a wave of maximum amplitude.

Constructive Interference of Longitudinal Waves

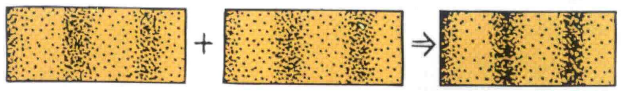


Fig. 10.3.2. Two longitudinal waves in phase undergo constructive interference at every point to produce a wave of increased intensity. In this case, since the amplitudes of the 2 longitudinal waves are the same, the resultant amplitude will be double that of each individual wave.

Destructive Interference:

Destructive interference **occurs at a point** when **two waves** meet in **antiphase** (phase difference of 180°). The resultant **amplitude** of the oscillation at that point is therefore a **minimum**.

- Phase difference between the two waves at that point = π, 3π, 5π,…
- Resultant amplitude $A_R = A_1 - A_2$
- If the waves have the **same amplitude**, then the **resultant amplitude** would be **zero**.

Destructive Interference of Transverse Waves

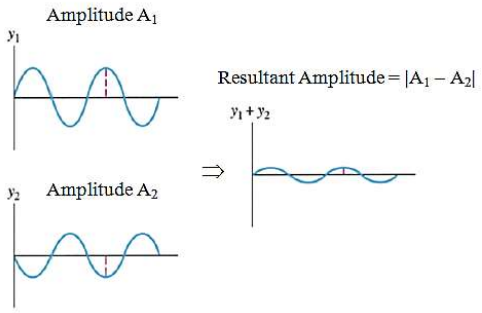


Fig. 10.3.3 Two identical transverse waves that are out of phase undergo destructive interference at every point when they superimposed

Destructive Interference of Longitudinal Waves

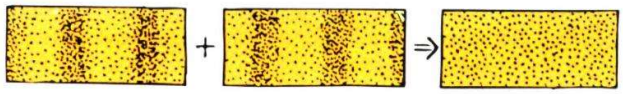
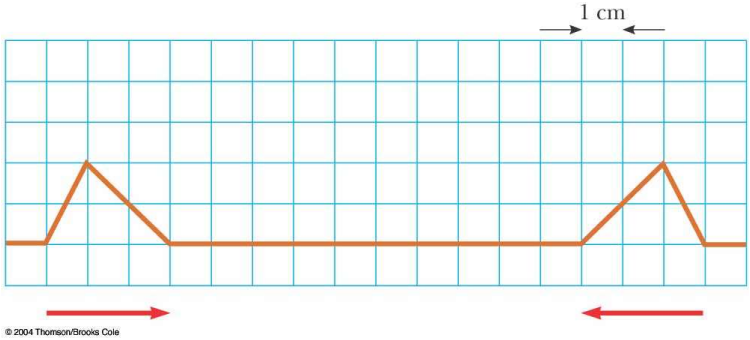


Fig. 10.3.4. Two identical longitudinal waves that are out of phase undergo destructive interference at every point when they are superimposed

Note that when interference is neither fully constructive nor fully destructive, it is called *intermediate interference* or *partially constructive or destructive interference*. The amplitude of the resultant waveform will be between the maximum and minimum amplitudes possible (not inclusive).

Example 10.3.1

Two pulses are travelling toward each other, at 10 cm s^{-1} on a long string, as shown in the figure below. Sketch the shape of the string after 0.6 s.



Solution:

10.4 Interference of Two Wave Sources

10.4.1 Interference of Waves from Two Sources that are in Phase

Consider two identical sources of monochromatic waves S_1 and S_2 as shown in Fig. 10.4.1. The two sources produce waves of *same amplitude, same frequency* and are permanently oscillating *in phase*. Both waves have the same speed.

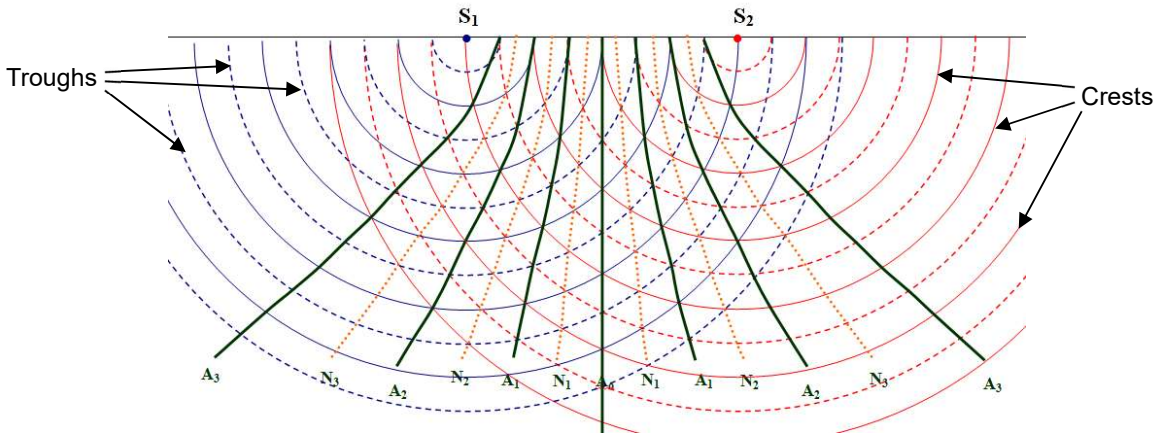


Fig. 10.4.1. Interference pattern of 2 waves sources oscillating in phase. Solid lines represent the crests while dotted lines represent the troughs of the waves produced by the two sources

Path Difference between the Two Wave Sources

Before we derive the general conditions for *constructive interference* or *destructive interference* to be observed at any point **P**, we first define what we mean by *path difference*:

Path difference δ between the two waves,

$$\delta = \left|\text{Distance travelled by wave from } S_1 \text{ to P} - \text{ Distance travelled by wave from } S_2 \text{ to P}\right|$$

$$= \left|S_1P - S_2P\right| \qquad \mathbf{A_o}$$

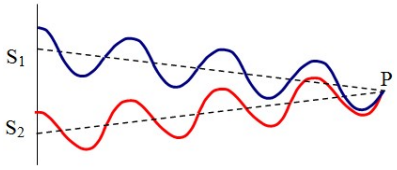


Fig. 10.4.2. Path difference between waves from S_1 and S_2 is the difference between the distances travelled by wave from S_1 and wave from S_2.

Conditions for Constructive Interference

Consider the lines **A_o, A_1, A_2** and **A_3.**

Along these lines, the crests of the wave from S_1 (See Fig. 10.4.1) will always meet with crests from S_2. Similarly the troughs of the wave from S_1 will meet with the troughs from S_2.

From the Principle of Superposition, at the points where a crest meets another crest, we will get a resultant crest that is twice as high of that of the component waves. Similarly, at the points where a trough meets another trough, the resultant wave will have a trough that is twice as low as that of each component wave.

Therefore along the lines **A_o, A_1, A_2** and **A_3**, **the waves will always meet in phase** and hence **constructive interference** occurs. **A_o, A_1, A_2** and **A_3** are also known as the **antinodal lines**.

For a point P on the antinodal line A_o, two waves from S_1 and S_2 travel to P over the same distance, therefore they meet in phase. We see that for all points on line A_o

Path difference between the 2 waves at P: $\delta = |S_1P - S_2P| = 0$

Phase difference between the 2 waves at P: $\Delta\phi = 0$

For any point P on the antinodal line A_1, we see that the path (distance) travelled by one wave to P is always longer than the other wave by **one wavelength**.

Path difference between the 2 waves, $\delta = |S_1P - S_2P| = \lambda$

Phase difference between the 2 waves at P: $\Delta\phi = 2\pi$

Similarly for antinodal lines A_2, for all points P on the line, the path travelled by one wave is always longer than the other wave by **two wavelengths.**

Path difference between the 2 waves, $\delta = |S_1P - S_2P| = 2\lambda$

Phase difference between the 2 waves at P: $\Delta\phi = 4\pi$

In general,

Conditions for Constructive Interference:

1) **Looking at Phase difference (independent of phase difference at the sources):**

   Phase difference between the two waves **meeting at point P** is

   $$\Delta\phi = 0, 2\pi, 4\pi, 6\pi, ... = 2m\pi$$

2) **Looking at Path difference (dependent on phase difference at the sources):**

   For **two waves sources S_1 and S_2** that are oscillating **in phase**,

   Path difference between the two waves travelling from the sources to the point P is:

   $$\delta = |S_1P - S_2P| = 0, \lambda, 2\lambda, 3\lambda, ... = m\lambda$$

   where m = 0,1,2,3,... and λ is the wavelength of the individual waves.

Note:

1. The points of **constructive interference** are also known as points of **maxima**.
2. m is also commonly known as the *order*.

- The central maximum is known as the *zeroth order maximum.* (m = 0)
- The two maxima on either side of the zeroth order maximum are known as the **1st order maxima** (m = 1). The next maxima further away are then known as the **2nd order maxima** and so on.
- Note that sometimes, we assign signs (+ or -) to distinguish the "directions" of the order e.g. the orders observed to the right of the zeroth order may be given a positive sign (+) and therefore those orders observed to left of the zeroth order may be given a negative (-) sign.

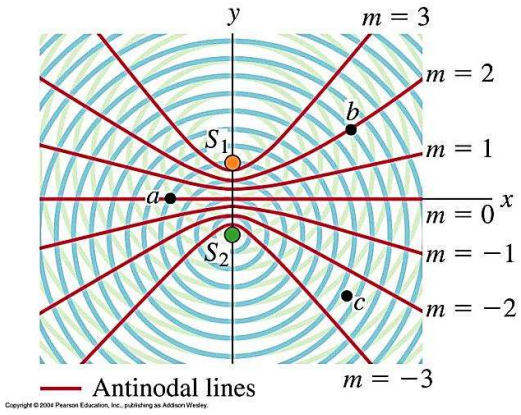


Fig. 10.4.3. Two wave sources interfere with each other to produce points of constructive and destructive interference in a two dimension area. Point *b* is a point where constructive interference is observed. Point *c* is a point where destructive interference is observed. The antinodal lines (lines that join points of constructive interference of the same order) are indicated as well with the value of the order *m* shown. Here orders which lie above the zeroth order maximum are given a (+) sign and orders which lie below the zeroth order maximum are given a (-) sign.

Destructive Interference

Consider each point along lines **N_1, N_2**, and **N_3** (See Fig 10.4.1). We notice that, along these lines, the crest of a wave from S_1 will always meet with a trough from S_2.

Hence, the two waves will always meet in **antiphase** and therefore **destructive interference occurs. N_o, N_1, N_2** and **N_3** are also known as **nodal lines.**

We follow the same procedure to derive the general condition for a destructive interference to be observed at any point **P** from the two sources by examining the nodal lines.

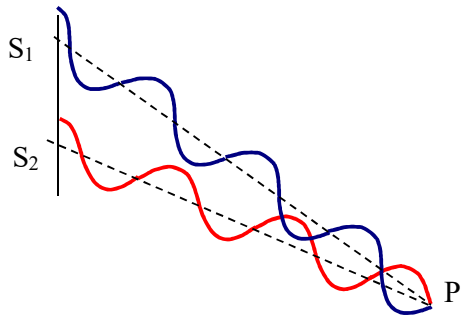


Fig. 10.4.4

For a particular nodal line N_1, the wave from S_1 travels one-half of a wavelength ($\lambda/2$) more than the wave from S_2 to P, so if initially the waves were produced at the source to be in phase (phase difference is zero), this additional $\lambda/2$ will cause an equivalent phase difference of 180° and hence the two waves will meet in anti-phase. Therefore a destructive interference will be observed at P.

Now if P is a point on N_1,

Path difference between the 2 waves,

$$\delta = \left|S_1P - S_2P\right| = \frac{\lambda}{2}$$

If P is a point on N_2,

Path difference between the 2 waves,

$$\delta = \left|S_1P - S_2P\right| = \frac{3\lambda}{2}$$

Finally, if P is a point on N_3,

Path difference between the 2 waves,

$$\delta = \left|S_1P - S_2P\right| = \frac{5\lambda}{2}$$

Conditions for Destructive Interference:

1) Looking at Phase difference (independent of phase difference at the sources):

Phase difference between the two waves meeting at point P is given by:

$$\Delta\phi = \pi,\ 3\pi,\ 5\pi,\ 7\pi, ... = (m - \frac{1}{2})\ 2\pi$$

2) Looking at Path difference (dependent on phase difference at the sources):

For **two waves sources S_1 and S_2** that are oscillating **in phase**,

Path difference between the two waves travelling from the sources to the point P is:

$$\delta \ = \ \left|S_1P - S_2P\right| \ = \ \frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, ... \ = (m - \frac{1}{2})\lambda$$

where the order, m = 1,2,3,... and λ is the wavelength of the individual waves.

Note :

1. The points of **destructive interference** are also known as points of **minima**.
2. Notice that because there is no destructive interference in the central line, there is **no zeroth order minimum observed**.

10.4.2 Accounting for Phase Difference at the Sources

Question:

The conditions for constructive and destructive interference stated above are for the case that ***the sources started out in phase***. If the **original sources *are 180° out of phase*** (i.e. in anti-phase or equivalent to an additional path difference of $\frac{\lambda}{2}$), will the **conditions for path difference** change and why?

Answer:

(The phase difference between the 2 wave trains arriving at P)

= (Phase difference at the source) + (Phase difference due to the equivalent path difference)

All the points where the waves meet in phase previously, the waves will now meet in antiphase.

All the points where the waves meet in antiphase previously, the waves will now meet in phase.

Hence, the conditions for constructive and destructive interference **looking at path difference** are interchanged

For Constructive Interference :	For Destructive Interference :
For **two waves sources S_1 and S_2** that are oscillating **in anti-phase**, **Path difference** between the 2 waves trains arriving at P, $\delta = \lvert S_1P - S_2P \rvert = \frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, \ldots$ $= (m\text{-}1/2)\lambda$ where m = 1, 2, 3, 4,.... *m*: *m*th order maxima **Note: Phase difference** between the two waves **meeting at point P** is still $\Delta\phi = 2\pi, 4\pi, 6\pi, \ldots = 2m\pi$	For **two waves sources S_1 and S_2** that are oscillating **in anti-phase**, **Path difference** between the 2 wave trains arriving at P, $\delta = \lvert S_1P - S_2P \rvert$ = 0, λ, 2λ, 3λ, … $= m\lambda$ where m = 0, 1, 2, 3, 4,.... *m*: *m*th order minima **Note: Phase difference** between the two waves **meeting at point P** is still given by: $\Delta\phi = \pi, 3\pi, 5\pi, 7\pi, \ldots = (m - \frac{1}{2})\, 2\pi$

Note : In this case, the zeroth order is a minimum and there is no zeroth order maximum.

Example 10.4.1: Two Source Interference

The figure (Fig. 10.4.5) shows two sources X and Y which emit identical sound waves of wavelength 2.0 m. The two sources emit in phase, and the waves emitted have equal amplitudes, each *A*.

What is the amplitude of the sound wave

(a) at R,

(b) at Q?

Suppose the source X is 180° out of phase with source Y. What does an observer hear

(c) at R,

(d) at Q?

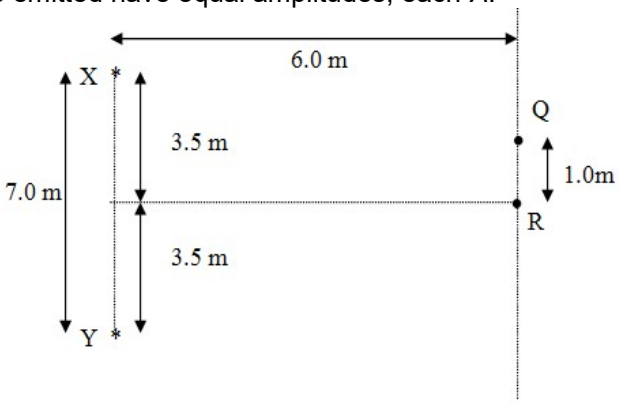


Fig. 10.4.5

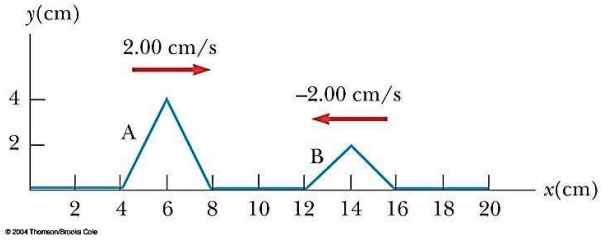
Solutions:

➢ **You may now complete Tutorial 10A on Superposition and Interference.**

Tutorial 10A: Superposition & Interference

Self-Practice Questions:

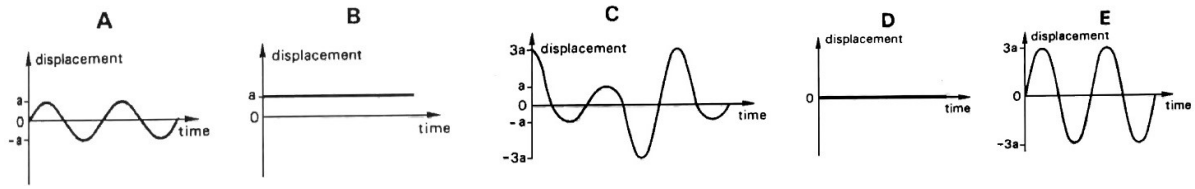
S1 (Serway 6th Edition. P18.2) Two pulses A and B are moving in opposite directions along a taut string with a speed of 2.00 cm s^{-1}. The amplitude of A is twice the amplitude of B. The pulses at t = 0 are shown in the figure on the right. Sketch the shape of the string at t = 1.0, 1.5, 2.0, 2.5 and 3.0 s.



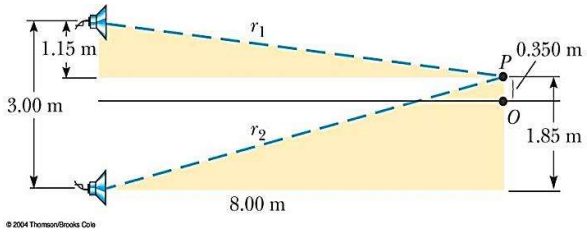
S2 N77/II/9; J89/I/10; J83/II/10.

Wave generators S_1 and S_2 generate waves of equal wavelength. At a point P, S_1 by itself produces an oscillation of amplitude 2*a*, and S_2 produces an oscillation of amplitude *a*, and there is a phase difference of π between the oscillations.

Which graph best represents the resultant oscillation at P when both the generators are switched on?



S3 Serway Example 18.1. A pair of speakers placed 3.00 m apart are driven by the same oscillator. A listener is originally at point O, which is located 8.00 m from the centre of the line connecting the two speakers. The listener than walks to point P, which is a perpendicular distance 0.350 m from O, before reaching the *first minimum* in sound intensity. (Assume speed of sound in air is 330 m s^{-1}.)



(a) What is the frequency of the oscillator?

(b) What if the speakers were connected in antiphase? What happens at point P?

Discussion Questions:

D1 **Serway 6th Edition.** Two pulses move in opposite directions on a string and are identical in shape except that one has positive displacements of the elements of the string and the other has negative displacements. At the moment that the two pulses completely overlap on the string,

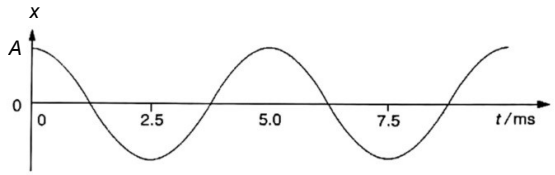
A the energy associated with the pulses has disappeared

B the string is not moving

C the string forms a straight line

D the pulses have vanished and will not appear

D2 **J95/II/2.** A particle in a medium is oscillating because of a transverse wave T_1 of intensity I. The figure below shows the variation with time t of the displacement of x of the particle. The amplitude of oscillation is A.



A second, similar transverse wave T_2 has the same frequency, but the amplitude of the oscillation due to T_2 alone is $3A/2$.

(a) Calculate

(i) the frequency of the waves,

(ii) the intensity, in terms of I, of the wave T_2.

(b) State two conditions which are necessary for the waves T_1 and T_2 to interfere.

(c) **(i)** What additional condition must be satisfied when the resultant intensity is to be

1. a minimum, and

2. a maximum.

(ii) Calculate, in terms of I, this minimum and maximum intensities.

D3 **J2000/II/4** Two microwave sources A and B are in phase with one another. They emit waves of equal amplitude and of wavelength 30.0 mm. They are placed 140 mm apart and at a distance 810 mm from a line OP along which the detector is moved.

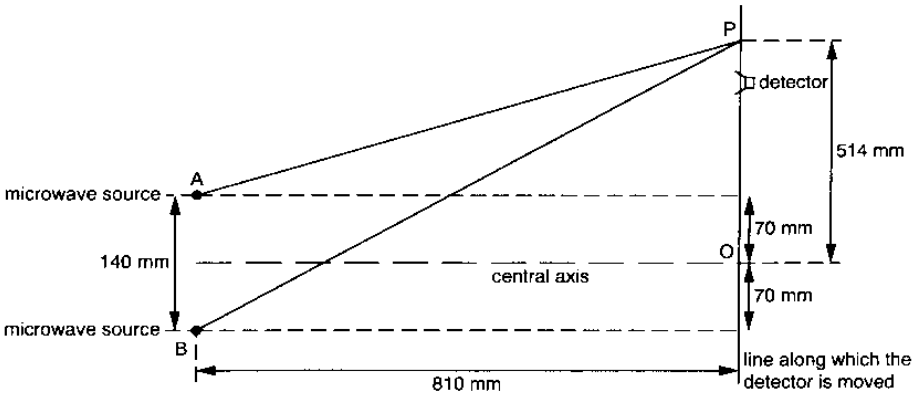


Fig. 4.5 (not to scale)

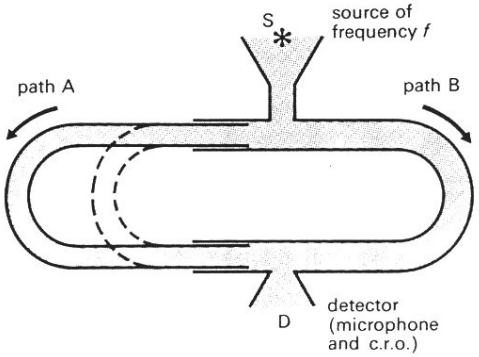
(i) Using Pythagoras' Theorem, it can be shown that the distance AP is 923.7 mm. Calculate the number of wavelengths between source A and point P.

(ii) Show that there are 33.3 wavelengths between source B and point P.

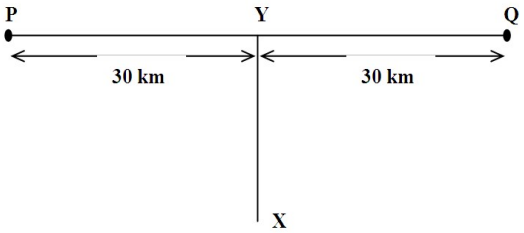
(iii) 1. State what intensity of microwaves will be received by the detector at P.

2. Describe how the intensity of reception varies as the detector is moved from P to point O on the central axis.

D4 **The Quincke tube.** Refer to the diagram below. In this method for measuring the speed of sound in air, a source of frequency 1.7 k Hz was used. The sliding tube had to be moved 0.10 m between consecutive positions of minimum intensity. Calculate the speed of sound.



D5 A ship at X is equidistant from two shore-based radio transmitters P and Q as shown in the figure below. Both transmitters operate on a wavelength of 300 m and radiate signals of equal amplitude.



(a) In the figure, the ship at X detects zero signal amplitude. What information does this give about the signals from P and Q?

(b) The ship moves in a straight line from X to Y. Throughout the journey the amplitude of the signal detected by the ship is zero. Explain this.

(c) The ship moves in the direction YQ until the signal detected has amplitude twice that from either transmitter alone. How far has the ship moved?

(d) When the ship sails from Y to the harbour at Q, the detected signal rises and falls in amplitude, calculate how many dips in intensity will be passed.

Numerical Answers to Select Questions in Tutorial 10A: Superposition & Interference

S3 1280 Hz

D2 (a)(i) 200 Hz (ii) $\frac{9}{4}I$ (c) (ii) $\frac{1}{4}I, \frac{25}{4}I$

D3 (i) 30.8 (ii) 33.29

D4 340 m s^{-1}

D5 (c) 75 m (d) 200

10.5 Interference of Light Waves - Young's Double Slit Experiment

The interference pattern produced by two *water waves* of the same wavelength can be readily seen in a ripple tank[1]. However, interference pattern is generally not easily observable when two light waves overlap.

So can light waves interfere with each other?

When two independent light sources are placed side by side, no interference pattern is observed because the light waves emitted by each of the sources are not **coherent** (i.e. they do not maintain a constant phase relationship with each other over the time of observation).

Light from an ordinary source undergoes random phase change about once every 10^{-8} s. Therefore, conditions for constructive interference, destructive interference, or some intermediate state last for times of the order of 10^{-8} s.

Since the eye cannot follow such short-term changes, no interference pattern is observed.

Conditions for two waves to produce an Interference Pattern

1. The sources must be **coherent**; i.e. they must maintain a *constant phase difference* with respect to each other.
2. The two wave sources must also emit waves of roughly the same amplitude.
3. For transverse waves, they must be unpolarised, or share a common direction of polarisation.

Note:

1 Coherence:

Two waves are said to be **coherent** if they have a **constant phase difference** between them.

- When two sources are coherent, it does not necessarily imply that they are in phase with each other.
- The constant phase difference may be non-zero.
- Therefore, if 2 sources are coherent, this necessarily implies that the frequencies and wavelengths (and hence speeds) of the waves emitted must be the same.

[1] Read more about the Ripple Tank in Appendix C (Pg. 48).

10.5.1 Description on Young's Double Slit Experiment

It was Thomas Young who first demonstrated that light can also interfere with each other to produce an interference pattern as predicted by wave theory and hence demonstrating that light is a wave. A schematic diagram of the apparatus that Young used is shown in Fig. 10.5.1.

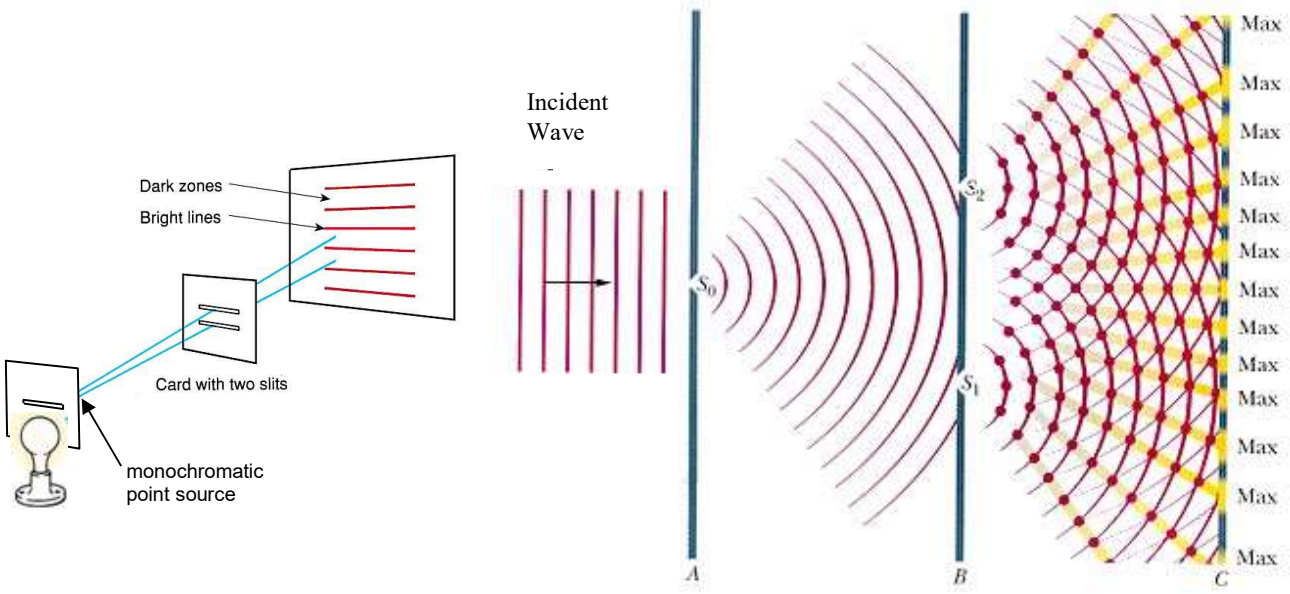


Fig. 10.5.1 Setup for Young's Double Slit Experiment

- To produce light sources which are coherent in Young's interference experiment, a point source of monochromatic light is incident on 2 slits. As light reaches the two slits, it is diffracted[2] by slits S_1 and S_2, which then act as 2 point sources. Since there is only one primary point source, the light waves emerging from S_1 and S_2 are always coherent. If S_1 and S_2 are equidistant from S_0, then the 2 sources are in phase.

- The light waves travelling from S_1 and S_2 overlap and undergo interference, forming an interference pattern of **maxima** and **minima**.

- On the screen, points of visible bright rows called *bright bands, bright fringes,* or *maxima* are formed. The dark regions are called *dark bands, dark fringes,* or *minima.*

- The pattern of bright and dark fringes on the screen is called an **interference pattern**. (Fig.10.5.2)

- A point source is obtained usually by placing another slit S_o between the double-slit and an extended light source.

[2] Diffraction will be discussed in more detail in Section 10.7

Conditions for Maxima and Minima

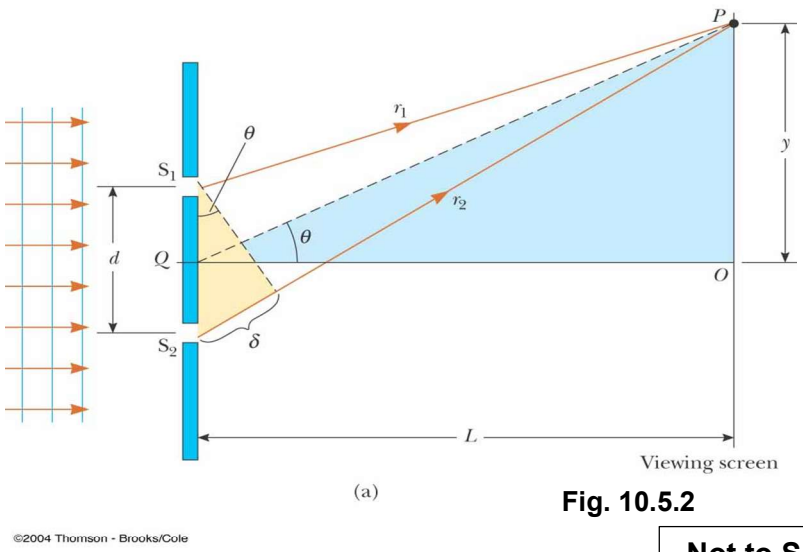
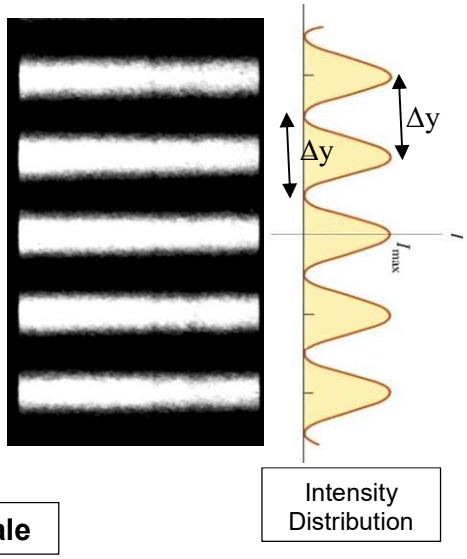


Photo of interference pattern



Intensity Distribution

Fig. 10.5.2

Not to Scale

Let us now derive the conditions for the positions of the maxima and minima of the Young's Double Slit experiment.

In Fig. 10.5.2, a screen is placed at a perpendicular distance *L* from the double-slits. The double slits are separated by a ***slit separation* of *d***. Because the wave through S_1 originate from the same wavefront as the wave through S_2, the two waves at S_1 and S_2 are always **in phase**.

Consider an arbitrary point **P** on the viewing screen,

- For **a maximum to be observed at P**, the wave from S_1 needs to **constructively interfere** with the wave from S_2. Hence, the path difference δ between the two waves when they meet at P needs to be:

$$\delta = |r_1 - r_2| = 0, \lambda, 2\lambda, 3\lambda, ... = m\lambda \qquad \text{where } m \text{ : order } (m = 0,1,2,3,...)$$

- Similarly, for **a minimum to be observed at P**, the wave from S_1 needs to **destructively interfere** with the wave from S_2. Hence, the path difference δ between the two waves when they meet at P needs to be:

$$\delta = |r_1 - r_2| = \frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, \frac{7\lambda}{2} = (m - \frac{1}{2})\lambda \qquad \text{where } m \text{ : order } (m = 1,2,3,...)$$

However, for the experiment, the **separation between the slits *d* is often negligible compared to the distance *L* between the screen and the slits**, i.e. *L* >> *d*. Therefore the rays from S_1 and S_2 to *P* are almost parallel to each other. (Fig. 10.5.3)

Point to Note: *What are the typical orders of magnitude for slit separation, d, slit widths S_1 and S_2 and distance L for Young's double slits experiment with light?*

Approximating the rays to be parallel to each other, the ***path difference*** δ may thus be *approximated* by:

$$\delta = d \sin\theta$$

where *d* is ***the separation between the slits***, and θ is the ***angular displacement of the position of interference fringe observed from the principal axis QO***.[3]

The conditions that define whether *bright* or *dark* fringes are observed at point P is thus given by

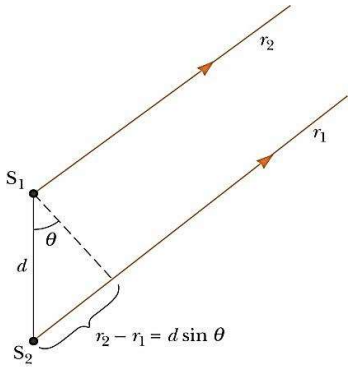


Fig. 10.5.3

1 Conditions for Constructive or Destructive Interference

Condition for Constructive Interference (Maxima or Bright Fringes):

$d \sin\theta = m\lambda$ where *m* : order number ($m = 0, 1, 2, 3, ...$)

Condition for Destructive Interference (Minima or Dark Fringes):

$$d \sin\theta = (m - \frac{1}{2})\lambda$$ where *m* : order number ($m = 1, 2, 3, ...$)

The assumption made during the derivation of these two equations is *L >> d*

Note:

- Similar to the two source interference introduced earlier, *m* is called the **order number.**
- *m* = 0 corresponds to θ = 0. In this case, a central bright fringe is observed and hence it is known as the *zeroth order maximum.* The first maximum on either side is called the *first-order maximum,* and so forth.
- Sometimes, a sign is assigned to the order to help distinguish the positions of the maxima or minima. For example, in this case, maxima located at one side (suppose above) of the zeroth order maximum may have values of *m* = 1, 2, 3, 4.... while maxima located at the other side (below) the zeroth order maximum may have values of *m* = -1, -2, -3, -4, ...

2 Positions *y* of Bright and Dark Fringes

It is often useful to obtain expressions for the positions *y* along the screen of the fringes measured vertically from O to P.

In a typical Young's Double Slit experiment, constructive interference occurs when

$$d \sin\theta = m\lambda \Rightarrow \sin\theta = m\lambda/d$$

If λ << *d*, the value of θ would be very small for low order fringes (i.e. small *m*), hence we can use small angle approximation ($\sin\theta \approx \tan\theta$) to arrive at $\tan\theta = m\lambda/d$

From triangle OPQ (Fig. 10.5.2), we have $\tan\theta = \frac{y}{L} \Rightarrow \frac{m\lambda}{d} = \frac{y}{L}$

[3] The *principal axis* is a line joining the midpoint of the two slits perpendicularly to the viewing screen.

Hence,

Positions of Bright Fringes: $y_{bright} = m\frac{\lambda L}{d}$ (m = 0, ±1, ±2, ±3, ...)

Positions of Dark Fringes: $y_{dark} = (m-\frac{1}{2})\frac{\lambda L}{d}$ (m = ±1, ±2, ±3, ...)

The assumptions made during derivation of the above 2 equations are:

- θ **is small. (Applies for low values of** m **where** $m\lambda << d$**)**
- $L >> d$

3 Fringe width or Fringe Separation, Δy:

The ***fringe width* or *fringe separation*** is defined to be the *distance between 2 consecutive bright fringes* or the *distance between 2 consecutive dark fringes*.

Hence, **Fringe separation**

Δy = Position of $(m+1)^{th}$ order bright fringe – Position of m^{th} order bright fringe

= $y_{m+1} - y_m$

= $(m+1)\frac{\lambda L}{d} - m\frac{\lambda L}{d}$

i.e. $\boxed{\Delta y = \frac{\lambda L}{d}}$

where λ is wavelength of light;

L is the distance of screen from slits;

d is the slit separation.

Notice that on the right hand side of the equation, λ, L and d are constants, hence we can make the following conclusion:

For ***a two-source interference, where d << L and λ << d,*** *the fringe separations close to the central maximum* **(**θ **is small)** *is approximately* ***constant****.*

This conclusion is further confirmed by the photo in Fig. 10.5.2, which captures the fringe pattern formed on the viewing screen during an experiment. The corresponding plot of the **intensity distribution** is also given.

Example 10.5.1 : Determination of wavelength of light using Young's Double Slit

A screen is separated from the double-slit source by 1.2 m. The distance between the two slits is 0.03 cm. The second-order bright fringe (m = 2) is measured to be 4.5 mm from the centre line

(i) What is the fringe separation between 2 neighbouring bright fringes formed on the screen?
(ii) Determine the wavelength of light. Ans: (i) 2.25 mm (ii) 5.63 x 10^{-7} m

Solution:

Effects of Diffraction on Double Slit

The slits used in the double slit experiment in practice cannot be considered as point sources. As a result, each slit produces a diffraction pattern as shown in Fig. 10.7.5. Fig. 10.5.4 illustrates the interference pattern of two slits superposed on the intensity pattern of a single slit. The interference fringe pattern due to the double slit gets moderated by the diffraction pattern.

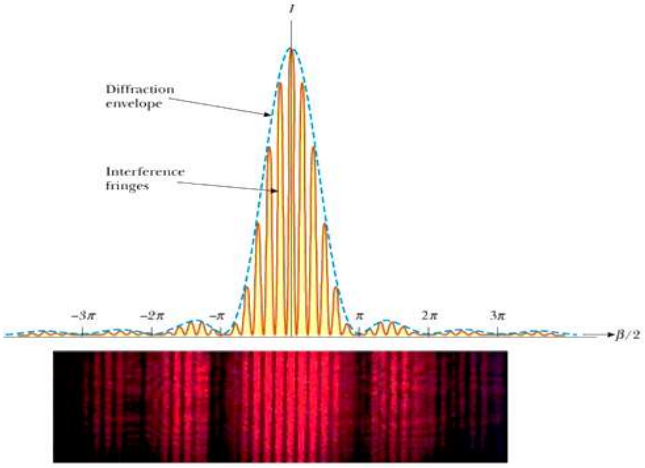


Fig. 10.5.4

This can sometimes result in missing orders in the interference fringe pattern. When the minima of the diffraction envelope coincide with the maxima of the interference pattern, the "maxima" of the interference pattern go missing because they are too dim to be seen.

10.6 Diffraction Grating

We have seen that the Young's double-slit experiment is able to allow us to determine the wavelength of light. However, a more commonly used tool is the ***diffraction grating.*** A diffraction grating has many fine, parallel, closely spaced and equidistant slits/rulings inscribed on a sheet of glass or metal. A typical diffraction grating has several hundreds to thousands of slits per mm.

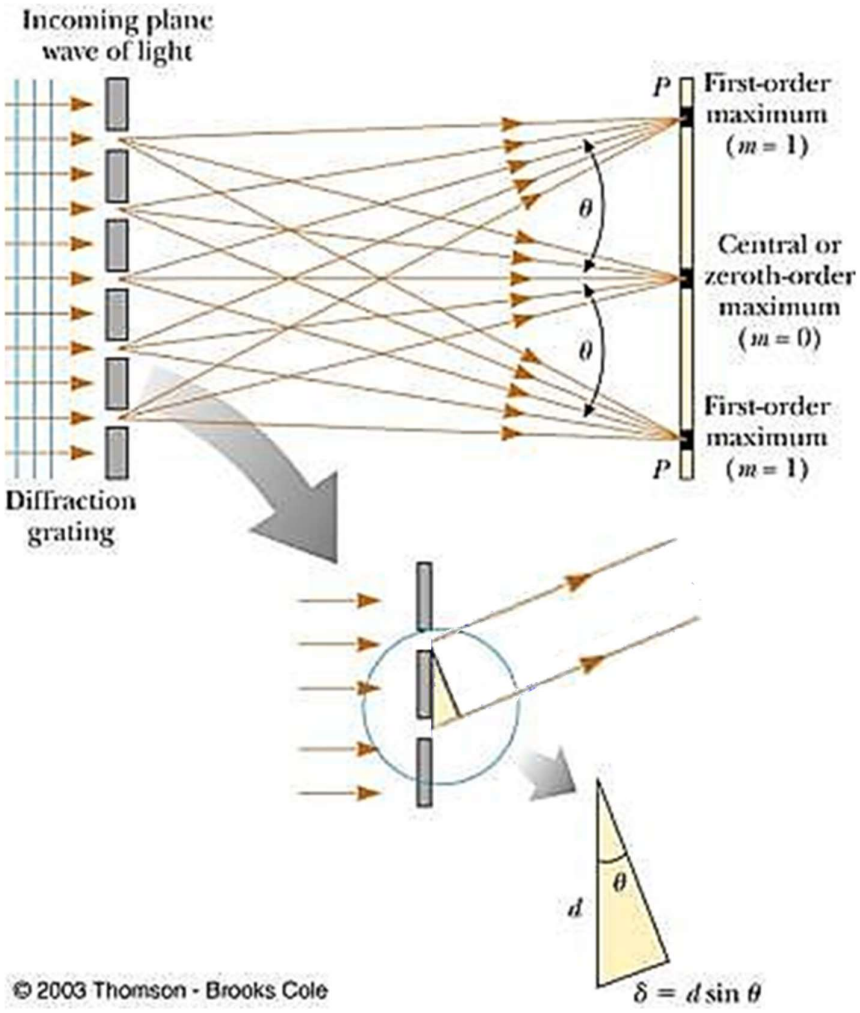


Fig. 10.6.1.

To see how a diffraction grating pattern differs from a Young's double-slit pattern and why it is so useful, we will analyse the diffraction pattern for various number of slits of the same slits-separation *d*.

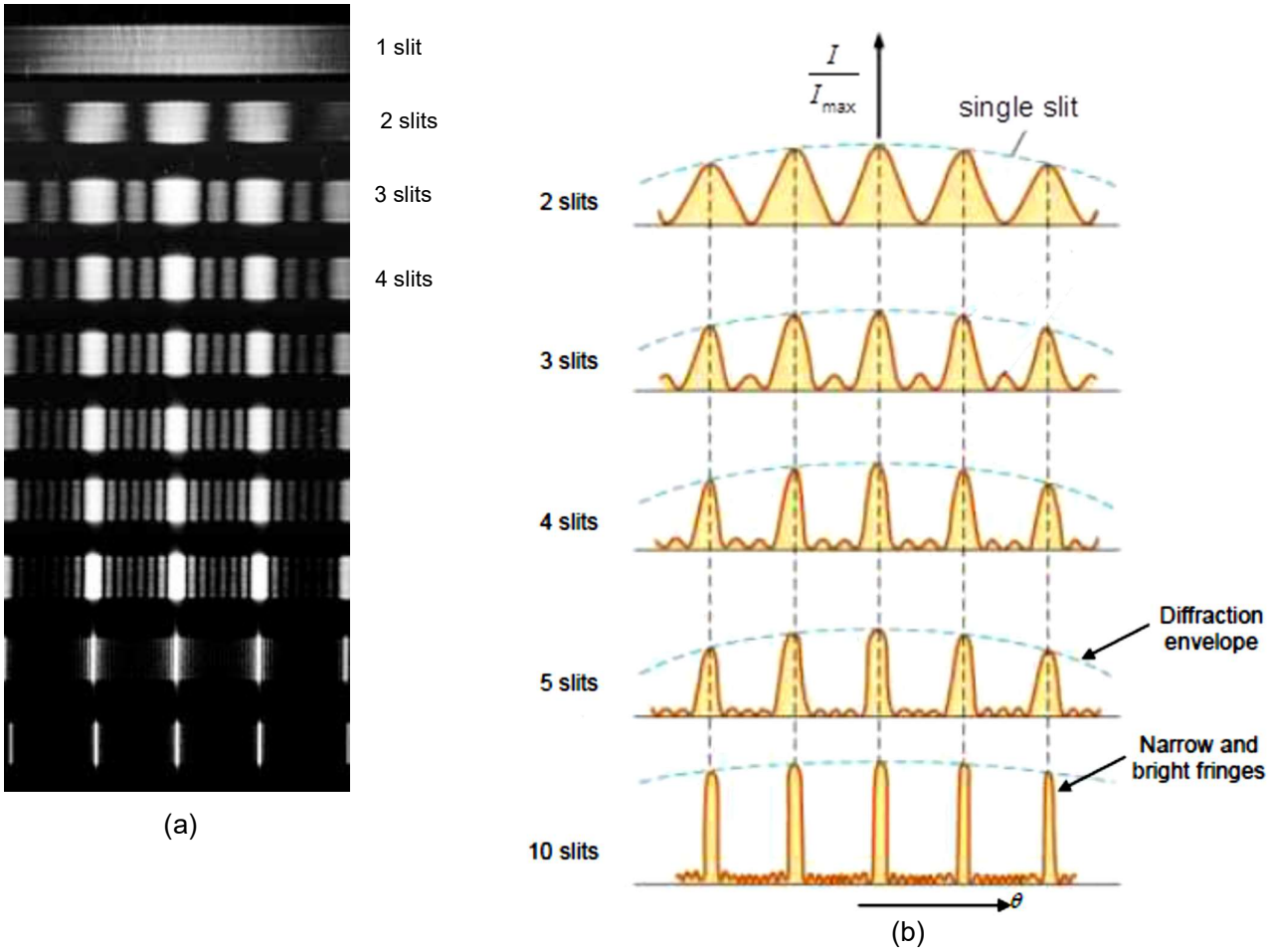


Fig 10.6.2. This shows the diffraction pattern of one slit, then two, then three, four, five, six, and more. When there are many slits, we call it a diffraction grating. The more slits, the narrower the fringes become. The pictures (a) above were taken with monochromatic light. The intensity distribution is shown in (b).

From Fig. 10.6.2(a) above, we make the following observations:

- For the same slit separation *d*, the **positions of the maxima remain the same**.
- As the number of slits increase, the maxima become narrower (i.e. the fringes become sharper on the screen).
- As the number of slits increase, the intensity of the maxima increases.

Therefore, we can infer that for a grating (no. of slits $\rightarrow \infty$) the **fringe pattern will have very good contrast** (narrower and brighter maxima separated by broad regions of dark minima).

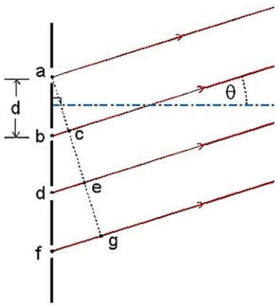
Condition for Maxima for a Diffraction Grating

- From Fig. 10.6.2(b) above, we notice that for same value of *d*, **the positions of the maxima (bright fringes) of a grating coincide with the positions of the maxima for a double slit:**
- We can therefore infer the equation that determines the positions of the maxima for a *diffraction grating pattern*.

Condition for Constructive Interference (Maxima) for a Diffraction Grating

For the *nth order* ***maxima***,

$$d \sin\theta = n\lambda$$



where d : separation between the slits
θ : angular position of the bright fringe from the principle axis
n : order number (n=0, 1, 2, 3,…)
λ : wavelength of light passing through the grating

Commercially available gratings are labelled by the *no. of lines per unit length, N*. The relationship between *N* and *d* are given by

$$N = \frac{1}{d}$$

Note :

1 In this case, we *do not define* the conditions for *destructive interference* as the width of the region between the bright fringes is very large.

2 For most of the gratings in the laboratory, the slit separations are much smaller than that of the double slits. Hence, *you will notice that the positions of the first and higher orders of the bright fringes lie much further away from the central bright fringe and hence small angle approximation (that θ is small) does not hold*. Therefore, in most questions for diffraction grating it is not meaningful to talk about the *fringe separation* as the angular separation between consecutive fringes are not constant, in fact *they increase as the order increases*.

Although the formula for constructive interference for the double slit and the diffraction grating is the same, the formula for fringe separation for double slits, $\Delta y = \frac{\lambda L}{d}$, CANNOT be used in a grating, unless the fringes are shown to be regularly spaced.

3 **Grating vs. Young's Double Slit**

The grating is often the preferred choice over the double-slit for the following reasons. For a diffraction grating,

- Fringes are sharper and positions are more precisely determined. There is therefore less uncertainty in measuring the position of the fringes, and less uncertainty in the computation of wavelength.

- Typically the angular displacements are large, so percentage uncertainty in measuring the angular displacement becomes relatively lower. This translates to a relatively lower percentage uncertainty in measuring wavelength as well.

4 Determination of Wavelength Using Grating[4]

- **Advantages and disadvantages of using the higher orders to determine the wavelength as compared to first order:**

  Using $d\sin\theta = m\lambda$, if we know the value of the slits separation ***d***, we can generally determine the wavelength of the light used by measuring the value of θ for a particular order of fringe. There are both advantages and disadvantages in using the higher orders to determine this wavelength.

  - **<u>Advantages :</u>**
    - angular displacement θ is larger, therefore there is less percentage uncertainty in measuring the angle.
  - **<u>Disadvantages :</u>**
    - May not be as bright as the first order bright fringe.

Example 10.6.1 : Diffraction Grating

White light from a source passes through a filter which transmits only wavelengths of 400 nm to 600 nm. When the filtered light falls normally on a diffraction grating, light of wavelength 600 nm in one order of the spectrum is diffracted at the same angle, 30°, as the 400 nm light in the adjacent order.

Find the number of lines per mm for the grating. [Ans: 417 lines per mm]

Solution:

[4] The spectrometer is a piece of equipment that is commonly used with the grating to determine the wavelength of light. Do read more about it in Appendix D (Pg. 49).

Example 10.6.2 : Maximum no. of Fringes Observable.

A monochromatic source of 495 nm is incident normally on a diffraction grating which has 500 lines per mm. How many diffraction lines can be observed on the screen? . [Ans: 9]

Strategy: The maximum angular deviation from the principle axis is 90^0. Using this condition, we can determine the maximum order and hence find the theoretical maximum number of fringes observable on the screen.

Solution:

- **You may now complete Tutorial 10B on Double Slits and Diffraction Grating.**

Tutorial 10B: Young's Double Slits and Diffraction Grating

Self-Practice Questions on Young's Double Slits

S1 In a Young's double slit experiment, sodium light of wavelength 0.59 x 10^{-6} m was used to illuminate a double slit with separation 0.36 mm. If the fringes are observed at a distance of 30 cm from the double slits, calculate the fringe separation.

S2 When red monochromatic light of wavelength 0.70 μm is used in the Young's double slit arrangement, fringes with separation 0.60 mm are observed. The slit separation is 0.40 mm. Find the fringe separation if (independently)

(a) yellow light of wavelength 0.60 μm is used;
(b) the slit separation becomes 0.30 mm;
(c) the slit separation is 0.30 mm and slits to fringe distance is doubled.

S3 **J88/I/10; N92/I/11.** Under what conditions will the bright fringes of a double-slit light interference pattern be farthest apart?

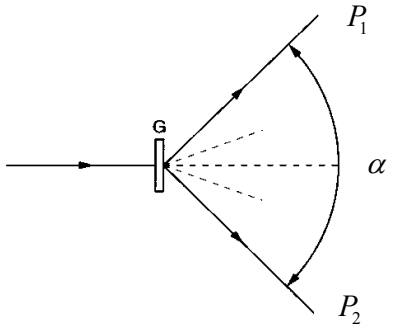
	distance between slits	*distance from slits to screen*	wavelength of source
A	small	small	short
B	small	large	short
C	small	large	long
D	large	small	short
E	large	small	long

Self-Practice Questions on Diffraction Grating

S4 What is the wavelength of light which gives a first order maximum at an angle of 22.5° when incident normally on a grating with 600 lines mm^{-1}?

S5 **N99/I/11**

A parallel beam of monochromatic light of wavelength λ is incident normally on a diffraction grating G. The angle between the directions of the two second-order diffracted beams at P, and P_2 is α, as shown.



What is the spacing of the lines on the grating?

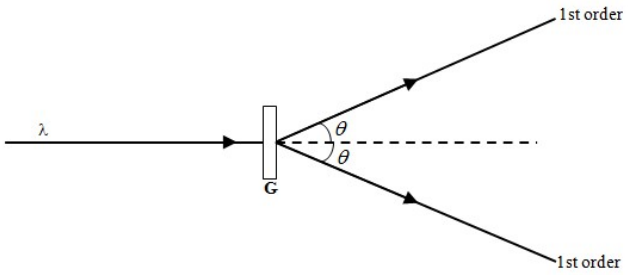
A $\frac{2\lambda}{\sin\alpha}$ **B** $\frac{\lambda}{\sin\alpha}$ **C** $\frac{2\lambda}{\sin(\alpha/2)}$ **D** $\frac{\lambda}{\sin(\alpha/2)}$

S6 In a diffraction grating experiment, the first order image of the 435.8 nm blue light from a commercial mercury vapour discharge lamp occurred at an angle of 15.8°. A first order red line was also observed at 23.7°, thought to be produced by an impurity in the mercury.
The wavelengths of red lines of various elements are listed below. Which element is the impurity in the mercury lamp?

	element	wavelength / nm
A	zinc	636.0
B	cadmium	643.3
C	hydrogen	656.3
D	neon	670.8
E	caesium	697.8

S7 A source emits spectral lines of wavelength 589 nm and 615 nm. This light is incident normally on a diffraction grating having 600 lines per mm. Calculate the angular separation between the first order diffracted waves. Find the maximum order for each of the wavelengths.

S8 Monochromatic light of wavelength λ, in the visible light range, is incident normally on a diffraction grating of about 500 lines per mm.



If the angular displacement from the principal axis, of the first order diffraction fringe is found at θ,

(i) Where would the angular displacement θ_2 of the second order diffraction fringe be most likely found?

A $0 < \theta_2 < \theta$ **B** $\theta_2 < 2\theta$ **C** $\theta_2 > 2\theta$ **D** $\theta_2 = 2\theta$

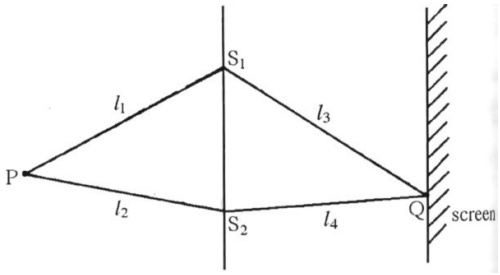
(ii) The light is now replaced by another monochromatic source of wavelength $\lambda' > \lambda$, still in the visible light range. Would the 1st order diffraction fringe be found at an angular displacement θ' such that

A $\theta' = \theta$ **B** $\theta' < \theta$ **C** $\theta' > \theta$

Discussion Questions

Young's Double Slits:

D1 N80/II/9 Two identical narrow slits S_1 and S_2 are illuminated by light of wavelength λ from a point source P.



If, as shown in the diagram above, the light is allowed to fall on a screen, and if *m* is a positive integer, the condition for destructive interference at Q is that

A $(l_1 - l_2) = (2m+1)\lambda/2$

B $(l_3 - l_4) = (2m+1)\lambda/2$

C $(l_3 - l_4) = m\lambda$

D $(l_1 + l_3) - (l_2 + l_4) = (2m+1)\lambda/2$

E $(l_1 + l_3) - (l_2 + l_4) = m\lambda$

D2 **N06/II/6**. Coherent light is incident normally on a double slit, as shown in Fig. 6.1.

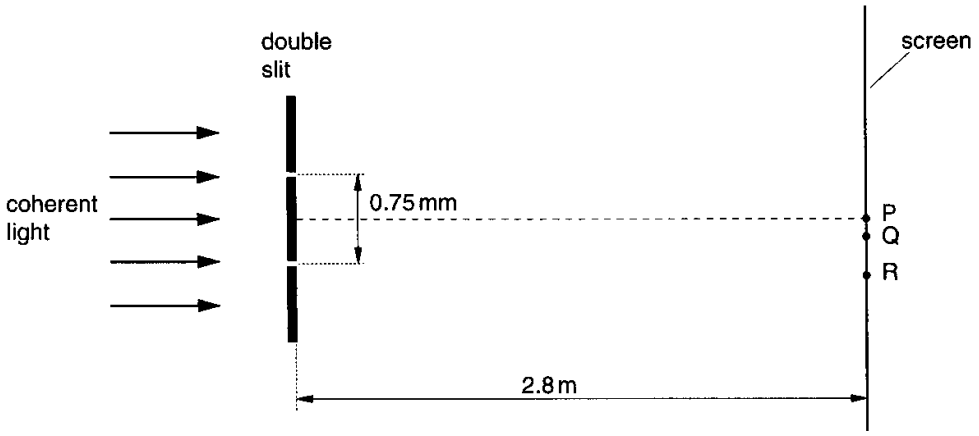


Fig. 6.1 (not to scale)

The separation of the slits in the double-slit arrangement is 0.75 mm. A screen is placed parallel to, and at a distance of 2.8 m from, the double slit. P is a point on the screen that is equidistant from the two slits. Fig. 6.2 (below) shows the variation with distance from P of the intensity I of the light on the screen.

(a) Calculate the wavelength, in nm, of the coherent light. [4]

(b) Points Q and R are points on the screen. Their positions are indicated on Fig. 6.2.
Determine the phase angle between the waves from the double slit when the waves meet at
(i) point Q, [1]
(ii) point R. [2]

(c) Suggest why the maxima on Fig. 6.2 are not all of the same intensity. [2]

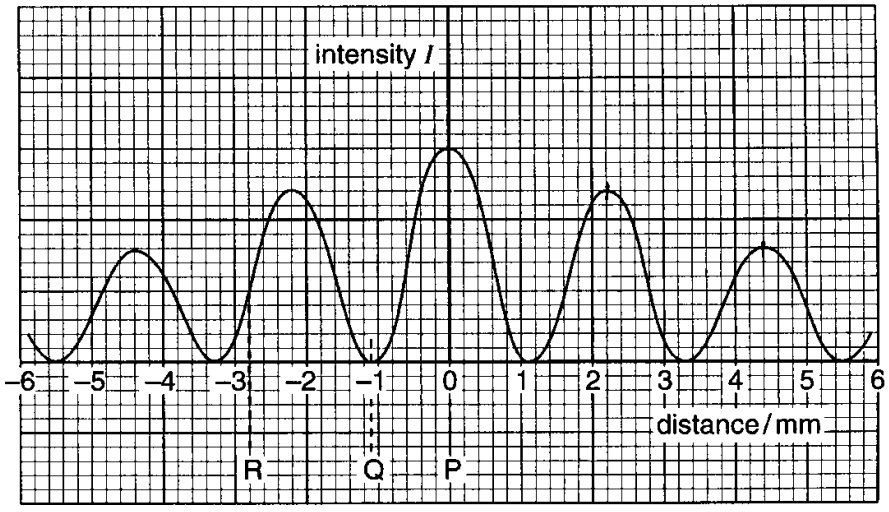


Fig. 6.2

D3 N04/III/3

(b) (i) Show how the principle of superposition of waves can be used to explain the formation of two-source interference fringes. [3]

(ii) Two-source interference fringes using light can only be obtained if light from the two sources is coherent. Explain

1. the meaning of the term *coherent,* [1]

2. why, in practice, interference fringes can be seen only if light from a single source is split into two. [2]

(iii) Coherent, monochromatic light from two narrow slits a distance 0.38 mm apart causes an interference pattern on a screen 1.20 m from the slits, as illustrated in Fig. 3.1.

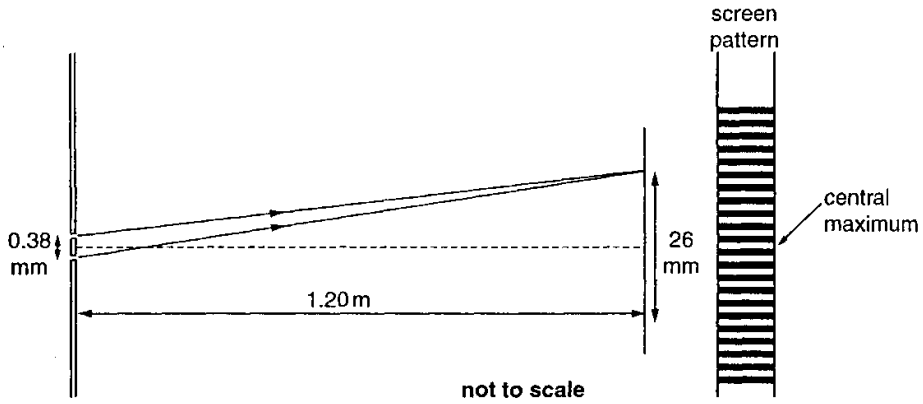


Fig. 3.1

The distance from the sixth bright fringe on one side of the pattern to the sixth bright fringe on the other side of the pattern is found to be 26 mm. Calculate the wavelength of the monochromatic light. [3]

(iv) State the experimental advantage gained by determining the fringe width in the way that was used in **(iii)**. [1]

(v) Another way of obtaining fringes similar to those described in **(iii)** is illustrated in Fig. 3.2.

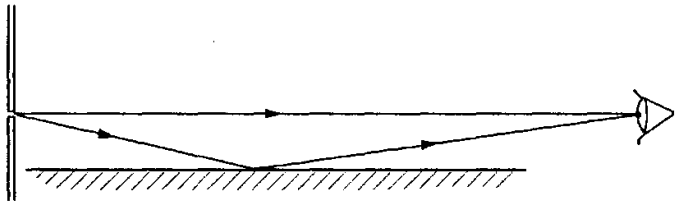
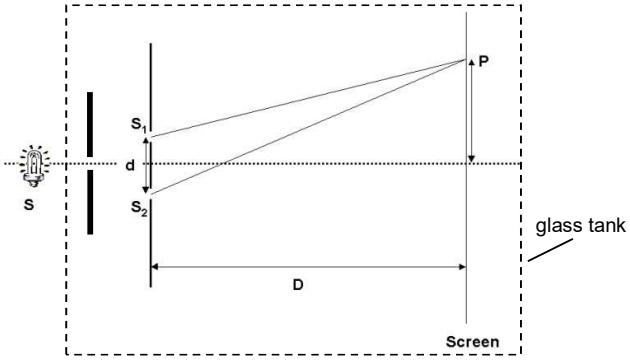


Fig. 3.2

A single slit is viewed both directly and by reflection from a mirror surface. Explain why this system produces a fringe pattern. [2]

D4 N99/II/3 (modified)



The wavelength of a monochromatic light source is to be determined by means of a double-slit interference experiment. Part of the setup is placed in a transparent glass tank as shown in the diagram above.

Initially, the glass tank is empty. State and explain what changes, if any, occur in the pattern of the fringes observed when each of the following changes is made separately:

(a) increasing the intensity of the light source incident on the double slits,

(b) increasing the distance between the double slit and the screen,

(c) reducing the intensity of light incident on one slit of the double slit.

(d) the light source is replaced with one of white light.

Diffraction Gratings:

D5 J85/I/11

Monochromatic light of wavelength λ is incident normally on a diffraction grating consisting of alternate opaque strips of width x and transparent strips of width y. The angle between the emerging zero-order and first-order spectra depends on

A *x*, *y* and λ **B** *x* and λ only **C** *y* and λ only

D *x* and *y* only **E** λ only

D6 Hutchings Pg 231, Question 13.10

The visible spectrum extends from 400 nm to 750 nm.

(a) If observed with a diffraction grating having 480 lines per mm, find the angular width of the first and third order spectra.

(b) Give two reasons why such a spectral will normally be viewed in the first order.

D7 J97/III/3

A diffraction grating with 250 lines per millimetre is placed in front of a monochromatic source of red light. A screen placed 200 cm beyond the grating has red light images measured at certain positions on a scale on the screen, as shown in Fig. 3.1.

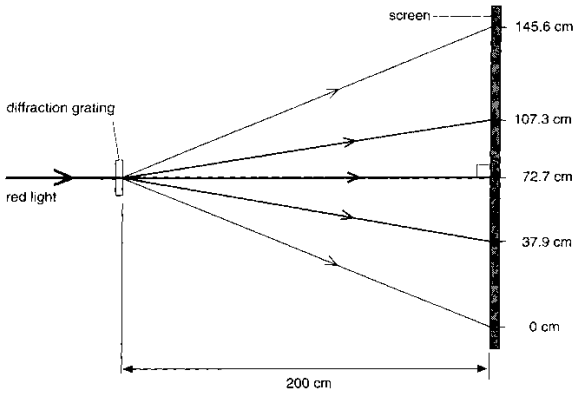


Fig. 3.1

(a) **(i)** Use the first order spectrum to deduce a value for the wavelength of the red light.
(ii) Make a check, using the second order spectrum, to show that your calculation is correct. [6]

(b) How would the pattern obtained be different if blue light were used in place of red light? You are not expected to make any calculations when answering this part of the question. [3]

(c) What main problem would arise if the experiment were repeated with infra-red radiation? Suggest how this problem could be overcome. [2]

D8 N2000/III/3(part), J96/III/3 (part)- (modified)

(a) An experiment is set up to demonstrate the behaviour of water waves in a ripple tank. Draw a diagram, one in each case, to show the diffraction of water waves through
(i) a narrow gap.
(ii) a wide gap. [3]

(b) Light from a low pressure sodium lamp consists mostly of two wavelengths, 588.99 nm and 589.59 nm. This light is allowed to fall normally on a diffraction grating with 500.00 lines per millimetre.
(i) Calculate the maximum angular separation between the light of the two wavelengths.
(ii) What problem would be likely to arise in observing the spectral lines in the order in (i)? [8]

D9 TYS Pg 127 Qn 28 N2000/II/4. N94/III/2 (part)

Light from a distant source of monochromatic light of wavelength 590 nm passes through a fine nylon mesh. The light is then incident on a screen, as illustrated in Fig. 4.1.

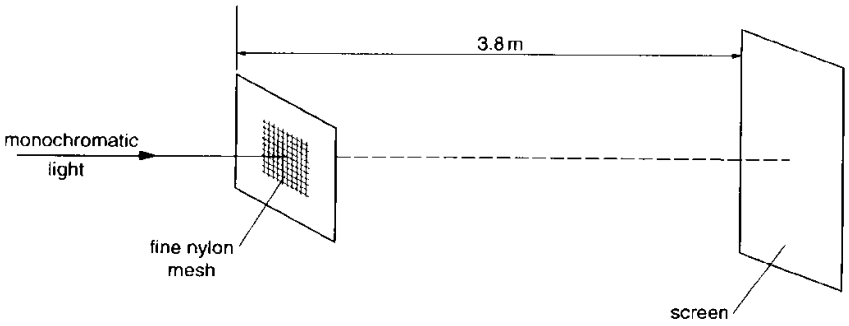


Fig. 4.1

The threads of the nylon mesh act as a diffraction grating with lines in the horizontal and in the vertical direction. **Part of the pattern of spots** of light on the screen is shown in Fig. 4.2.

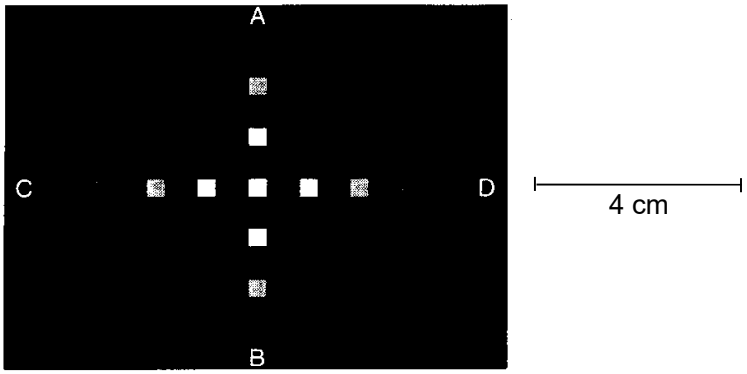


Fig. 4.2

(a) Explain what is meant by the *diffraction* of light. [2]

(b) State which line of spots of light, AB or CD, is produced by the *horizontal* nylon threads.

(c) Calculate the angle, in radians, between the orders of the diffracted light. [2]

(d) Using your answer to (c), determine the number of nylon threads per millimetre of the mesh. [4]

(e) The spots are observed to be almost equally spaced apart both horizontally and vertically. What does this suggest about the structure of the nylon mesh?

(f) Patterns such as this can be used to determine the spacing of atoms in crystals. Suggest one problem that occurs in carrying out such an experiment with the wavelength used. Indicate how a different wavelength of radiation can overcome this problem.

Numerical Answers for Tutorial 10B: Young's Double Slit and Diffraction Grating

S1 4.92×10^{-4} m **S2**(a) 0.51 mm (b) 0.80 mm (c) 1.60 mm
S4 638 nm **S7** 0.959°, 2, 2
D2(a) 589 nm (b)(i) π rad (ii) 8.0 rad
D3 (iii) 6.86×10^{-7} m
D6 (a) First at 10.0°, second at 23.5°. Incomplete third starting at 35.2°.
D7 (a)(i) 684 nm **D8** (b)(i) 0.110° **D9** (c) 2.6×10^{-3} rad (d) 4.5

10.7 Diffraction of Waves

What is diffraction?

Diffraction is the ***spreading*** of a wave (into the "geometrical shadow") after passing through a slit or around an obstacle.

Diffraction effects become more apparent when the width of the opening is small enough compared to the wavelength.

10.7.1 Diffraction of water waves

When waves meets a barrier in which there is an opening, if the width of the opening is much larger than the wavelength of the waves, the wave emerging from the opening continues to move in a straight line (apart from some small spreading at the edge). If the width of the opening is comparable or smaller than the wavelength, then the bending is a lot more obvious, and emergent wave becomes more circular.

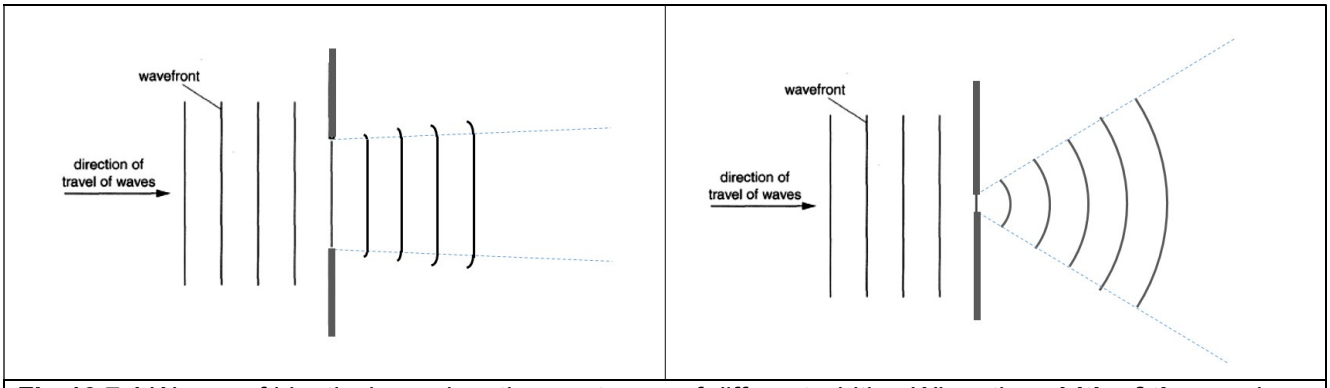


Fig 10.7.1 Waves of identical wavelengths meet gaps of different widths. When the **width of the gap** is **comparable** to the **wavelength**, diffraction effects are more obvious.

What is meant by "Geometrical Shadow"

Consider a series of particles projected (diagram below) separately into an aperture and obstacle.

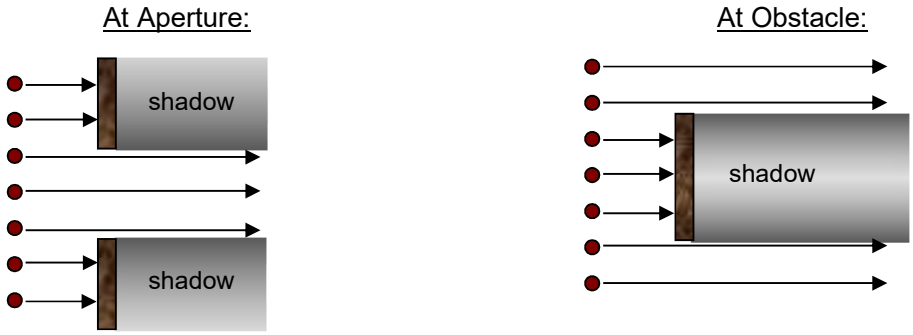


Fig. 10.7.2

As the particles are blocked by the barriers, there will be a sharp 'shadow' region where the particles do not reach. This shadow is known as the "geometrical shadow" of the barriers.

Diffraction is also observed for waves meeting an obstacle in its path. Again, **the relative size of the obstacle to the wavelength will determine the effect of diffraction** as shown in Figure 10.7.3. It is found that in order for the waves to spread around the edges of the obstacle to extend into the geometric shadow of the obstacle, the width of the obstacle should be small enough compared to the wavelength.

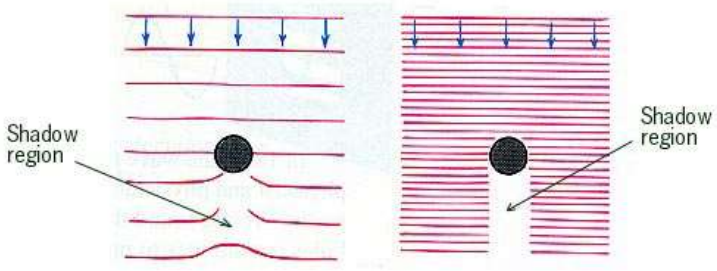


Fig. 10.7.3 The waves will bend if the length of the obstacle is comparable to the wavelength of the waves.

Since light and sound demonstrate diffraction effects, they are said to behave like waves. This diffraction effect cannot be explained by particle theory (i.e. diffraction effect is unique to waves).

10.7.2 Single Slit Diffraction Pattern

Diffraction Pattern of a Single Slit

We might expect that light passing through a small aperture would result in a broad region of light on a screen, due to the spreading of the light as it passes through the opening.

We actually find something more interesting.

A diffraction pattern consisting of light and dark areas is observed (see Figure 10.7.4), somewhat similar to the interference pattern discussed earlier.

This pattern consists of a broad intense central band (called the central maximum), flanked by a series of alternating bright and dark bands.

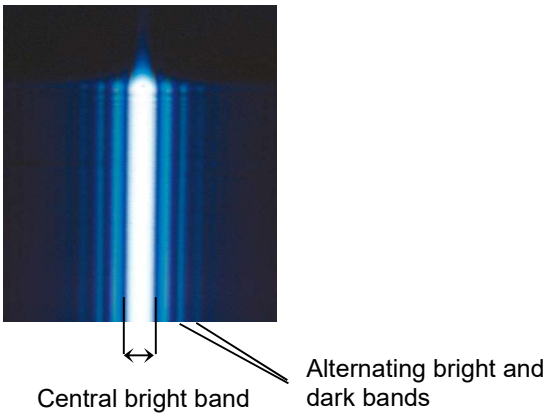


Fig. 10.7.4

Single-slit Diffraction of light

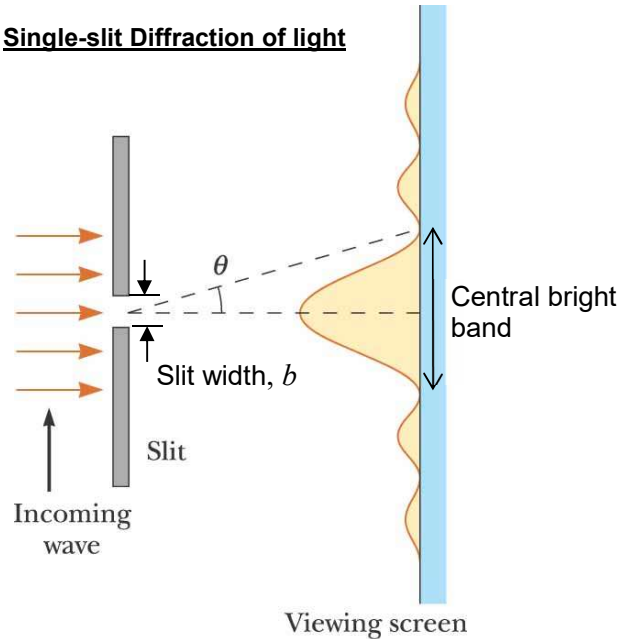


Fig. 10.7.5

In a single slit diffraction, **light waves of wavelength** λ passing through a **single slit of width** ***b*** undergoes diffraction to produce a broad intense central band with an intensity profile that is maximum at the centre, and tapers off towards both sides as shown in Figure 10.7.5.

You are not required to explain formally these observations for the A-level Physics syllabus.
Nonetheless, for those interested, these observations can be explained by means of Huygen's Principle, which explains how the slit can be modeled as row of secondary point sources, the waves from which interfere with one another to create an interference pattern.
A detailed explanation can be found in Appendix A.

The **first minimum intensity of the single slit diffraction pattern** occurs at **angle** θ from the central maximum where

$$\sin\theta = \frac{\lambda}{b}$$

This angle θ is important because it tells us the angular spread (2θ) which contains most of the energy of the wave. It can be seen from this that as ***b* (the slit width) is reduced, θ gets larger and the pattern is more spreads out**. The **pattern also gets dimmer as less light energy passes through a smaller slit**.
A more complete analysis of the diffraction pattern can be found in Appendix B.

Width of the single slit pattern

For small angles the **angular spread of the diffraction pattern** is inversely proportional to the **slit width *b*** or, more precisely, to the **ratio of *b* to the wavelength** λ**.** Figure 10.7.6 shows graphs of intensity *I* as a function of the angle for three values of the ratio b/λ.

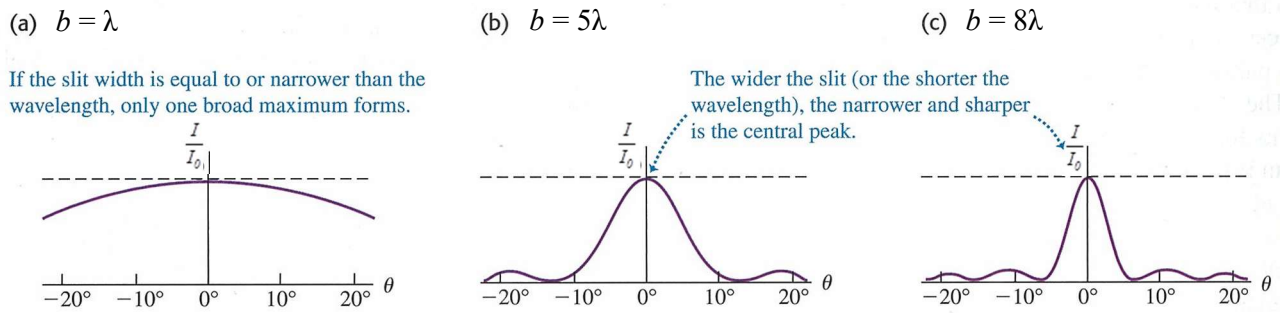


Fig. 10.5.6 The single slit diffraction pattern depends on the ratio of the slit width b to the wavelength λ.

Example 10.7.1: Single slit diffraction

The figure (Fig. 10.7.7a) shows light from a He-Ne laser of 633 nm is incident on a 2.0 x 10^{-4} m wide slit and the resulting diffraction pattern on the screen as shown in Fig 10.7.7b. What is the width of the central maxima (the distance between the dark fringes on either side of the central maxima) on a screen 2.0 m away? . [Ans:12.7 mm]

Solution:

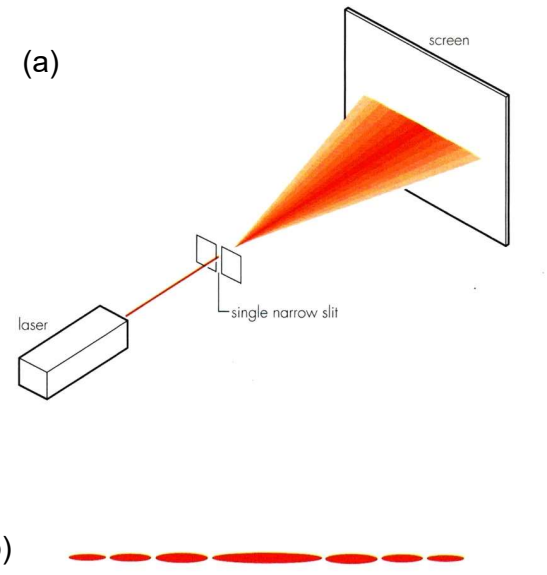


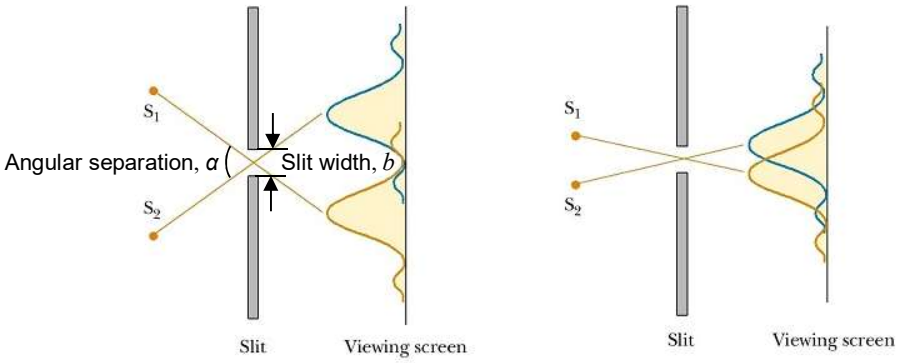
Fig. 10.7.7

10.7.3 Resolving Power and Rayleigh's Criterion

The resolving power of an optical instrument is the ability of the instrument to differentiate clearly between the images of two close objects, i.e., the two images are seen as two distinct images. The resolving power is measured in terms of angular separation (units in radians) of the two images.

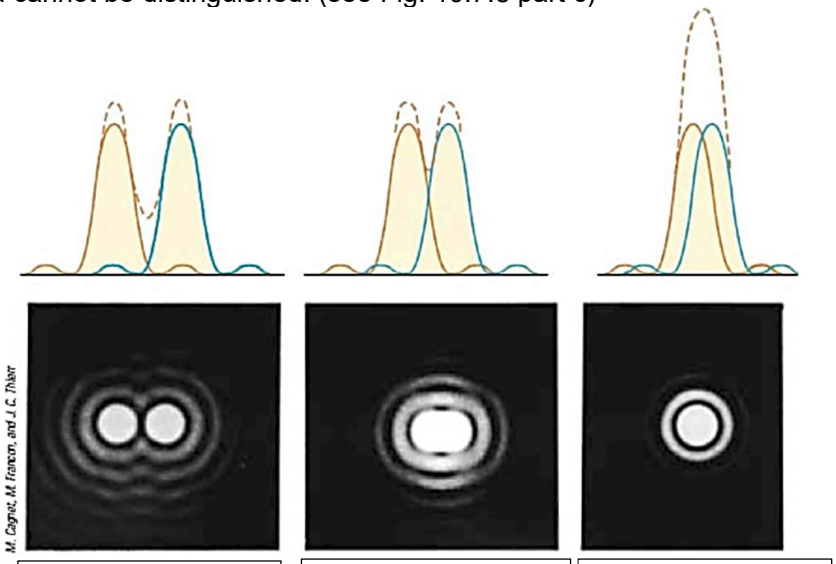
As most images are viewed through an aperture as in the case of microscopes and telescopes, it is identical to viewing the diffraction pattern produced. Hence, the diffraction of light affects greatly or limits the resolving power of an optical instrument.

To understand this limitation, let us consider two light sources far away from a narrow slit of **slit width *b***, with angular separation of *α*.



(a) Little overlap of diffraction patterns (b) Significant overlap of diffraction patterns

Since the two sources are usually incoherent (such as two distant stars), let's assume that no interference occurs between the two light passing through the slit. So each light source simply produces its own diffraction pattern on the screen. What is observed on the screen will be the summation of two diffraction patterns, one from each source. If the sources are close together, their diffraction patterns overlap; if their angular separation is small enough, their diffraction patterns overlap almost completely and cannot be distinguished. (see Fig. 10.7.8 part c)



(a) sources are far apart and **patterns are well resolved.** **($\alpha > \theta_{min}$)**

(b) The sources are closer together and the patterns are just resolved. **($\alpha \approx \theta_{min}$)**

(c) the sources are so close that the **patterns are not resolved**. **($\alpha < \theta_{min}$)**

Fig 10.7.8
Individual diffraction patterns of two point sources (solid curves) and the resultant patterns (dashed curves) for various separation of sources as light passes through a circular aperture. In each case, the dashed curve is a summation of the two solid curves.

The generally accepted criterion for whether two images are distinguishable is the **Rayleigh criterion**, which states that **two images are just resolved or distinguishable when the central maximum of one image coincides with the first minimum of the other image** (see Fig. 10.7.8 part b).

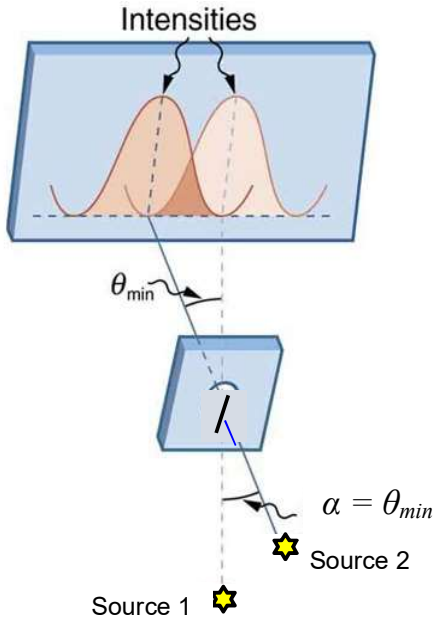


Fig. 10.7.9

(Image from http://cnx.org/contents/9ANhisjh@5/Limits-of-Resolution-The-Rayle)

Hence according to the Rayleigh criterion, **two images are just resolved or distinguishable** when the **angular separation between the two sources, *α*** is (approximately) equal to the **angular separation *θmin*** between the central maximum and the first minimum of the images (See Fig. 10.7.9).

From section 10.5.2 (single slit diffraction pattern), we learnt that the **first minimum** for a single slit diffraction occurs at an angle $\sin\theta_{min} = \frac{\lambda}{b}$ where *b* is the slit width.

When θ_{min} is small, we can use the approximation $\sin\theta_{min} \approx \theta_{min}$. The **Rayleigh criterion for the resolving power of a single aperture** is thus given by

$$\theta_{min} \approx \frac{\lambda}{b}$$

where ***b* is the slit width and *θmin* is in radians**.

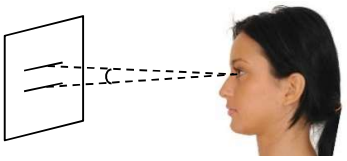
Hence we can also say that **the angle subtended by the two sources, *α* must be greater than $\frac{\lambda}{b}$ if the images are to be resolved**.

Note:

1. The resolving power of a telescope can be improved by using an objective lens of wide aperture (i.e. large diameter). Besides reducing the diffraction effects, more light is also transmitted through the telescope and a brighter image is observed.
2. In microscopes, the resolving power is improved by using radiation of much smaller wavelengths than visible light, such as ultra-violet. In electron microscopes, the wavelength of the electron beam is 10^5 times smaller than visible light, resulting in much higher resolution in the images. The wavelength associated with the electron beam will be covered in Chapter 17 Quantum Physics.
3. For circular lens aperture (e.g. telescope), the formula becomes $\sin\theta_{min} = \frac{1.22\lambda}{D}$, where *D* is the diameter of the circular lens. **Not required for A Level, for information only.**

Example 10.7.2

The resolving power of the eye may be determined by drawing two parallel lines at a distance of say 2 mm apart on a piece of card. The card is slowly moved away from the eye until the eye just cannot see the two lines as separate lines. The distance of the card from the eye is measured. Suppose for a particular person, the distance of the card from his eye is 5 m when his eye just fail to see the two lines distinctly. [Ans: (a) 0.0004 rad (b) 1.25 mm]



(a) Determine the angle subtended at the eye by the two lines.

(b) By approximating the wavelength of the visible light to be 500 nm, estimate the aperture of the pupil of the eye.

Solution:

- **You may now complete Tutorial 10C on Single Slit and Rayleigh's Criterion.**

Tutorial 10C: Single Slit and Rayleigh's Criterion

Self-Practice Questions

S1 Calculate the width of the central maximum in the single-slit diffraction pattern of yellow light of = 589.0 nm by a slit 0.250 mm wide viewed on a screen 2.50 m away.

S2 A single-slit pattern is formed by light passing through a narrow slit 0.050 mm wide. If the width of the central maximum is 3.8 cm on a screen 1.5 m away from the slit, what is the wavelength of the light?

S3 The images of two sources are just resolved. Which of the following is a correct statement of the Rayleigh criterion for this situation?

A The central maximum of the diffraction pattern of one source must coincide with the central maximum of the diffraction pattern of the other source.

B Light from the sources must pass through a circular aperture.

C Light from the sources must be coherent.

D The first minimum of the diffraction pattern of one source must coincide with the central maximum of the diffraction pattern of the other source

S4 Two binary stars emit radio waves of wavelength 6.0×10^{-2} m. The waves are received by a radio telescope whose collecting dish has a diameter of 120 m. The two stars are just resolved if their **minimum** angular separation in radians is of the order of

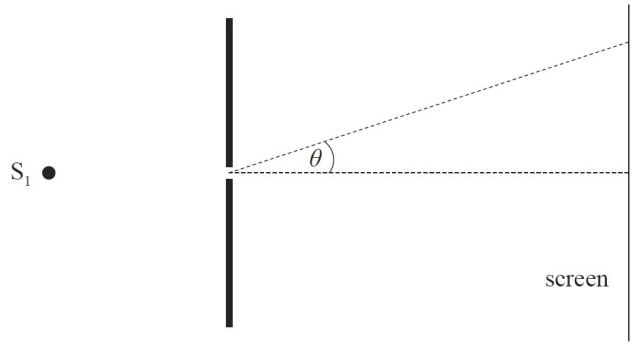
A 10^{4} **B** 10^{2} **C** 10^{-2} **D** 10^{-4}

Discussion Questions

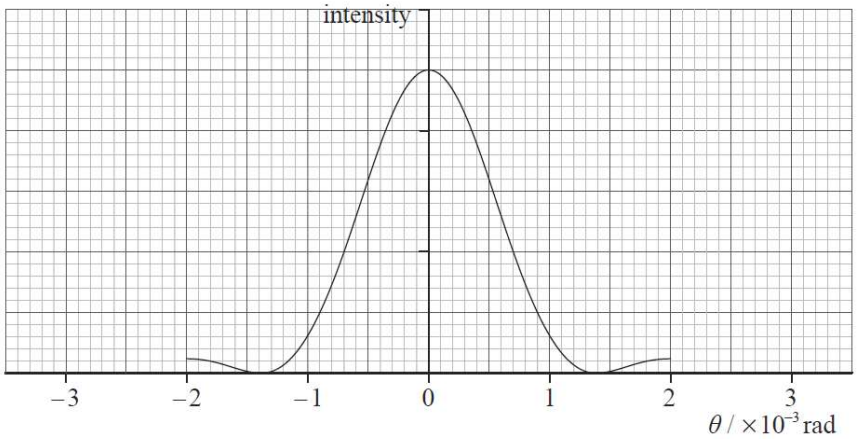
D1 **IB 2010**

This question is about diffraction and resolution.

(a) Light from a monochromatic point source S1 is incident on a narrow rectangular slit.

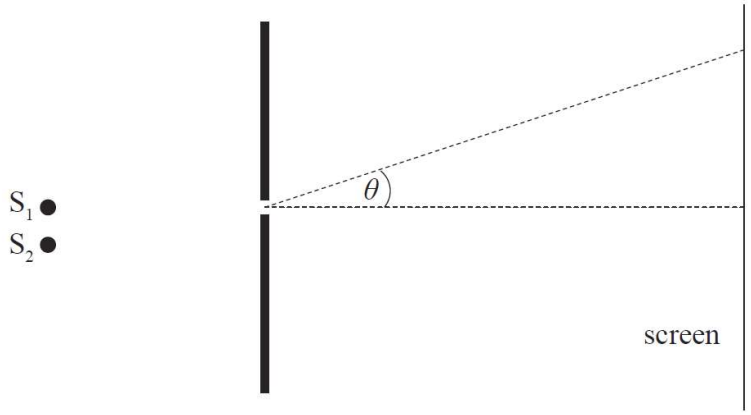


After passing through the slit, the light is incident on a screen some distance away from the slit. The graph shows how the intensity distribution on the screen varies with the angle θ shown in the diagram.



(i) The width of the slit is 4.0 ×10^{-4} m. Use data from the graph to calculate the wavelength of the light. [2]

(ii) An identical source S2 is placed close to S1 as shown.

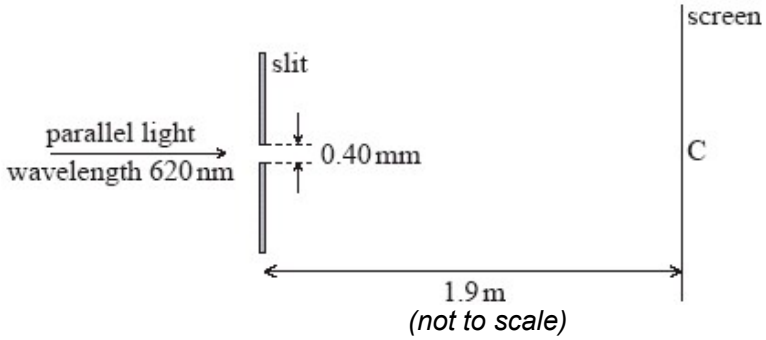


The images of the two sources on the screen are just resolved according to the Rayleigh criterion. On the graph above, draw the intensity distribution of the second source. [1]

D2 IB

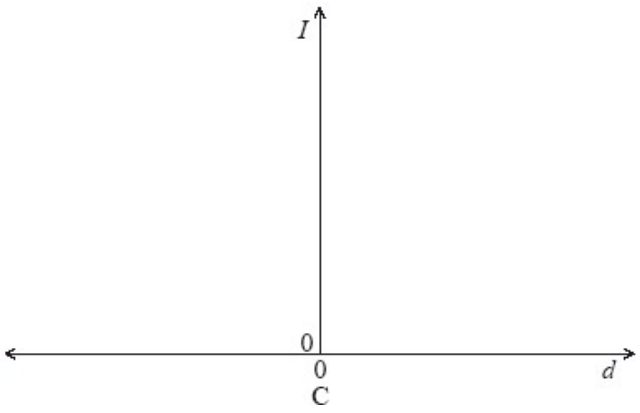
(a) (i) Explain what is meant by the diffraction of light. [2]

(ii) A parallel beam of monochromatic light from a laser is incident on a narrow slit. The diffracted light emerging from the slit is incident on a screen.



The slit width is 0.40 mm and it is 1.9 m from the screen. The wavelength of the light is 620 nm. Determine the width of the central maximum on the screen. [3]

(iii) The centre of the diffraction pattern produced on the screen is at C. On the axes sketch a graph to show how the intensity *I* of the light on the screen varies with the distance *d* from C. [3]

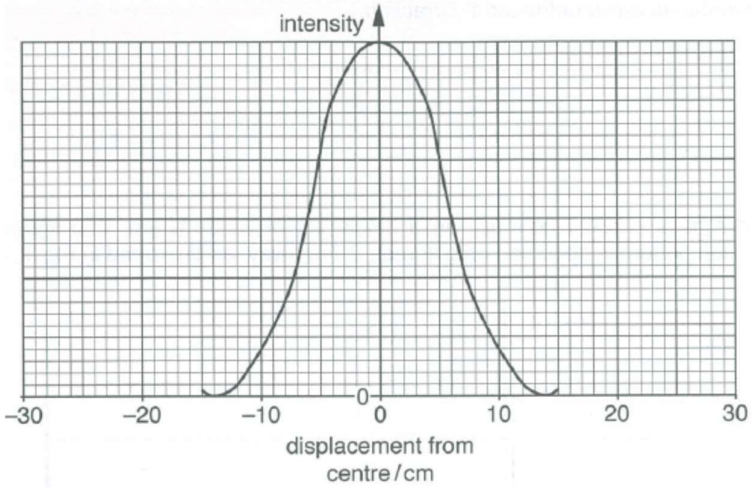


(b) (i) When two separate lasers are used as sources, the images of the slit formed by the light from each laser are resolved. State what is meant by the term resolved in this context. [1]

(ii) On the same axes, sketch a graph to show how the intensity *I* of the light from the second laser varies with distance along the screen if the two images are just resolved. Label this curve B. [2]

(iii) A car, with its two headlights switched on, is approaching an observer who has good eyesight. Outline why, at a long distance from the observer, the images of the headlights of the car are not resolved by the observer. [4]

D3 **Modified N2018/P2/Q3.** When laser light is incident normally on a single slit of width 12 μm, the light diffracts and a diffraction pattern is observed on a screen 2.7 m away. The screen is parallel to the plane of the single slit.
A diffraction pattern is formed on the screen. Part of the diffraction pattern showing the variation with displacement from the centre of the pattern of the intensity of the light is shown below.



(i) State and explain the changes to the observed diffraction pattern above when visible light of a longer wavelength is used.

(ii) When white light is incident on a single slit, the central fringe is coloured at the edges and has a white central region. Explain this observation.

D4 HCI 2019 Prelim P3 Q9(part)

Light of wavelength 650 nm is incident normally on a double slit such that the waves emerge from X and Y in phase, and reach a screen 1.5 m away, as shown in Fig. 9.4.

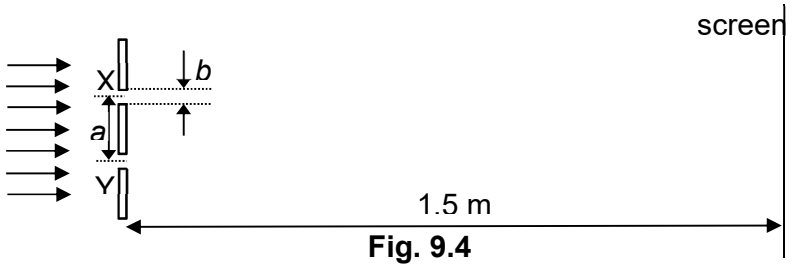


Fig. 9.4

The variation of intensity with distance along the screen is shown in Fig. 9.5.

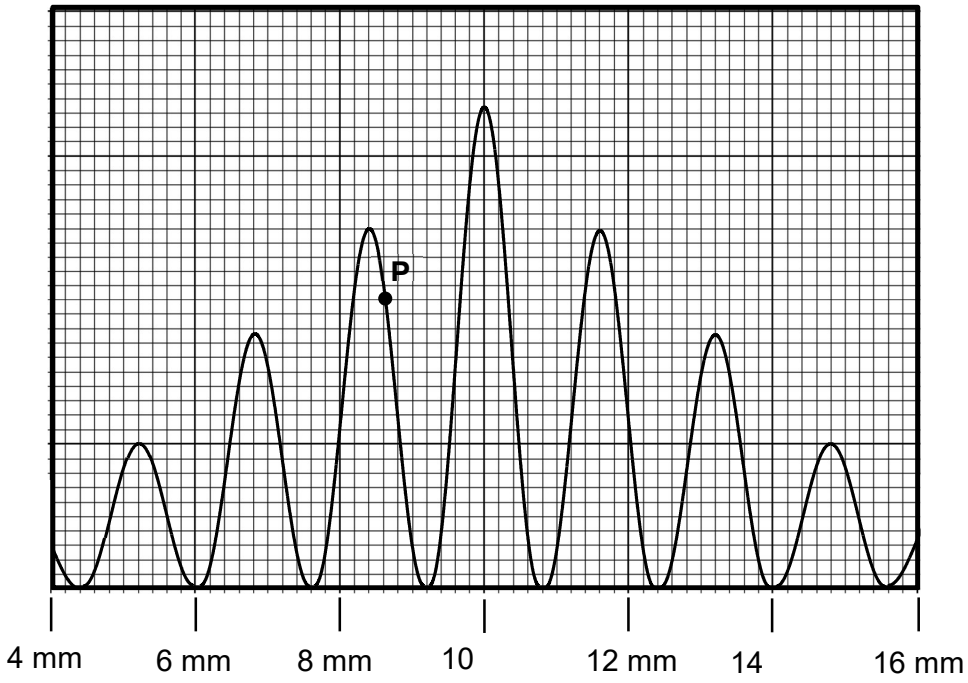


Fig. 9.5

(i) Explain how it can be deduced from Fig. 9.5 that the waves from the two slits are coherent. [1]

(ii) Determine the phase angle between the waves from the slits when the waves meet to produce the intensity at point **P** on the pattern of Fig. 9.5. [2]

(iii) Calculate the separation *a*, between the slits. [2]

(iv) Given that the 6th order bright fringe is the first missing order due to the diffraction envelope, calculate the width *b*, of each slit. [3]

Numerical Answers for Tutorial 10C: Single slit and Rayleigh's Criterion

S1 0.012 m **S2** 633 nm
D1 (a)(i) 5.6 x 10^{-7} m
D2 (a)(ii) 5.89 x 10^{-3} m
D4 (ii) 5.5 rad
(iii) 0.61 mm
(iv) 0.10 mm

10.8 Stationary Waves

Consider two identical waves moving in opposite directions overlapping in the same region of space. Figure 10.8.1 shows wave P moving towards the right and wave Q moving towards the left. By applying the principle of superposition, we know that during the time the two waves overlap, the resultant displacements is given by the vector sum of the displacements of the primary waves at that location.

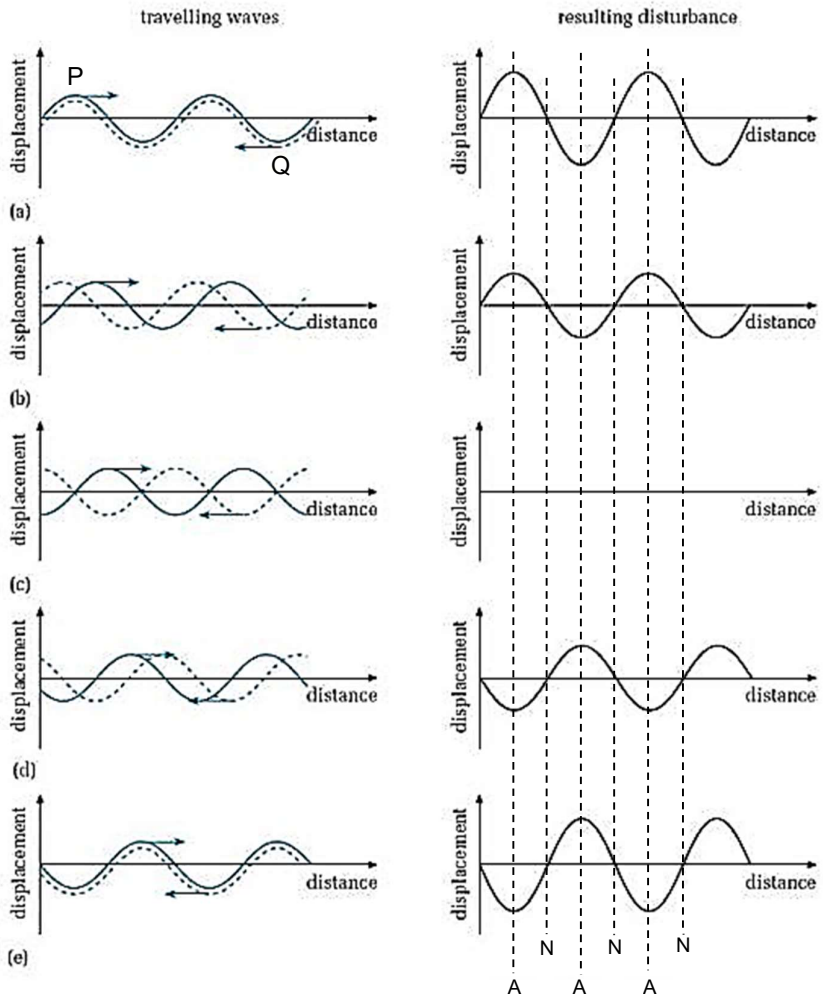


Fig. 10.8.1

The diagram shows how the superposition of 2 identical waves moving in opposite directions overlap to form a stationary wave.

Because the wave pattern does not appear to move in either direction along the wave, the resultant waveform is known as a **standing wave** or **stationary wave**.

We can observe from the diagram that amplitude of vibration for every point of the resultant wave varies.

- There are particular points called **nodes** (*N*) that never move, i.e. the amplitude of vibration is zero.
- Midway between the nodes are points called **antinodes** (*A*), where the amplitude of vibration is greatest.

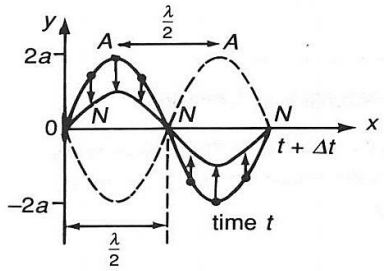
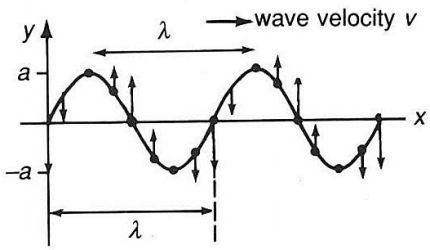
Formation of Stationary Wave (or Standing Wave)

Stationary Waves

When two identical waves of the **same amplitude**, **frequency** and **speed** but travelling in **opposite directions** are superposed together, the resultant wave obtained is called a stationary wave.

10.8.1 Characteristics of Stationary Wave

Let us examine the properties of the stationary wave by comparing it with a travelling wave.

	Standing Wave	Progressive Wave
Diagram		
Waveform	Does not advance.	Waveform advances with the velocity of the wave.
Amplitude	Varies according to position from zero at the nodes (permanently at rest) to a maximum of 2 *A* at the antinodes. (*A* is the amplitude of each of the component travelling waves that form the stationary wave.)	Apart from attenuation, it is the same for all particles in the path of the wave.
Frequency	All the particles vibrate in SHM with the same frequency as the component waves (except for those at the nodes which are at rest).	All particles vibrate in SHM with the frequency of the wave.
Phase	A segment (loop) refers to points between two adjacent nodes. Phase of all particles within a segment is the same (in phase). Particles in adjacent segments have a phase difference of π rad (in antiphase).	All particles within one wavelength have different phase.
Wavelength	2 x distance between a pair of adjacent nodes or antinodes.	Distance between two adjacent particles that are vibrating in phase.
Energy	No propagation of energy, but there is energy associated with the wave.	Energy propagates in the direction of the wave travel.

Example 10.8.1 : Stationary Waves

Fig 10.8.2 shows a car driven at a speed of 30 m s^{-1} along a straight road between 2 radio transmitters. The transmitters T_1 and T_2 are sending out the same programme, using a frequency of 1.50 MHz. The radio is heard to fade and strengthen regularly. [Ans: 3.3 s]

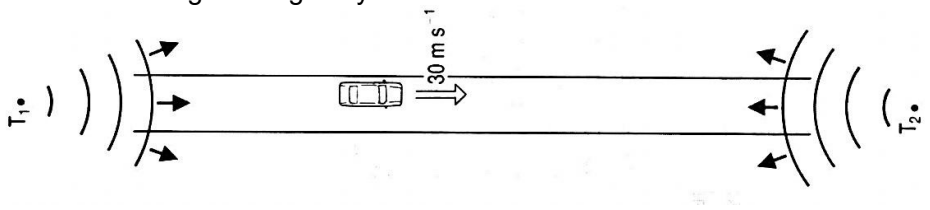


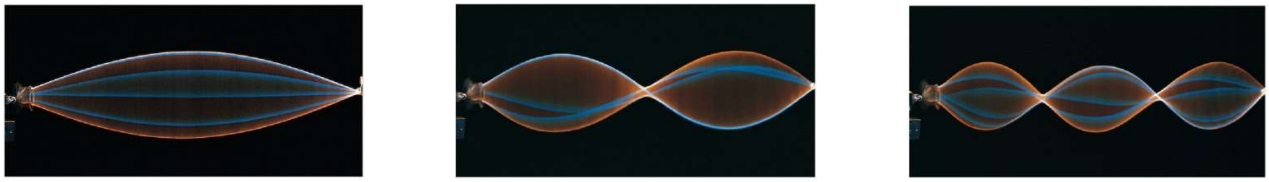
Fig. 10.8.2

What is the period of this regular fading? .

Solution:

10.8.2 Stationary Waves in Strings

Consider a string, fixed at one end, and the other end attached to an oscillator. If the oscillator moves up and down in simple harmonic motion, it will produce a continuous wave-train down the string. As the wave-train reaches the fixed end, it gets reflected back. The reflected wave-train will interfere with the wave-train from the source to form a resultant waveform on the string. It is observed that at certain frequencies of vibration of the source, we will be able to see some unique waveforms on the string as illustrated in Fig. 10.9.3(a)-(c).



(a) (b) (c)

Fig. 10.8.3

Time exposures of standing waves in a stretched string. At certain driven frequencies of vibration, the amplitude of vibration of the string become very large and the resultant waveform does not appear to progress along the string. (a) to (c) show standing waveforms at increasing frequencies of vibration.

The frequencies at which standing waves are produced are known as the **natural frequencies** or **resonant frequencies** of the string.

The different standing wave patterns shown in Fig. 10.8.3 are the different **"resonant modes of vibration"**. When the driving frequency matches that of the natural frequency of the string, the amplitude of oscillation generated are the largest.

This principle is very important as it is used in the generation of music in many stringed instruments like guitar and violin.

Resonant modes of vibrations in a stretched string fixed at two ends

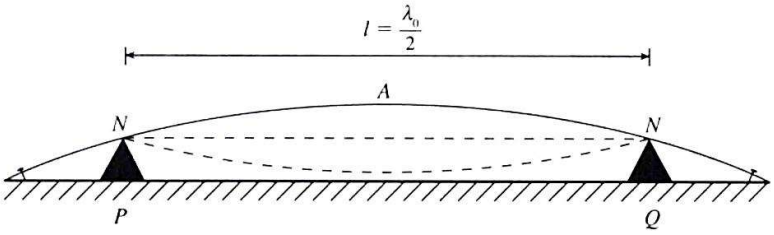


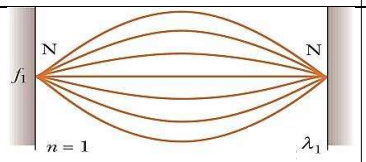
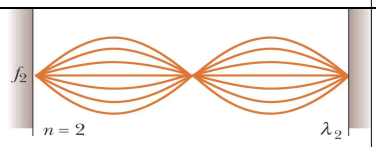
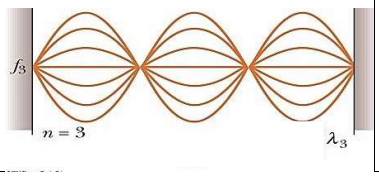
Fig. 10.8.4

Consider a wire stretched between 2 points and plucked in the middle, a transverse wave travels along the wire and is reflected at the ends. The reflected and re-reflected waves from the ends will continually interfere with each other to produce a stationary wave on the string.

The two *ends of the strings are fixed and thus must have zero displacement* and hence must be displacement nodes.

This additional constraint or boundary condition results in the string being allowed to have only certain wavelengths and frequencies of oscillations.

Fig. 10.8.4 shows the largest possible wavelength of stationary wave that can be produced on the wire - corresponding to the smallest frequency. This mode of oscillation is known as the **fundamental mode** and the lowest frequency of vibration f_1 is known as the **fundamental frequency.**

Mode of Vibration	Wavelength	Frequency	Harmonic	Overtone
	$\lambda_1 = 2L$	$f_1 = \frac{v}{\lambda_1} = \frac{v}{2L}$	1st	–
	$\lambda_2 = L$	$f_2 = \frac{v}{\lambda_2} = \frac{v}{L}$ $f_2 = 2f_1$	2nd	1st
	$\lambda_3 = \frac{2}{3}L$	$f_3 = \frac{v}{\lambda_3} = \frac{3v}{2L}$ $f_3 = 3f_1$	3rd	2nd

Generally the *nth* harmonic is given by: $f_n = n\left(\frac{v}{2L}\right)$, where n = 1, 2, 3, 4, ...

Note:

1. Fundamental Frequency and Harmonics

All resonant frequencies of vibrations that can be generated are called harmonics. The harmonics are numbered according to the ratio of their frequencies to the first natural frequency f_1. Thus if the 1st harmonic is 110 Hz, then 220 Hz, 330 Hz, 440 Hz etc. are the 2nd, 3rd , 4th etc harmonics respectively.

- The first natural frequency is the lowest possible frequency and is known as the *fundamental frequency* (or 1st harmonic)
- The frequencies of the harmonics are related to the fundamental frequency f_1 by $f_n = nf_1$ where f_n is the frequency of the n^{th} harmonic.
- Note that all harmonics are present for the string shown above.

2. Overtones

Overtones are the frequencies (excluding the fundamental) that can be produced by an instrument accompanying the 1st harmonic that is played.

The tuning fork and frequency generator are the only examples that can produce a pure note.

The lowest overtone is known as the 1st overtone, the next overtone is the 2nd overtone. Hence, in this case, the 1st overtone of the string corresponds to the 2nd harmonic of the string.

3. Speed of a wave on a string

The speed of a wave on a string is given by $v = \sqrt{\frac{T}{\mu}}$ where T is the tension in the string and μ is the mass per unit length of the string. **Not required for A Level, for information only.**

Example 10.8.2: Resonant Wavelengths of a Wire

A 1.0 m wire stretched between 2 points is plucked near one end. What are the three longest wavelengths present on the vibrating wire? [Ans: 2.0 m, 1.0 m, 0.67 m]

Solution:

10.8.3 Standing Waves in Air Column

Like standing transverse wave on a stretched string, a standing longitudinal wave is obtained by *the superposition of two trains of progressive longitudinal waves of equal amplitude, and frequency, travelling with the same speed in the opposite directions*.

Stationary transverse and longitudinal waves share similar properties except for the *direction of the displacement of the particles* where the displacement of air molecules are parallel to the direction of propagation of the component waves that form the stationary wave.

The formation of stationary waves in pipes and tubes is the basis of production of musical notes in most wind instruments. When stationary waves are set up in the tubes, **the air column resonates at its natural frequencies.**

The boundary conditions that determine the natural frequencies of the air column depends on whether that end is open or closed.

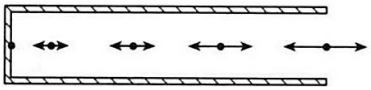


Fig. 10.8.5

Looking at Fig. 10.8.5,

At the closed end,

- the longitudinal movement of the air molecules is prevented and is always a **displacement node.**

At the open end,

- air molecules can vibrate freely, and is always a **displacement antinode**.

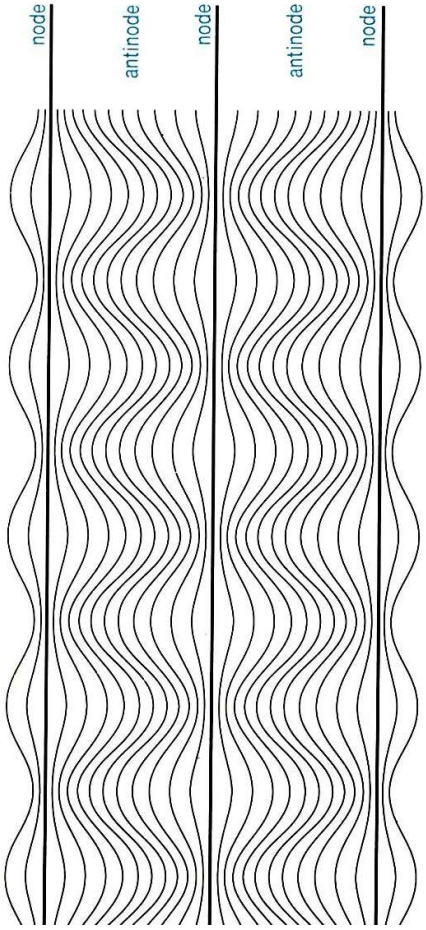


Fig. 10.8.6

The diagram can be used to show a longitudinal standing wave. To see the effect, move a piece of stiff card with a slot of no more than 1 mm cut in down over the diagram.

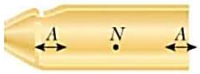
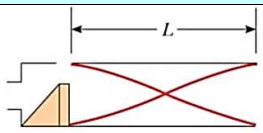
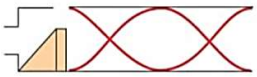
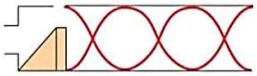
Resonant Frequencies of Air Columns

There are generally 2 types of air columns we want to concern ourselves with:

- open pipe
- closed pipe (closed at one end)

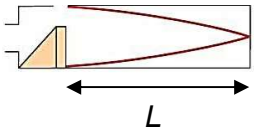
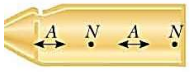
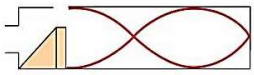
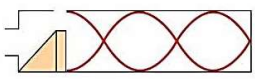
Recall that the open end of the pipe, we have a displacement antinode, and at the closed end of the pipe, we have a displacement node.

Resonant Vibrations in an Open Pipe

	(Displacement) Wave Profile	Wavelength	Frequency	Harmonic	Overtone
		$\lambda_1 = 2L$	$f_1 = \frac{v}{\lambda_1} = \frac{v}{2L}$	1st (fundamental)	–
		$\lambda_2 = L$	$f_2 = \frac{v}{\lambda_2} = \frac{v}{L}$ $f_2 = 2f_1$	2nd	1st
		$\lambda_3 = \frac{2}{3}L$	$f_3 = \frac{v}{\lambda_3} = \frac{3v}{2L}$ $f_3 = 3f_1$	3rd	2nd

The *nth* harmonic can be expressed as: $f = n\left(\frac{v}{2L}\right)$, where n = 1, 2, 3, 4, ...

Resonant Vibrations in a Closed Pipe

	(Displacement) Wave Profile	Wavelength	Frequency	Harmonic	Overtone
		$\frac{\lambda_1}{4} = L$ $\lambda_1 = 4L$	$f_1 = \frac{v}{\lambda_1} = \frac{v}{4L}$	1st (fundamental)	–
		$\frac{3\lambda_3}{4} = L$ $\lambda_3 = \frac{4L}{3}$	$f_3 = \frac{v}{\lambda_3} = \frac{3v}{4L}$ $f_3 = 3f_1$	3rd	1st
		$\frac{5\lambda_5}{4} = L$ $\lambda_5 = \frac{4L}{5}$	$f_5 = \frac{v}{\lambda_5} = \frac{5v}{4L}$ $f_5 = 5f_1$	5th	2nd

The *nth* harmonic is given by: $f_n = n\left(\frac{v}{4L}\right)$, where n = 1, 3, 5,...

Note:

1. The fundamental frequency of an open pipe is twice that of a closed pipe of the same length.
2. Only odd harmonics are present in closed pipe.
3. Assuming that we have a closed pipe and an open pipe with characteristics the same in all aspects inclusive of length, the following table presents a possible situation for the combination of frequencies emitted.

Harmonics	Frequency emitted from a Closed Pipe /Hz	Frequency emitted from an Open Pipe / Hz
1	256	512
2	-	1024
3	768	1536
4	-	2048
5	1280	2560

The quality of sound emitted is a mixture of these harmonics for each system. Hence we see that the "note" from an open pipe is "richer" than that from a closed pipe.

Example 10.8.3 **Serway p18.37.**

Calculate the length of a pipe that has a fundamental frequency of 240 Hz if the pipe is
(a) closed at one end and
(b) open at both ends.

Take the speed of sound to be 343 m s^{-1}. [Ans: (a) 0.357 m (b) 0.715 m]

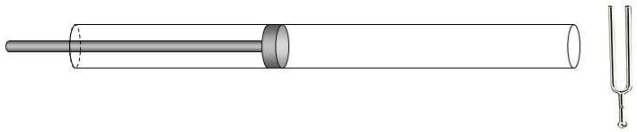
Solution:

Example 10.8.4 Serway P18.45.

An air column in a glass tube is open at one end and closed at the other by a movable piston. The air in the tube is warmed above room temperature, and a 384-Hz tuning fork is held at the open end. Resonance is first heard when the piston is 22.7 cm from the open end and again when it is 68.2 cm from the open end. [Ans: (a) 349 m s^{-1} (b) 1.14 m]

(a) What speed of sound is implied by these data?

(b) How far from the open end will the piston be when the next resonance is heard?

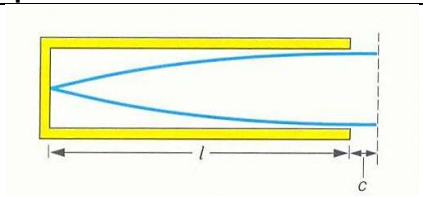
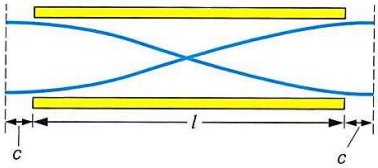


Solution:

End Corrections[5]

In practice, the air just outside the open end of a pipe is set into vibration and the displacement antinode of a standing wave occurs a distance *c* – called the **end correction** beyond the end of the tube.

The effective length of the resonant air column is larger than the length of the pipe. The following figure shows the respective corrected lengths for the fundamental frequencies of a closed and open pipe.

Closed Pipe	**Open Pipe**
	
Making Amendments for End Correction $\frac{\lambda}{4} = L + c$	Making Amendments for End Correction $\frac{\lambda}{2} = L + 2c$

- Note that the end correction for a cylindrical pipe is approximately 0.6*r* where *r* is the radius of the pipe. Hence, for the larger pipes, it is necessary to take end correction into consideration.

[5] The knowledge of end correction is not stipulated in the syllabus. But this can be found in many questions in A-levels especially in the earlier years.

10.8.4 What about the pressure?

Earlier in Waves topic, we discussed the pressure variations in a progressive sound wave; now we can consider how pressure varies in a stationary sound wave.

Fig. 10.8.7(a) shows the two extreme curves of the displacement-position graph for a stationary wave for times when the displacements are greatest. As before, we will take displacements of air molecules to the right to be positive.

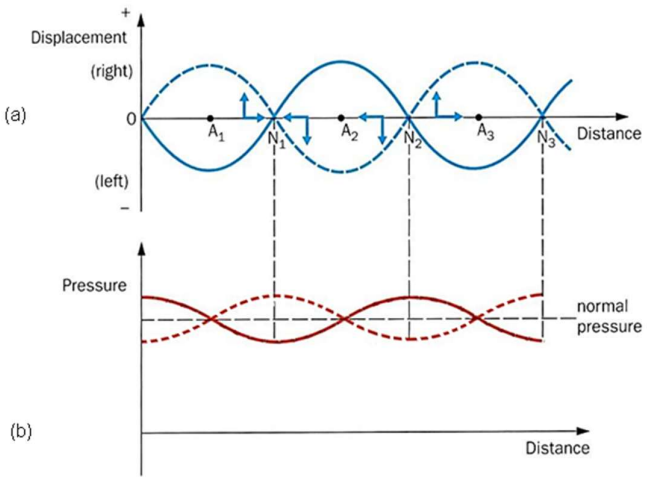


Fig. 10.8.7

Pressure variation for a stationary sound wave

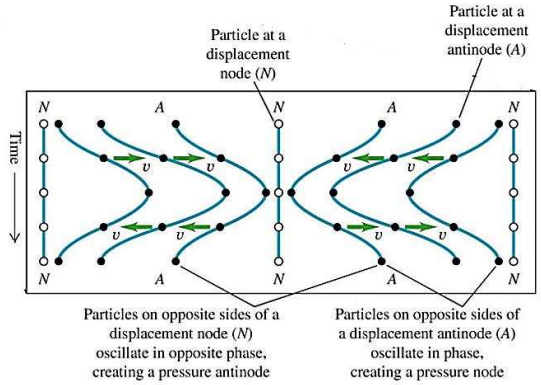


Fig. 10.8.8

In a standing longitudinal wave, a displacement node *N* is a pressure antinode (a point where the pressure fluctuates the most) and a displacement antinode A is a pressure node (a point where pressure does not fluctuate).

How does the pressure vary at displacement node N_1?

- Consider the instant in time when the wave corresponds to the dashed line in Fig. 10.8.7 (a).
- At position node N_1, air molecules to the left of N_1 are displaced positively and hence to the right towards N_1; air molecules to the right of N_1 has a negative displacement and hence to the left toward N_1 as well; air molecules at N_1 has zero displacement and remain at their equilibrium positions. **At this instant in time, the air at N_1 is compressed and the pressure is greater than normal.**
- Half a period later, the wave would correspond to the solid line in Fig. 10.8.7 (a),
- At position node N_1, air molecules to the left of N_1 are displaced negatively and hence to the left away from N_1; air molecules to the right of N_1 has a positive displacement and hence to the right away from N_1 as well. Air molecules at N_1 has zero displacement and remain at their equilibrium positions. **At this instant in time, the air at N_1 is rarefied and the pressure is lower than normal.**
- This illustrates the fact that **pressure variation is greatest at displacement nodes.**
- **Displacement node** is subjected to large pressure changes and hence it is a **pressure antinode.**

How does the pressure vary at displacement antinode A_1?

- At position antinode A_1, air molecules to the left of A_1, to the right of A1 and at A1 are always displaced in the same direction. This means that at antinode A_1, the separation between air molecules is maintained constant and **the pressure remains at normal all the time**.
- **Displacement antinode** is subjected to zero pressure changes and hence it is a **pressure node.**

Example 10.8.5: End Corrections

A source of sound of frequency 250 Hz is used with resonant tube, closed at one end, to measure the speed of sound in air. The two shortest resonant lengths are found to be 0.30 and 0.96 m. Calculate

(a) the speed of the sound, and

(b) the end correction of the tube. [Ans: (a) 330 m s^{-1} (b) 3.0 cm]

Solution:

10.8.5 Investigating the speed of sound using stationary waves

We can determine the speed of sound wave by setting up a stationary wave and measure the corresponding frequency and wavelength as described by the experiment below. A stationary sound wave is set up by reflection from a reflector.

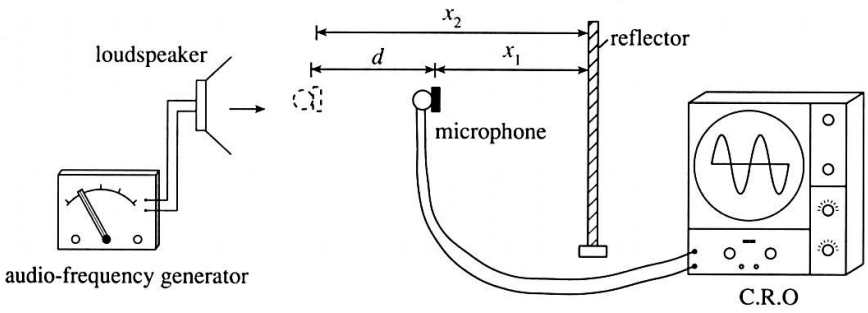


Fig. 10.8.9 Measurement of speed of sound

The experimental setup is shown in Fig. 10.8.9.

- The signal generator (See Appendix E, Pg. 49) connected to a speaker delivers a note of frequency *f* towards a reflector (e.g. a metal sheet).
- The sound wave gets reflected off the reflector and interference occurs between the forward and reflected waves and a stationary wave pattern with nodes and antinodes is established between the speaker and reflector.
- The positions of antinodes are determined using a microphone connected to the oscilloscope (See Appendix A of waves lecture notes for the write-up on oscilloscopes). The microphone is slowly moved away from the reflector until a trace of maximum amplitude is obtained. This denotes the position of a pressure antinode. The distance x_1 of the microphone from the reflector is measured using a metre rule.
- The microphone is next moved slowly further away from the reflector until the amplitude of the trace on the C.R.O. is once again maximum. The distance of the microphone from the reflector, x_2, is again measured.

  Therefore, Distance between 2 successive pressure antinodes, $d = x_2 - x_1 = \lambda/2$

  Wavelength of the sound wave, $\lambda = 2d$

- If the frequency *f* of the sound wave is known, the speed of sound *v* can be determined from

$$v = f\lambda$$

➢ **You may now complete Tutorial 10D on Stationary Waves.**

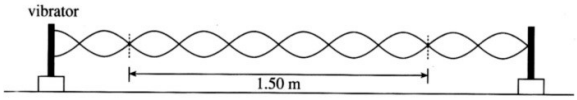
Tutorial 10D: Stationary Waves

Self-Practice Questions

S1 J84/II/12. Which one of the following correctly compares characteristics of travelling and stationary plane waves?

	Travelling Wave	*Stationary wave*
A	no medium required	requires a material medium
B	separation between two adjacent points of corresponding phase is one wavelength	separation between a node and the adjacent antinode is half a wavelength
C	the amplitude of vibration is the same at all points	amplitude of vibration varies with position
D	energy at any point is always kinetic	energy at any point changes from kinetic and potential and back again.
E	energy is transported at a speed given by the frequency divided by the wavelength	No net transport of energy

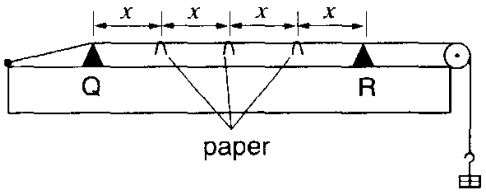
S2 A vibrator produces a stationary wave on a stretched string as shown in the figure below. If the frequency of the vibrator is 20 Hz, what is the speed of the wave?



A 3.0 m s^{-1} **B** 10.0 m s^{-1} **C** 15.0 m s^{-1} **D** 22.5 m s^{-1}

E 43.0 m s^{-1}

S3 J97/I/12. A wire is stretched over two supports, Q and R, a distance 4*x* apart. Three light pieces of paper rest on the wire, as shown.



When the wire is made to vibrate at one particular frequency, the middle piece of paper stays on, but the others fall off the wire.

What is the wavelength of the vibration produced on the wire.

A 2*x* **B** 3*x* **C** 4*x* **D** 8*x*

S4 A tall vertical cylinder is filled with water and a tuning fork of frequency 512 Hz is held over its open end. Calculate the position of the water level below the open end when (a) the first resonance and (b) second resonance are heard. (Assume speed of sound in air is 330 m s^{-1})

S5 N84/II/12.

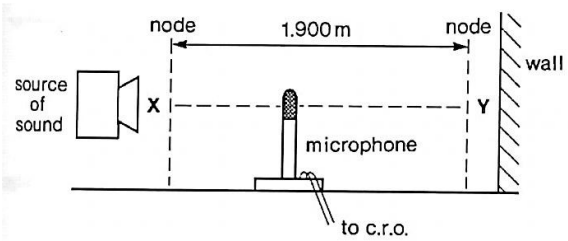
A stationary wave in the gas in a resonance tube can be described in terms either of amplitude Δx of oscillation of the particles of the gas from their mean positions or of the fluctuation of pressure Δp above and below the average. Which one of the following correctly describes the situation at resonance in a tube which is closed at one end? (Neglect end correction at the open end.)

	At closed end		At open end	
	Δx	Δp	Δx	Δp
A	zero	maximum	zero	maximum
B	zero	maximum	maximum	zero
C	maximum	zero	maximum	zero
D	maximum	zero	zero	maximum
E	zero	zero	maximum	maximum

Discussion Questions:

D1 N89/I/7; J95/I/10

A source of sound of frequency 2500 Hz is placed several metres from a plane reflecting wall in a large chamber containing a gas. A microphone, connected to a cathode-ray oscilloscope, is used to detect nodes and antinodes along the line **XY** between the source and the wall.



The microphone is moved from one node through 20 antinodes to another node, a distance of 1.900 m.

What is the speed of sound in the gas?

A 238 m s^{-1} **B** 250 m s^{-1} **C** 330 m s^{-1} **D** 475 m s^{-1}

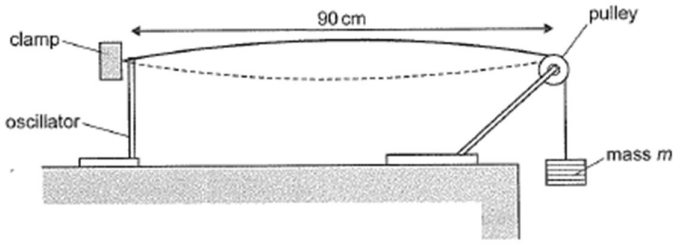
D2 N79/II/14; J85/I/12.

A suspension bridge is to be built across a valley where it is known that the wind can gust at 5 s intervals. It is estimated that the speed of transverse waves along the span of the bridge would be 400 m s^{-1}. The danger of resonant motions in the bridge at its fundamental frequency would be greatest if the span had a length of

A 2 000 m **B** 1 000 m **C** 400 m **D** 80 m **E** 40 m

D3 N2017/I/17

The speed of a wave on a stretched string is proportional to the square root of the tension in the string. The string is held under tension between a clamp and a pulley 90 cm apart by a mass *m* of 400 g hanging from the end of the string.



When the frequency *f* of the mechanical oscillator close to the clamped end is 15 Hz, the stationary wave shown is set up between the clamp and the pulley. The hanging mass *m* and frequency *f* are varied until another stationary wave is formed.

Which combination of *m* and *f* gives a stationary wave?

	m/g	*f*/Hz
A	500	30
B	600	40
C	800	60
D	900	90

D4 J99/III/4 (part).

(a) One end of a horizontal string is attached to the oscillating plate. The string passes over a pulley and the string is kept under tension by means of a weight, as illustrated in the figure below.

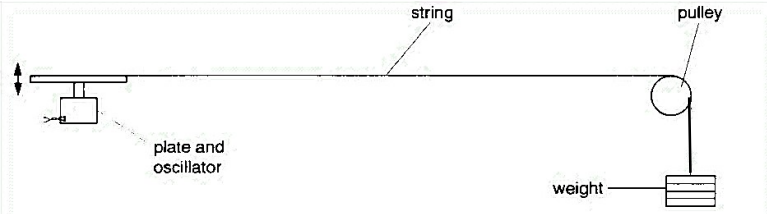


Fig. 4.2

The frequency of oscillation of the plate is increased and at certain frequencies, stationary waves are produced on the string.

(i) Copy the above figure and on your diagram show the stationary wave on the string when the frequency is such that the distance between the plate and the pulley corresponds to two wavelengths of the wave on the string.

(ii) On your diagram, label the position of a node on the string.

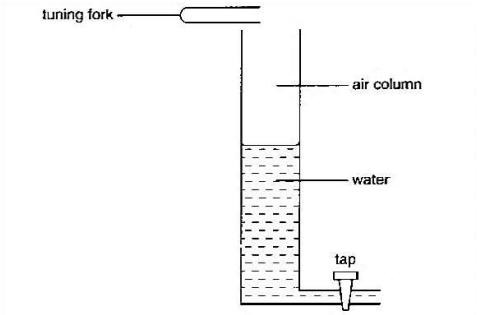
(iii) Briefly explain why a stationary wave is observed on the string only at particular frequencies of vibration of the plate. [4]

(b) Some musical instruments rely on stationary waves on strings in order to produce sound.

Suggest why strings made of different materials or with different diameters are sometimes used. [2]

D5 N2000/I/10

The diagram shows an experiment to produce a stationary wave in an air column. A tuning fork, placed above the column, vibrates and produces a sound wave. The length of the air column can be varied by altering the volume of the water in the tube.



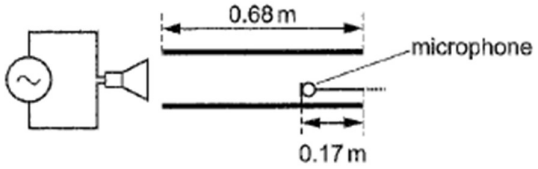
The tube is filled and then water is allowed to run out of it. The first two resonances occur when the air column lengths are 0.14 m and 0.46 m.

What is the wavelength of the sound wave?

A 0.32 m **B** 0.56 m **C** 0.60 m **D** 0.64 m

D6 N2018/I/18

A stationary sound wave is formed inside an open tube of length 0.68 m. A small micophone is inserted into the tube. It detects the first node at a distance of 0.17 m from the end. The microphone is then fixed in this position. The speed of sound in the tube is 340 m s^{-1}.

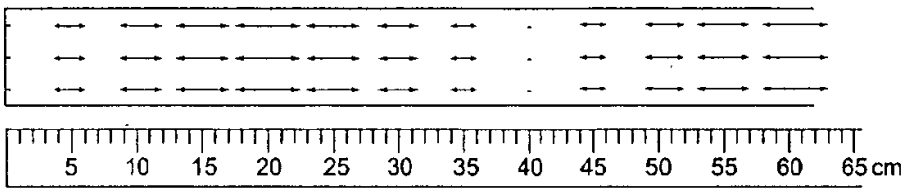


The frequency of the signal generator is now increased until the microphone again detects a node. What is this new frequency?

A 750 Hz B 1000 Hz C 1500 Hz D 3000 Hz

D7 N2004/I/18.

A stationary sound wave is set up in a pipe using an oscillator of frequency 440 Hz. The extent of the vibration of the molecules in the pipe is illustrated. A centimetre scale is shown alongside the pipe.



What is the speed of sound in the pipe?

A 176 m s^{-1} **B** 328 m s^{-1} **C** 337 m s^{-1} **D** 352 m s^{-1}

D8 (Jun 07 P2/5) Light reflected from the surface of smooth water may be described as a polarised transverse wave.

(a) By reference to the direction of propagation of energy, explain what is meant by
(i) a *transverse wave,* [1]
(ii) *polarisation.* [1]

(b) A glass tube, closed at one end, has fine dust sprinkled along its length. A sound source is placed near the open end of the tube, as shown in Fig. 5.1.

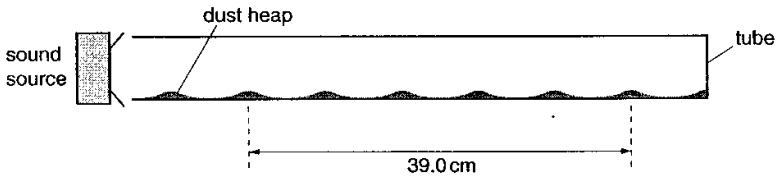


Fig. 5.1

The frequency of the sound emitted by the source is varied and, at one frequency, the dust forms small heaps in the tube.

(i) Explain, by reference to the properties of stationary waves, why the heaps of dust are formed.[3]

(ii) One frequency at which heaps are formed is 2.14 kHz. The distance between six heaps, as shown in Fig. 5.1, is 39.0 cm. Calculate the speed of sound in the tube. [3]

(c) The wave in the tube is a stationary wave. Explain, by reference to the formation of a stationary wave, what is meant by the speed calculated in **(b)(ii)**. [3]

Numerical Answers to Selected Questions in Tutorial 10D: Stationary Waves
S4(a) 0.161 m (b) 0.483 m
D8 (b)(ii) 334 m s^{-1}

Appendices

Appendix A: Huygens' Principle

But how can one explain the bending of waves due to diffraction? Why does the spreading of waves become increasingly more obvious with the width of the aperture decreases to about the same order as the wavelength? This phenomenon may be explained using *Huygens' Principle.*

The *Huygens' Principle* is a geometrical method for finding, from the known shape of a wave front at some instant, the shape of the wave front at some later time. Huygen assumed that

> ***every point of a wave front may be considered the source of secondary wavelets that spread out in all directions with a speed equal to the speed of propagation of the wave.***

The new wave front at a later time is found by constructing a surface *tangent to the secondary wavelets,* called the *envelope* of the wavelets.

The following diagram shows how Huygens' construction is used to determine the new wavefronts for a plane wave and for a spherical wave.

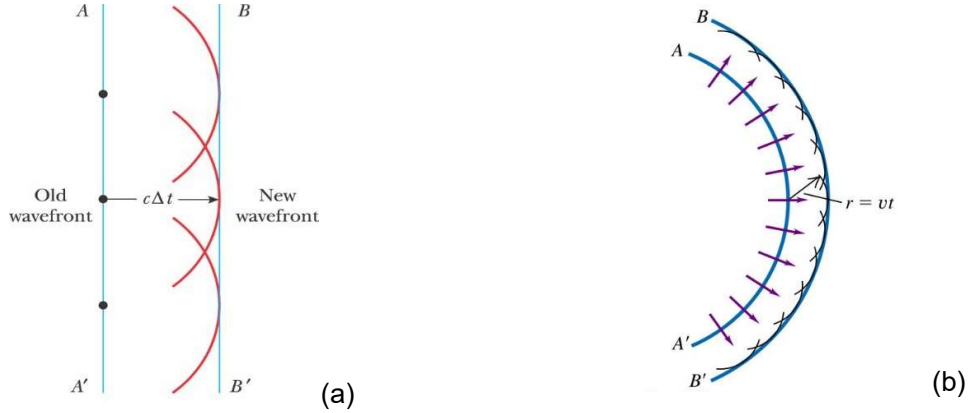


Fig. 1: Huygens' construction for (a) a plane wave propagating to the right and (b) a spherical wave propagating to the left.

Through Huygens' construction, we can now see how the subsequent wavefronts 'reconstruct' themselves when incident plane waves pass through a slit.

Let us now apply Huygens' construction to explaining diffraction.

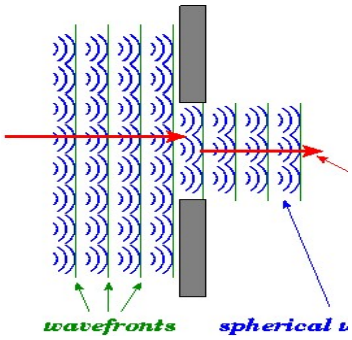


Fig. 2

Consider first the case where the obstacle or opening is much larger than the wavelength of the waves (Fig. 2), we find that the wavelets emerging forms wavefronts that is essentially straight. Diffraction is not very apparent.

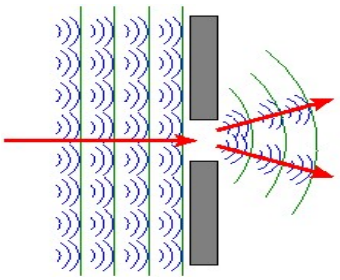


Fig. 3

If, however, the size of the opening becomes comparable to the wavelength, the waves proceed to "bend through" or around the opening or obstruction as shown in Fig. 3.

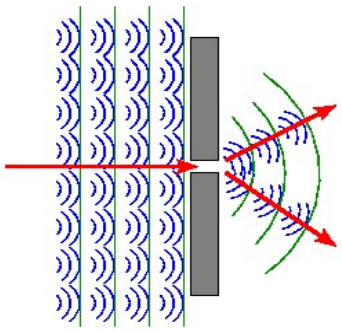
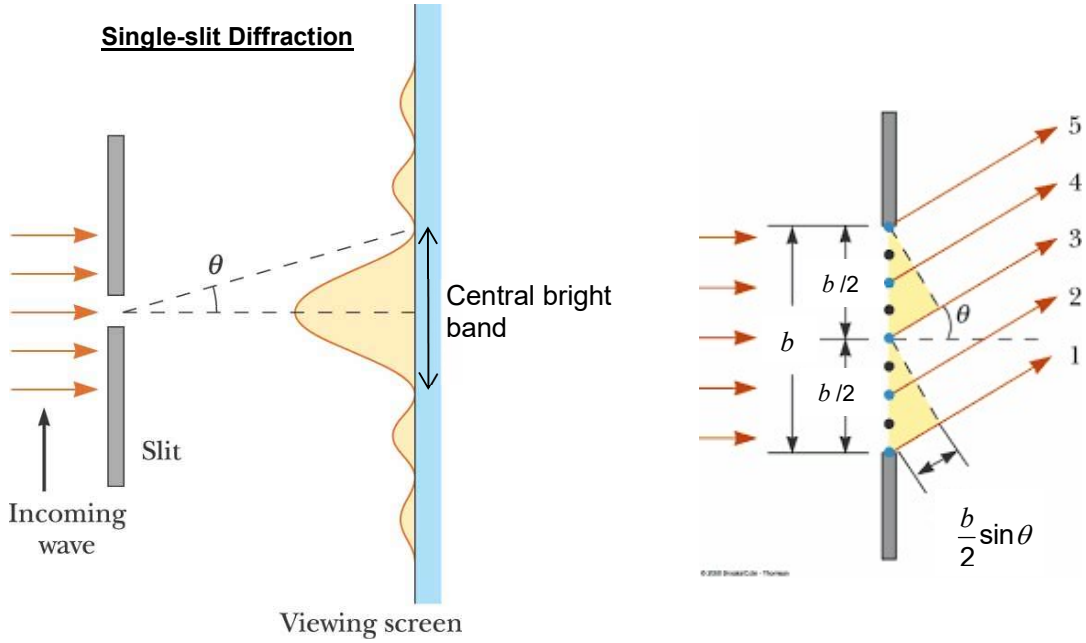


Fig. 4

Finally, as the obstruction or opening equals the wavelength, the diffraction becomes quite pronounced as shown in Fig 4.

Appendix B: Single Slit Diffraction Pattern

To analyze the diffraction pattern of the single slit, let us divide the slit into two halves as shown.



Consider ray 1 and 3. As the two rays travel towards the screen, ray 1 travels an additional path difference of $\frac{b}{2}\sin\theta$ where b is the width of the slit and θ is the angle with the principle axis. Similarly, the path difference between rays 2 and 4, as well as between rays 3 and 5, is also $\frac{b}{2}\sin\theta$.

If this path difference is exactly half a wavelength, then destructive interference occurs. This cancellation occurs for any two rays that originate at the points separated by half a slit width because the phase difference between 2 such points is 180°. Therefore the waves from the upper half of the slit interfere destructively with the waves from the lower half when

$$\frac{b}{2}\sin\theta = \pm\frac{\lambda}{2}$$
$$\text{or } b\sin\theta = \pm\lambda$$

Dividing the slit into 4 equal parts and using the same reasoning, we find that destructive interference also occurs when $\frac{b}{4}\sin\theta = \pm\frac{\lambda}{2}$ or $b\sin\theta = \pm 2\lambda$.

This same argument can be applied when the slit is divided into 6, 8 and so on equal parts, and we obtained the general condition for **destructive interference** (or **dark region**) to be

$$b\sin\theta = m\lambda \text{ where m = } \pm 1, 2, 3, \ldots$$

Appendix C: Ripple Tank

One way to observe interference of water waves is in a ripple tank. Generally a ripple tank could be used to investigate various properties of water waves.

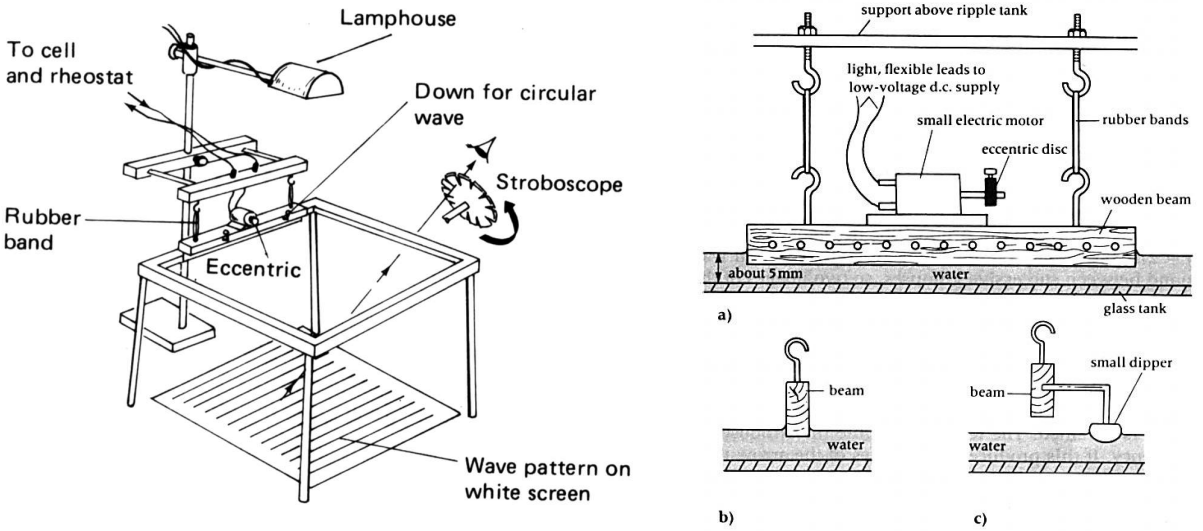


Fig. 1 Set-up for the ripple tank

1. **Eccentric:** An eccentric (off-centre) metal disc on the axle of the motor causes the beam to vibrate at the same frequency as the revolutions of the motor. The speed of the motor is controlled by a rheostat in series with a low-voltage d.c. supply, or a dry battery (usually 3.0 or 4.5 V).

2. To generate continuous straight wave, a vibrating beam just touching the water is used.

3. To generate circular waves, a small spherical dipper is fixed to the beam instead.

4. The speed of the motor may be adjusted to vary the frequency and wavelength of continuous wave train.

5. The progressive waves move quite quickly across the water surface and any attempt to measure the wavelength or draw the wave diagrams from the moving water shadows would be very difficult. A standard method of freezing or slowing down rapid motion would be to use a stroboscope.

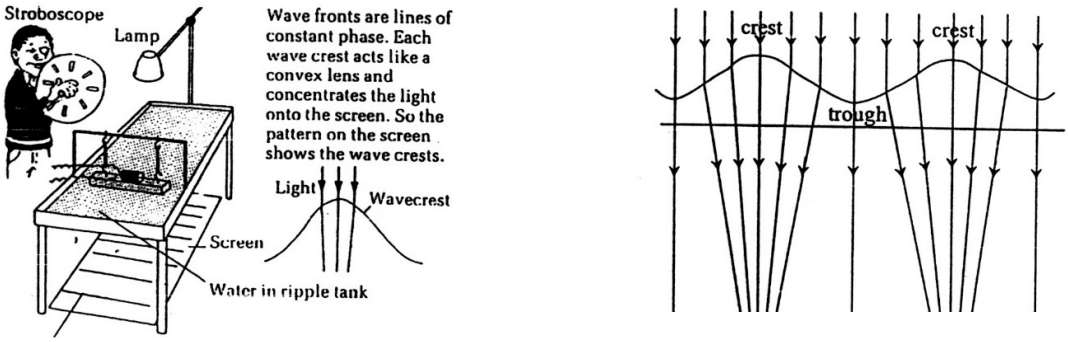


Fig. 2 Observations of wavefronts in a ripple tank.

Appendix D: Spectrometer

The grating or double slit is usually mounted on a device known as a *spectrometer* to allow us to easily take measurements of θ so that we can study about the light source. Fig. 1 below shows you a typical laboratory spectrometer and Fig. 2 shows a schematic sketch of the spectrometer.

Fig. 1
A laboratory spectrometer

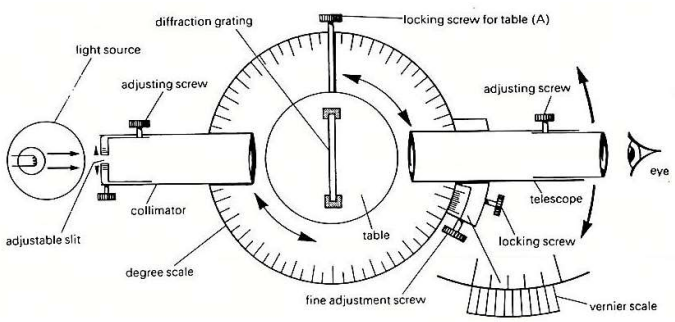


Fig. 2
The schematics of the spectrometer

The essential parts of a spectrometer are

(a) A fixed *collimator* with a movable slit of adjustable width to produce a parallel beam of light from the source illuminating the slit. The light source to be studied is placed between the collimator and the turntable.

(b) a *turntable* on which the prism, double slit or grating can be mounted, and

(c) a *telescope* attached to the turntable such that it can be rotated about same vertical axis as the turntable. Usually there is also vernier scale attached below the turntable so that you to read off the *angle* though which the telescope is turned through from this scale. The vernier scale allows you to read the angular position of the telescope to the nearest 0.1°.

When used with a grating,

- we usually start off by locating the position of the *zeroth order maximum*. It is the brightest fringe that has the brightest intensity. This provides a reference point for which angular displacement can be measured. The angular displacement of the m^{th} *order* maximum,

  θ = |Angular position of zeroth order bright fringe - Angular position of m^{th} order bright fringe|

  $$\Rightarrow \qquad \theta = \left|\theta_o - \theta_m\right|$$

  Alternatively, by noting the angular positions of the *mth* order bright fringe at either side of the central maximum ($\theta_{m,left}$ and $\theta_{m,right}$), we can also determine the angular displacement θ of m^{th} order maximum from the central maximum. In this case, $2\theta = \left|\theta_{m,left} - \theta_{m,right}\right|$

- We also move the telescope to each side of this central maximum to check that several orders of *diffraction* can be seen and there is some symmetry about this central maximum. If not, the plane of the surface of the grating is not perpendicular to the plane in which telescope moves and tilt of the table needs to be adjusted.

Chapter 11

Thermal Physics

Based on a thermoscope invented by Galileo Galilei in the early 1600s, a Galileo thermometer is a simple, fairly accurate thermometer, today it is mostly used as decoration. The Galileo thermometer consists of a sealed glass tube that is filled with water and several floating bubbles. The bubbles are glass spheres filled with varying amounts of a colored liquid mixture.

Playlist of lecture examples:

https://youtube.com/playlist?list=PL_b5cjrUKDlahSMbhCQQxK5vpqGuwKte2

H2 Physics 9749 Syllabus

Temperature and Ideal Gases

Content

- Thermal equilibrium
- Temperature scales
- Equation of state
- Kinetic theory of gases
- Kinetic energy of a molecule

Learning Outcomes

Candidates should be able to:

(a) show an understanding that regions of equal temperature are in thermal equilibrium

(b) explain how empirical evidence leads to the gas laws and to the idea of an absolute scale of temperature (i.e. the thermodynamic scale that is independent of the property of any particular substance and has an absolute zero)

(c) convert temperatures measured in kelvin to degrees Celsius: T / K = T / °C + 273.15

(d) recall and use the equation of state for an ideal gas expressed as $pV = nRT$, where n is the amount of gas in moles

(e) state that one mole of any substance contains 6.02×10^{23} particles and use the Avogadro number $N_A = 6.02 \times 10^{23}$ mol^{-1}

(f) state the basic assumptions of the kinetic theory of gases

(g) explain how molecular movement causes the pressure exerted by a gas and hence derive the relationship $pV = (1/3)\,Nm\langle c^2\rangle$, where N is the number of gas molecules (a simple model considering one-dimensional collisions and then extending to three dimensions using $1/3\,\langle c^2\rangle = \langle c_x^2\rangle$ is sufficient)

(h) recall and apply the relationship that the mean kinetic energy of a molecule of an ideal gas is proportional to the thermodynamic temperature (i.e. $\frac{1}{2}\,m\langle c^2\rangle = (3/2)\,kT$) to new situations or to solve related problems

First Law of Thermodynamics

Content

- Specific heat capacity and specific latent heat
- Internal energy
- First law of thermodynamics

Learning Outcomes

Candidates should be able to:

(a) define and use the concepts of specific heat capacity and specific latent heat

(b) show an understanding that internal energy is determined by the state of the system and that it can be expressed as the sum of a random distribution of kinetic and potential energies associated with the molecules of a system

(c) relate a rise in temperature of a body to an increase in its internal energy

(d) recall and use the first law of thermodynamics expressed in terms of the increase in internal energy, the heat supplied to the system and the work done on the system

Contents

Independent Study Guide 4
Introduction to Thermal Physics 5
1 Temperature 6
1.1 Thermal Equilibrium & the Concept of Temperature 6
1.2 Temperature Measurement & Temperature Scales 7
1.2.1 From an Empirical Scale to an Absolute Scale 7
1.2.2 Other Absolute Scales 8
2 Specific Heat Capacity and Internal Energy 9
2.1 Heat Transfer 9
2.2 Conservation of Energy 10
2.3 Microscopic View of Energy 12
Tutorial 11A 14
3 Ideal Gas 21
3.1 A Really Huge Number – The Avogadro Number, N_A 21
3.2 Molecular Interpretation of Pressure 23
3.2.1 Derivation of $\boldsymbol{pV = \frac{1}{3} Nm < c >^2}$ 24
3.3 Molecular Interpretation of Temperature 27
Tutorial 11B 29
4 State of A System 34
4.1 First Law of Thermodynamics 34
4.2 Four Basic Thermodynamic Processes 36
4.3 Cyclic Process 37
Tutorial 11C 39
Appendix A: Temperature Measurement & Temperature Scales 45
Appendix B: Empirical Scale 46
Appendix C: Energy and Phases of Matter 47

Independent Study Guide

O Level Physics

- Temperature: Principles of thermometry
- Thermal Properties of Matter
  - Internal energy
  - Specific heat capacity
  - Melting, boiling & evaporation
  - Specific latent heat
- Kinetic Model of Matter
  - States of matter
  - Brownian motion
  - Kinetic Model

② **To be covered in C2 in 2024**

Read Chapter 11 Thermal Physics Lecture Notes (p. 21 - 28)

- Equation of State of an ideal gas
- Kinetic theory of gases
- Kinetic energy of a molecule

Attempt Tutorial 11B (p. 29 - 33)

Read Chapter 11 Thermal Physics Lecture Notes (p. 34 - 38)

- First law of thermodynamics
- Thermodynamic processes

Attempt Tutorial 11C (p. 39 - 44)

① **Year-End Self-Review & Independent Study (Recommended 5 hrs)**

Read Chapter 11 Thermal Physics Lecture Notes (p.1 – 13)

- Temperature & Thermal Equilibrium
- Temperature scales (Thermometry)
- Specific heat capacity & specific latent heat
- Internal energy

Attempt Tutorial 11A (p.14 -20)

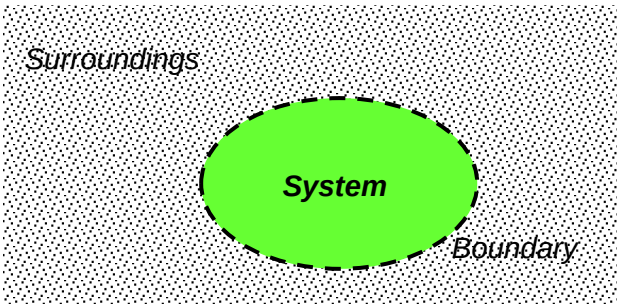
A-level H2 Physics (9749) – Thermal Physics

Temperature	Ideal Gases	First Law of Thermodynamics
Thermal equilibrium ①	Equation of State ②	Specific heat capacity & specific latent heat ①
Temperature scales ①	Kinetic theory of gases ②	Internal energy ①
	Kinetic energy of a molecule ②	First law of thermodynamics ②

Introduction to Thermal Physics

Defining Our System

When we are studying a physical phenomenon, the first thing we can do is to do some kind of 'isolation' i.e. we 'pluck' some part of that thing out from its surroundings. This 'plucked out' part is what we commonly call the 'system'. The surface that separates the system from its surroundings is called the boundary. It is important to be strict about the boundary between the system and its surroundings i.e. we should not be sloppy or hazy when it comes to defining the system boundary at the outset of our study.



Thermodynamics Properties of A System – Macroscopic vs Microscopic

After we have defined our system, the next thing is of course to decide which aspect(s) of the system we wish to observe. In science, people are keen on those aspects that can be studied in the laboratory. Furthermore, we try to quantify such aspects i.e. ascribe numbers to them. In Thermodynamics, some of the aspects of the system we usually deal with are: length, volume, density, energy, pressure, temperature, mass etc. We call them properties or physical quantities of the system.

Such properties are separated into two categories: macroscopic and microscopic. Taking energy for example: macroscopic forms of energy are those in which a system possesses as a whole with respect to some outside reference frame, such as kinetic and potential energies of the centre-of-mass of the system. The microscopic forms of energy, on the other hand, are those related to the molecular structure of the system and the degree of molecular activity. They are independent of outside reference frames. The sum of all the microscopic forms of energy of a system is its *internal energy*.

State of A System – A Set of Its Properties

If a system is not undergoing any change, the set of all the properties of the system that can be measured or calculated will specify the condition of this system. We call this the ***state*** of the system. At a given state, all the properties of a system have fixed values. If the value of even one property changes, the state will change to a different one. The state of a system is hence a condition of the system specified in terms of the totality of its properties.

1 Temperature

1.1 Thermal Equilibrium & the Concept of Temperature

Equilibrium
The word equilibrium implies a state of balance. In an equilibrium state, there is no unbalanced driving 'force' to disturb the system. There are many types of equilibrium e.g. in mechanics, when we say a system is at translational equilibrium, we mean that there is no net force on this system.

We will introduce another type of equilibrium – thermal equilibrium.
Two objects in thermal contact are said to be in thermal equilibrium when there is ***no net exchange of heat*** between them.

Note that 'heat' is energy in transit and not the property of an object, it is inappropriate to make statements like 'this object has more heat than the other object' etc. It is more appropriate to say this object has more thermal energy than the other object. Heat transfer between two objects can be done via conduction, convection and radiation. In the last case, the two objects need not be in physical contact.

Introducing Temperature
How do we know if any heat transfer has occurred between the two objects? We need some kind of parameter to characterise each object. That is to say that

i) If the two objects are separately characterised by the same value of the parameter, then these two objects will be in thermal equilibrium when brought into thermal contact with each other.

ii) If they are, on the other hand, characterised by different values of the parameter, then they will not be in thermal equilibrium when brought into thermal contact with each other.

So what is this parameter that characterises each object? – Temperature! If two objects are in thermal equilibrium, they have the same temperature. So heat is defined as the ***transfer of energy*** between two objects due to a difference in temperature between them.

Now suppose we have three objects A, B and C. If A and C remain in thermal equilibrium when brought into thermal contact, we know that $T_A = T_C$. Similarly if B and C also remain in thermal equilibrium when brought into thermal contact, $T_B = T_C$. Hence we can conclude that $T_A = T_B$ and be confident that A and B will also remain in thermal equilibrium when brought into thermal contact.

This idea of thermal equilibrium (even without the concept of temperature) is so fundamental in thermodynamics that it is placed as the most basic law in thermodynamics:

Zeroth Law of Thermodynamics
If bodies A and B are ***separately in thermal equilibrium*** with a third body C, then A and B are also in thermal equilibrium with each other.

Since temperature is the parameter that will indicate if a system is in thermal equilibrium with another system, it is worth measuring. The zeroth law makes possible the use of test systems commonly called 'thermometers', which allow measurements to decide whether any two systems will or will not remain in thermal equilibrium when brought into thermal contact with each other.

1.2 Temperature Measurement & Temperature Scales

1.2.1 From an Empirical Scale[1] to an Absolute Scale

The Problem with the Empirical Scale

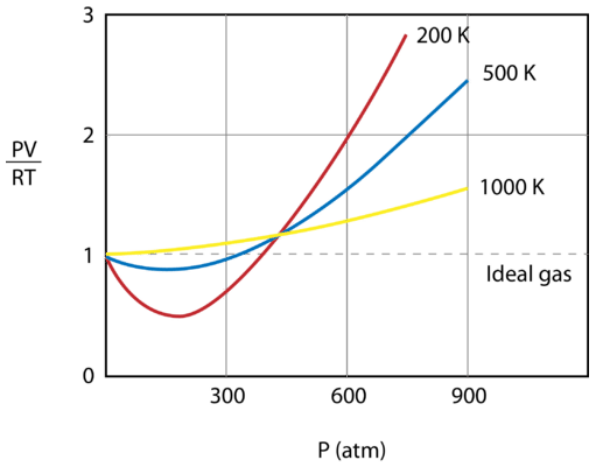
Consider this scenario: two different thermometers working on different thermometric properties[2] X and Y are calibrated on the same fixed points – melting ice and boiling water.

Say we assign the number 0 and 100 at these two fixed points for the two thermometers, so we have an empirical Centigrade Scale. Now, so long as the two thermometers measure the temperature of melting ice and boiling water we will get perfect agreement in temperature readout from these two thermometers but what if we use them to measure other temperatures? Clearly, the two different thermometric properties X and Y do not have to vary with temperature linearly and certainly not in the same way.

Ideal Gas Scale

This disagreement between thermometers using different thermometric properties is clearly a disadvantage of the empirical scale. One way of solving this is to take one scale as a standard in which all temperatures are expressed. This specially chosen standard scale must be independent of the thermometric property used hence such a scale is sometimes known as an **absolute scale**.

An ideal gas is one that follows the gas laws at all conditions of temperature and pressure. To do so, the gas needs to completely abide by the kinetic-molecular theory. There is no such thing as an ideal gas in real life. A real gas deviates most from an ideal gas at low temperatures and high pressures. Gases are most ideal at high temperature and low pressure. The figure below shows a graph of PV/RT plotted against pressure for 1 mol of a gas at three different temperatures of 200 K, 500 K, and 1000 K. The magnitude of the deviations from ideality is greatest for the gas at 200 K and least for the gas at 1000 K.

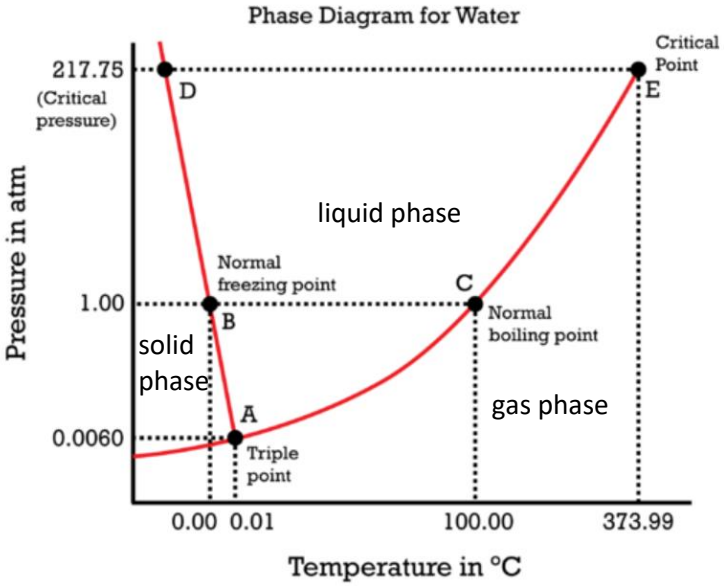


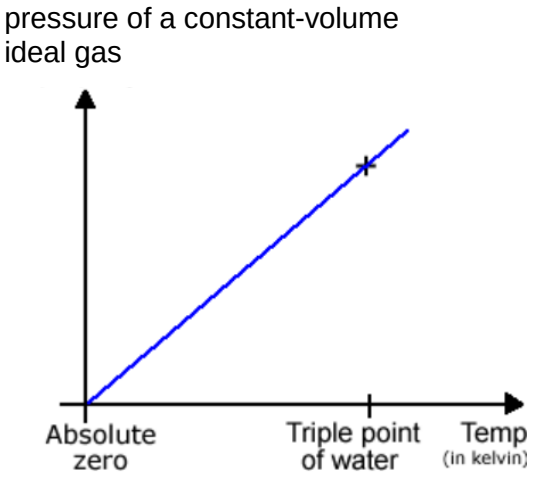
One way to minimize disagreement between thermometers is to use the gas under conditions when it behaves approximately like an ideal gas where the ideal gas equation of state ($PV = nRT$) holds. In this case, the volume of a fixed mass of gas V will be directly proportional to the absolute temperature T if the gas pressure P is kept constant. This relationship works for any gas under ideal conditions (hence independent of the type of gas) which in theory gives the most undisputed temperature measurement. Temperature measurement here is read in the **Ideal Gas Scale**.

[1] See Appendix B: Empirical Scale

[2] See Appendix A: Temperature Measurement and Temperature Scales

The unit for this scale is kelvin. The symbol is K. The reference temperature (fixed point) is set at zero kelvin (0 K) – the ***absolute zero***. This is a theoretical temperature as it has not been attained experimentally. How about another fixed point? The state in which ice, liquid water, and water vapour coexist in equilibrium – a state known as the triple point of water was chosen as this temperature can be accurately and reproducibly measured. It was assigned the value 273.16 K.





So the interval between the two fixed points is 273.16 K. If you remember the basic SI unit for temperature, does this strike a bell? The kelvin, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.

1.2.2 Other Absolute Scales

The ideal gas scale represents a big advancement as it relies on the universal property of any gas under ideal behaviour (rather than on a particular gas). It is hence an absolute scale.

Thermodynamic Scale
Another absolute scale known is the **Thermodynamic Scale**. It is based on the theoretical efficiency of a perfectly reversible heat engine. This engine can work on any working fluid hence this is the foundation for an absolute scale (as opposed to an empirical scale).

The thermodynamic scale is equivalent to the ideal gas scale.

Celsius Scale
If the temperature of ice is 0^{o} on the Centigrade Scale (most of us roughly have an idea how cold it is), then on the thermodynamic scale, it would be 273.15 K which is a totally unfamiliar number to most people. To make the numbers on the thermodynamic scale resemble those in the empirical scale more familiar to us, an arithmetical adjustment is made to it and the **Celsius Scale** is created. Thus the Celsius Scale is actually a "displaced" Thermodynamic Scale made to look like the familiar Centigrade Scale. Note that Celsius is an absolute scale designed to produce the temperature readout like the (familiar) empirical Centigrade Scale.

To convert temperature measured in degree Celsius (*Celsius Scale*) to kelvin (*Thermodynamic Scale*):

$$T / \mathrm{K} = t / {}^{o}\mathrm{C} + 273.15$$

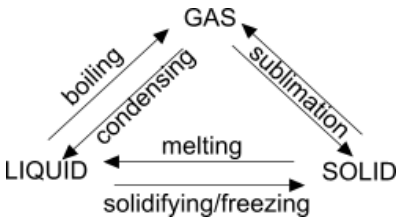
2 Specific Heat Capacity and Internal Energy

2.1 Heat Transfer

We have talked about the concept of heat transfer between two systems being driven by the temperature difference between them. In this section, we will focus on quantifying this amount of heat. Heat transfer can potentially affect the state of matter. We know that matter normally exists as either a solid, a liquid or a gas.

The formula to quantify the amount of heat depends on:
i) the physical state (or phase) of the object, and
ii) whether the object undergoes a change in phase.

Below shows changes in phase without a change in temperature.



Decide which object to focus on

At the end of the heat transfer, does the object undergo any change in phase?

No

Heat transfer
= (mass)(specific heat capacity)(temp. change)

$$\boldsymbol{Q = m\,c\,(\Delta T)}$$

The **specific heat capacity** of a substance c is the energy per unit mass required to raise the temperature of the substance by one unit of temperature.
SI unit : J kg^{-1} K^{-1}

Or

Heat transfer
= (heat capacity)(temp. change)

$$\boldsymbol{Q = C\,(\Delta T)}$$

The **heat capacity** of an object C is the amount of energy that is required to raise the temperature of the object by one unit of temperature. SI unit: J K^{-1}

$T_{final} > T_{initial} \Rightarrow$ heat absorbed
$T_{final} < T_{initial} \Rightarrow$ heat lost

Yes

Heat transfer
= (mass) (specific latent heat)

$$\boldsymbol{Q = mL}$$

The **specific latent heat of fusion** L_f is defined as the <u>energy per unit mass</u> required to change a substance from solid to liquid <u>without a change in temperature</u>.

SI unit : J kg^{-1}

The **specific latent heat of vaporisation** L_v is defined as the <u>energy per unit mass</u> required to change a substance from liquid to gas <u>without a change in temperature</u>.

SI unit : J kg^{-1}

solid → liquid → gas ⇒ heat absorbed
gas → liquid → solid ⇒ heat lost

2.2 Conservation of Energy

Heat transfer is one form of energy transfer. Hence to solve any problem involving heat transfer, we always consider conservation of energy, presented in a simple word equation as follows:

Total energy or power supplied = total energy/power consumed + energy/power lost

It is important for us to identify the supplier and consumer of energy/power during the energy exchange.

Example 1

Suppose 0.400 kg of water initially at 40.0 °C is poured into a 0.300 kg glass beaker having a temperature of 25.0 °C. A 0.500 kg block of aluminum at 37.0 °C is placed in the water and the system is insulated. Calculate the final equilibrium temperature of the system. [37.8 °C]
Specific heat capacities of water, glass and aluminum are 4186, 837 and 900 J kg^{-1} K^{-1} respectively.

Example 2

A 5.00 kg block of ice at 0 °C is added to an insulated container partially filled with 10.0 kg of water at 15.0 °C.
(a) Find the final temperature, neglecting the heat capacity of the container. [0 °C]
(b) Find the mass of the ice that has melted. [1.89 kg]
Specific heat capacity of water is 4186 J kg^{-1} K^{-1}, specific latent heat of fusion of ice is 3.33 x 10^5 J kg^{-1}

Example 3

In an electrical method to determine the specific latent heat of vaporisation of a liquid, the liquid was heated by means of a heating coil and the steady-state quantities measured under two different operating conditions are given:

	Trial 1	Trial 2
Current in heating coil	2.5 A	3.0 A
Potential difference across coil	15.0 V	18.0 V
Mass of liquid vaporised in t = 250 s	38.5 g	59.6 g

(a) Calculate the specific latent heat of vaporisation of the liquid. [1.95×10^5 J kg^{-1}]

(b) Estimate the rate of loss of heat to the surroundings. [7.39 W]

2.3 Microscopic View of Energy

What type of changes will occur to a system when heat enters or leaves its boundary? To answer this question, we need to adopt a microscopic view of our system. The internal mechanism of a system is explained by the kinetic theory of matter – it states that all matter is made up of a large number of tiny atoms or molecules which are in continuous motion.

A neat way to present the molecular information of a system is to use the concept of energy – in this case, energy related to the molecules within the boundary of our system. At the molecular level, we are concerned primarily with two types of energy:

i) the kinetic energy due to the random movement of the molecules – microscopic kinetic energy, and

ii) the potential energy among molecules due to the intermolecular forces between them – microscopic potential energy. Gases have zero potential energy (or close enough to zero for real gases and zero for ideal gases). In relation to this, solids and liquids have negative potential energy. When heat is supplied to a solid or liquid, potential energy increases, so the value of potential energy becomes less negative and gets closer to zero.

Internal Energy

Energy is a scalar quantity. If we sum up all the above two types of energy for all molecules within the boundary of our system, the result of this summation is known as the ***internal energy***.

The internal energy *U* is the summation of microscopic kinetic energy due to random motion of the molecules and the microscopic potential energy due to intermolecular forces.

$$U = \Sigma (\text{K.E.})_{\text{mic}} + \Sigma (\text{P.E.})_{\text{mic}}$$

Microscopic Kinetic Energies

Due to the random movement of molecules.

Average microscopic K.E. of the molecules $\propto$ Absolute temperature of the system.

Microscopic Potential Energies

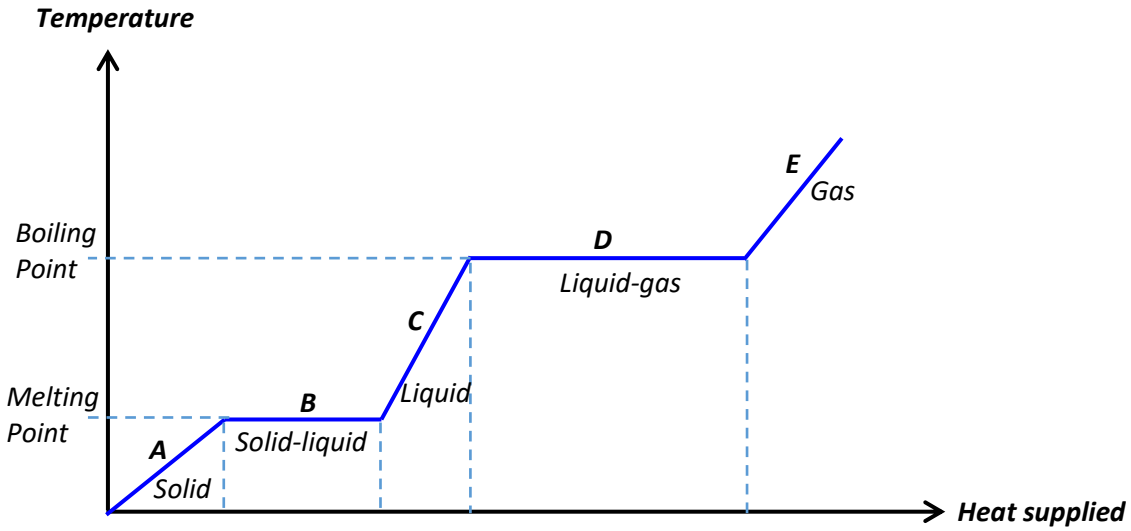
The microscopic potential energies arise from intermolecular forces between the molecules and they depend on the intermolecular separation and the type of bond.

Using this concept of internal energy and its associated concepts, we can explain the following:

i) At different phases, the molecules within the system exhibit different characteristics.
The existence of three phases of matter is determined primarily by the intermolecular forces and the type of motion which the molecules will have because of their internal energy.

ii) The temperature of a body is a measure of the **average kinetic energy** of the molecules that make up the body. Increases in this microscopic kinetic energy of a body can be detected by a rise in its temperature. Increases in microscopic potential energy, on the other hand, does not result in any rise in temperature. This is most dramatically demonstrated by boiling water; despite all the heat being supplied, the reading on the thermometer remains constant.

We are now ready to explain what heat transfer does to our system.
The diagram below shows how the temperature of a solid varies as heat is transferred to it at some initial temperature below its melting point as it turns into a liquid and then into a gas over time.



We can easily explain the feature of this graph using the concept of internal energy.
Note that as time progresses, heat is continuously supplied to the same matter in every part of the graph above.

Part	Phase	Temperature	K.E.$_{mic}$	P.E.$_{mic}$	Internal Energy
A	Solid	Increases	Increases	Marginal change	Increases
B	Solid-liquid	Constant at melting point	Constant	Increases	Increases
C	Liquid	Increases	Increases	Marginal change	Increases
D	Liquid-Gas	Constant at boiling point	Constant	Increases	Increases
E	Gas	Increases	Increases	Increases	Increases

See Appendix C: Energy and Phases of Matter.

Tutorial 11A

Unless otherwise stated,			
	molar gas constant	R	$8.31\ J\ K^{-1}\ mol^{-1}$
	Avogadro constant	N_A	$6.02 \times 10^{23}\ mol^{-1}$
	Boltzmann constant	k	$1.38 \times 10^{-23}\ J\ K^{-1}$
	Gravitational field strength	g	$9.81\ N\ kg^{-1}$
	density of mercury		$13\ 600\ kg\ m^{-3}$

Self-Review Questions

S1. Must two objects in thermal contact be physically touching each other? Explain your answer.

S2. **(N77/II/25)** A solid X is in thermal equilibrium with a solid Y, which is at the same temperature as a third solid Z. The three bodies are of different materials and masses.

Which of the following statements is certainly correct?

A X and Y have the same heat capacity.

B Y and Z have the same internal energy.

C There is no net transfer of energy if X is placed in thermal contact with Z.

D It is not necessary that Y should be in thermal equilibrium with Z.

E It is not necessary that X should be at the same temperature as Y.

S3. **(J88/I/26; J95/I/23)** Which thermodynamic temperature is equivalent to 501.85°C?

A 775.00 K **B** 228.85 K **C** 774.85 K **D** 228.70 K

S4. At absolute zero of temperature,

I the Celsius temperature is zero;
II the average kinetic energy of the particles of a substance is minimum;
III the Kelvin temperature is -273 K.

Which statement(s) is/are correct?

A **I** only **B** **II** only **C** **I** and **II** **D** **III** only

S5. **(N97/I/23, N00/I/24)** The temperature of a body at 100°C is increased by $\Delta\theta$ as measured on the Celsius scale. How is this temperature change expressed on the Kelvin scale?

A $\Delta\theta$ + 373 **B** $\Delta\theta$ + 273 **C** $\Delta\theta$ + 100 **D** $\Delta\theta$

S6. **(Serway)** The spirit-in-glass Galileo thermometer, invented in Florence, Italy, around 1654, consists of a tube of liquid (the spirit) containing a number of submerged glass spheres with slightly different masses. At sufficiently low temperature, all the spheres float. As the temperature increases, the spheres sink one after another. Hence this device is a crude but interesting tool for measuring temperature. *See the photograph on the cover page.*

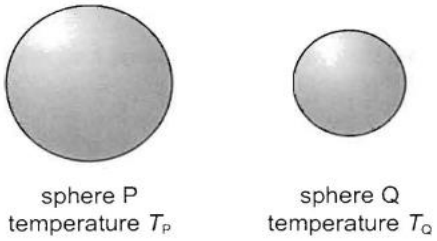
Suppose the tube is filled with ethyl alcohol, whose density is $0.78945\ g\ cm^{-3}$ at 20.0 °C and decreases to is $0.78097\ g\ cm^{-3}$ at 30.0 °C .

i) If one of the spheres has a radius of 1.000 cm and is in equilibrium halfway up the tube at 20.0 °C, determine the mass of this sphere.

ii) When the temperature increases to 30.0 °C, what should be the mass of another sphere of the same radius such that it is held in equilibrium at the halfway point?

iii) How do we read temperature from this thermometer?

S7. (N10/I/19) Two spheres P and Q are both made of steel. Sphere P has a radius that is larger than that of sphere Q.



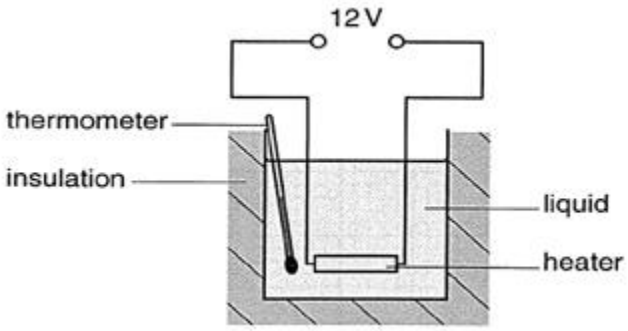
Initially, sphere P is at temperature T_P and sphere Q is at temperature T_Q, where $T_P > T_Q$.
The spheres are brought into contact and their final temperature is T. No thermal energy is transferred from the spheres to the surroundings.

Which expression gives the relation between T_P, T_Q and T?

A $(T_P - T) = (T - T_Q)$ **B** $(T_P - T) > (T - T_Q)$
C $(T_P - T) < (T - T_Q)$ **D** $(T_P - T) = (T + T_Q)$

S8. (N12/II/2)

a) A student uses the apparatus shown to determine the specific heat capacity of a liquid.



A power supply of potential difference 12V is connected to a heater of resistance 2.4 Ω. The mass of liquid is 140 g. The temperature of the liquid changes from 25°C to 45 °C in 210 s.

Calculate **i.** the power of the heater; [2]
ii. the student's value for the specific heat capacity of the liquid. [3]

b) The accepted value for the specific heat capacity of this liquid is 4200 J kg^{-1} K^{-1}. The uncertainties in the measurements in **(a)** cannot account for the difference. Suggest a cause for this difference.[1]

S9. (N10/III/1)

a) State what is meant by latent heat of vaporisation. [2]

b) A student carries out an experiment to determine the specific latent heat of vaporisation of water. Water is boiled in a beaker by means of an electric heater, as shown in Fig. 1.1.

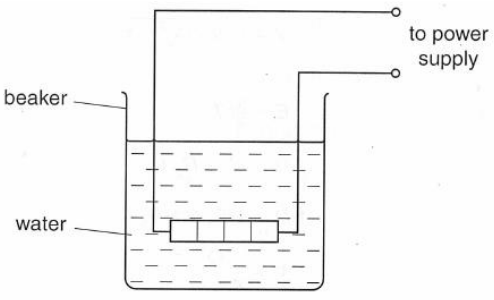


Fig. 1.1

P / W	m / g
140	14.1
95	8.2

Fig. 1.2

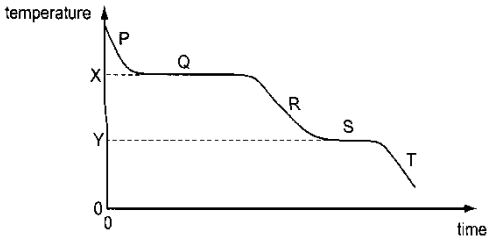
The power *P* supplied to the heater is measured. When the water is boiling at a constant rate, the mass *m* of the water evaporated in 5.0 minutes is determined.

Data for the power and the mass *m* for two different values of *P* are shown.

i. Suggest why, in order to obtain a reliable result for the specific latent heat, the mass *m* is determined for two different values of *P*. [1]

ii. Calculate the value for the specific latent heat *L* of vaporisation. [3]

S10. (N08/I/17) A vapour in a container is at a high temperature and loses heat to its surroundings. The graph shows how its temperature changes over the next few minutes.



Which feature of the graph indicates that the specific latent heat of vaporisation of the substance is greater than its specific latent heat of fusion?

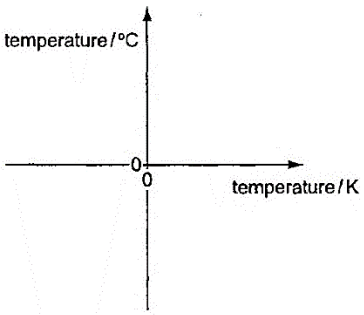
A The gradient of the graph at P is greater than the gradient at R.
B The gradient of the graph at T is greater than the gradient at R.
C The length of the line Q is greater than the length of the line S.
D The value of X is greater than the value of Y.

S11. The temperature of a hot liquid in a container of negligible thermal capacity falls at a rate of 2.0 °C per minute just before it begins to solidify. The temperature then remains steady for 20 minutes until the liquid has all solidified.
What is the ratio of the specific heat capacity of the liquid, *c*, to its specific latent heat of fusion, *L*?

Discussion Questions

D1. **(N05/I/13)** A student draws a linear graph on the axes shown in order to convert temperatures in kelvin to temperatures in degrees Celsius.

What is the intercept on the vertical axis and the gradient of the line?



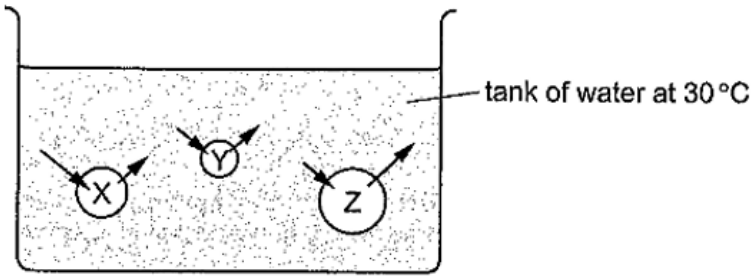
	intercept	gradient
A	0	1/273
B	–273	1
C	+273	1
D	0	273

D2. **(N17/I/11)** Which is the description of a graph relating temperatures in degrees Celsius (horizontal axis) to temperatures in kelvin (vertical axis)?

A a straight line of gradient 1/273 through the origin
B a straight line of gradient 273 through the origin
C a straight line of gradient 1 and vertical axis intercept of -273
D a straight line of gradient 1 and vertical axis intercept of 273

D3. **(N17/III/08) part**
State what is meant by the thermal equilibrium of two bodies by reference to
(i) temperature and (ii) movement of thermal energy

D4. **(N18/I/13)** Three objects X, Y and Z are put into a tank of water kept at 30 °C.
The arrows represent the input and the output of thermal energy for each object just after being placed in the tank. The length of each arrow corresponds to the quantity of energy transferred.



What are the initial temperatures of X, Y and Z?

	X	Y	Z
A	above 30 °C	at 30 °C	below 30 °C
B	above 30 °C	below 30 °C	above 30 °C
C	below 30 °C	above 30 °C	below 30 °C
D	below 30 °C	at 30 °C	above 30 °C

D5. **(N13/I/18)** A liquid is maintained at its boiling point by means of an electric heater. The constant rate at which the liquid boils away is measured for two different powers of the heater as shown.

power of heater	rate of loss of mass of liquid
P_1	m_1
P_2	m_2

For each power of the heater, P_1 or P_2, the rate of heat loss h to the environment is the same.
Which expression is correct for the specific latent heat of vaporisation of the liquid?

A $\frac{P_1}{m_1}$ **B** $\frac{1}{2}\left(\frac{P_1}{m_1}+\frac{P_2}{m_2}\right)$ **C** $\frac{(P_1-P_2)}{(m_1-m_2)}$ **D** $\frac{(P_1+P_2)}{(m_1+m_2)}$

D6. **(N21/I/13)** Thermal energy is supplied to 5.00 kg of water at 30.0 °C. All of this water becomes steam at 100 °C. There is no heat loss to the surroundings.
Specific heat capacity of water = 4190 J kg^{-1} K^{-1}.
Specific latent heat of vaporisation of water = 2260 kJ kg^{-1}.

How much thermal energy is supplied to the water?

A 1.48×10^6 J
B 3.73×10^6 J
C 1.13×10^7 J
D 1.28×10^7 J

D7. **(N21/I/14)** A canister of an ideal gas, at rest on the Earth's surface, has an internal energy of 3.0 kJ. The canister is then thrown upwards to a point X where the potential energy of the gas has increased by 0.5 kJ and its kinetic energy is 1.0 kJ. There is no heat transfer or work done on the ideal gas.

What is the internal energy of the gas when it is at point X?

A 2.5 kJ **B** 3.0 kJ **C** 4.0 kJ **D** 4.5 kJ

D8. (N17/III/8) part A student carries out an experiment with melting ice, using the apparatus illustrated in Fig. 8.1.

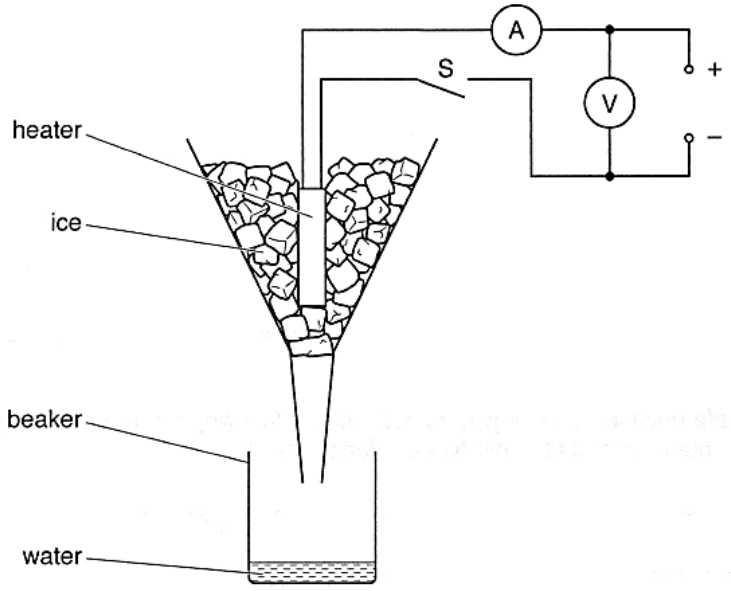


Fig. 8.1

Initially, the switch S is open and the funnel is kept filled with pure melting ice. Water drips out of the funnel at a constant rate.

(i) Suggest and explain whether the ice is in thermal equilibrium with its surroundings. [1]

(ii) The switch S is closed. When water drips out of the funnel at a different constant rate, the beaker is replaced with an empty beaker and the readings shown in Fig. 8.2 are obtained.

ammeter reading /A	voltmeter reading /V	mass of empty beaker/g	mass of beaker plus water/g	time of collection of water /minutes
6.3	12.0	32.4	114.0	5.0

Fig. 8.2

The specific latent heat of fusion of ice is 330 J g^{-1}. Calculate the rate R at which thermal energy is gained from the atmosphere by the ice in the funnel. [3]

D9. **(N15/III/7) part (b)** A student determines the specific heat capacity of water using the apparatus shown in Fig. 7.1.

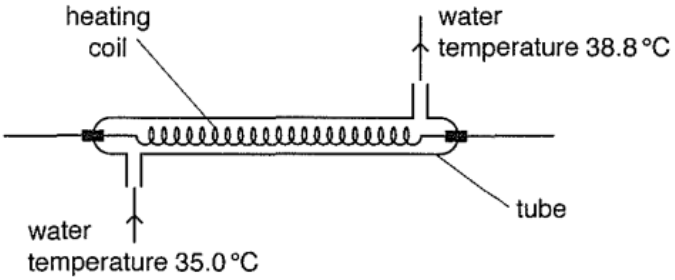


Fig. 7.1

Water enters the tube at a constant temperature of 35.0 °C. The electrical power *P* dissipated in the heating coil and the flow of water through the tube are varied so that the output temperature of the water is always 38.8 °C. The mass of water flowing through the tube per unit is m.

The variation with power *P* of *m* is shown in Fig. 7.2.

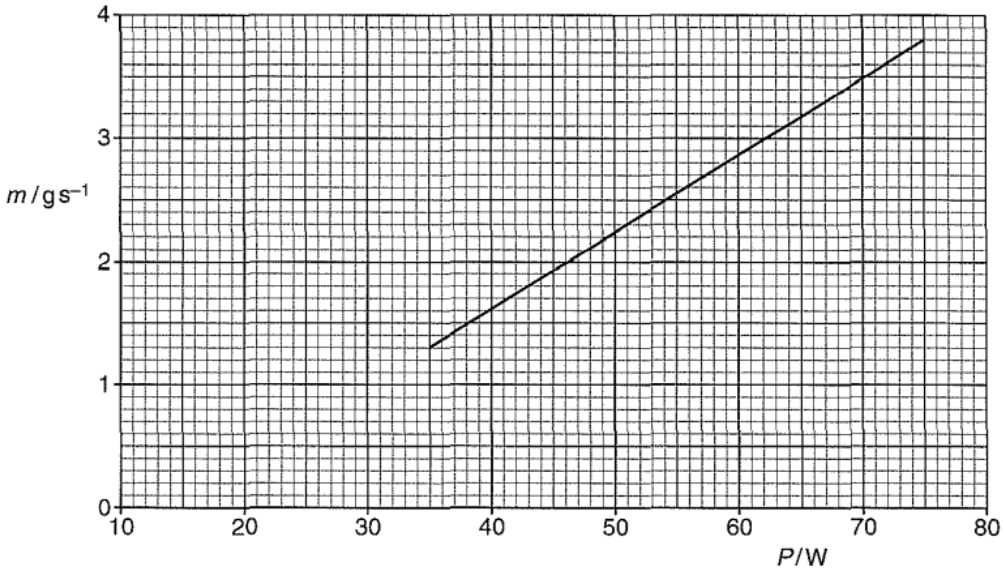


Fig. 7.2

(ii) Suggest why there is an intercept on the power axis of the graph of Fig. 7.2. [1]

(iii) Use Fig. 7.2 to determine value for the specific heat capacity, in J kg^{-1} K^{-1}, of the water. Show your working. [4]

(c) The experiment in **(b)** is repeated with water entering the tube at a temperature of 45.0 °C and leaving at 48.8 °C.
On the axes of Fig. 7.2, sketch a graph to show qualitatively the variation with power *P* of *m*. [2]

Numerical answers to selected questions

S6	S8	S9	S11
(i) 3.31 g (ii) 3.27 g	60 W 4500 J kg^{-1} K^{-1}	2290 J g^{-1}	0.025 K^{-1}

D8	D9
R = 14.2 W	4190 J kg^{-1} K^{-1}

3 Ideal Gas

Kinetic Theory
The kinetic theory of matter states that all matter is made up of a large number of tiny atoms or molecules which are in continuous motion. We can explain the three phases of matter and their various macroscopic properties in terms of the behaviour of their constituent molecules (microscopic properties).

Why Study Ideal Gases
The study of the three phases of matter is not an easy task as it involves various branches of knowledge in physics like mechanics (whether classical or quantum) and statistics etc. Among the three phases of matter, the gaseous phase presents itself as the easiest one to be studied especially when the gas molecules are spaced far apart from each other and not interacting with each other. Hence we will attempt to explain some macroscopic properties of such a gas via the collective behaviour of the gas molecules. This simplified gas system is known as the **ideal gas** and it has the following microscopic properties:

Assumptions/Properties of an Ideal Gas
(A1) A gas consists of a very large number of molecules.
(A2) The gas molecules are moving randomly.
(A3) Collisions of the gas molecules with one another and with the walls of the container are perfectly elastic.
(A4) There are no intermolecular forces of attraction.
(A5) The volume of the gas molecules is negligible, compared to the volume of the container, which is the volume of the gas.
(A6) The duration of collisions is negligible, compared to the time interval between collisions.

Ideal Gas Equation of State
For an ideal gas, the equation of state is:

$$pV = nRT$$

which can be expressed as $pV = NkT$, since $R = kN_A$,

where
pressure p, volume V, absolute temperature T, amount of ideal gas in moles n, amount of ideal gas in number of molecules N, molar gas constant R = 8.31 J K^{-1} mol^{-1} and Boltzmann constant k = 1.38 x 10^{-23} J K^{-1} are linked.

3.1 A Really Huge Number – The Avogadro Number, N_A

In the microscopic regime, we are going to deal with a large number of molecules. In fact, one mole of any substance is that amount of substance that contains 6.02 x 10^{23} of constituent particles (whether atoms, molecules or ions).

6.02 x 10^{23} is known as the Avogadro's number N_A. If we are given 0.012 kg of carbon-12, there will be 6.02 x 10^{23} of atoms in it – this is the technical definition of the Avogadro's number. Similarly, in the case of helium-4 atoms, its molar mass is 4.0 x 10^{-3} kg. If we are given 4.0 x 10^{-3} kg of helium gas, we know that there are 6.02×10^{23} helium-4 atoms in it. In the case of oxygen gas, we have to be a bit more careful because oxygen gas is made up of oxygen molecules (two oxygen atoms). So one mole of oxygen gas contains 6.02×10^{23} oxygen molecules and this gas has a mass of 32 g.

$$number\ of\ mole, n = \frac{mass\ of\ sample}{Molar\ mass\ of\ sample}$$

Looking at the periodic table, you would have noticed that the molar mass of each ATOM very closely resembles the mass number of that atom (the number of protons and neutrons).

Note that *R*, *k* and N_A are provided in the constants list.

Example 4

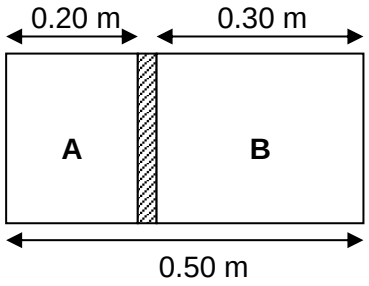
Show numerically that k is related to R via $k = \frac{R}{N_A}$.

Example 5

A cylinder, 0.50 m long and of cross-sectional area *S*, closed at each end, is fitted with a smooth thermally insulating piston, separating two gases **A** and **B** into two compartments as shown.

Initially, the piston is in equilibrium at a distance of 0.20 m from one end of the cylinder, and the gas in each compartment is at a temperature of 27 °C.

The temperature of the smaller volume of gas, **A**, is now increased to 177 °C, while that of the gas in the larger volume **B** is maintained at 27 °C. The piston moves through a distance *x* until a new position of equilibrium is obtained.



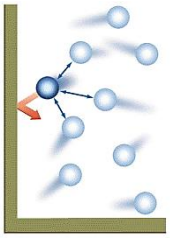
a) State the equation relating the initial and final pressures, volumes and temperatures of the gas in **A**.
b) State the equation relating the initial and final pressures, volumes and temperatures of the gas in **B**.
c) Hence find the new position of equilibrium. [0.25 m from the left]

3.2 Molecular Interpretation of Pressure

How Kinetic Theory Explains the Nature of Pressure Exerted by Gas

According to the kinetic theory, a gas confined in a container is made up of a large number of gas molecules moving randomly and colliding with each other as well as with the walls of the container.

When the gas molecules collide elastically with the container of the gas, they rebound off the wall with the same speed. Clearly, the walls of the container exert a force on the gas molecules to cause their momentum to change in direction. By Newton's 2nd Law, this force is equal to the rate of change of momentum of the molecule. By Newton's 3rd Law, the molecules exert an equal but opposite force on the walls.



Since there is a large number of molecules, the number of collisions at any instant is very large and practically constant, resulting in a constant gas pressure.

The **pressure by the gas on the container** is the **average force per unit area exerted by the molecules on the walls of the container.**

The pressure *P* due to the gas molecules on a wall of the container can be written as

$$P = \frac{\text{average force exerted on the wall by molecules}}{\text{area of the wall}}$$

$$= \frac{(\text{no. of collisions per unit time})(\text{change of momentum per collision})}{\text{area of the wall}}$$

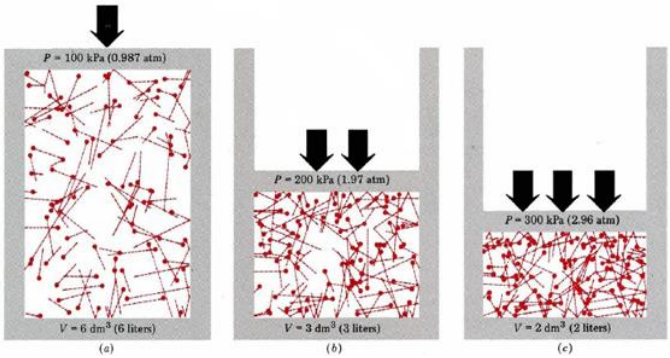
We can now explain how the following factors can lead to an increase in pressure of a gas:

- **No. of molecules in the container.** All other conditions being constant, a larger number of molecules will imply more collisions made with the wall of the container.

- **Volume of container.** Having a smaller container implies that the molecules need to travel less distance to reach the next wall after rebounding from the opposite wall. Therefore, the no. of collisions they make with the walls per unit time increases.

- **Mass of the molecule.** The type of gas used will determine the mass of a molecule which in turn determines the momentum that each molecule possesses. The larger the mass of a molecule, the greater the momentum it will have. Thus, the change of momentum of the molecules per collision will be greater.

- **Velocities of the molecules.** The momentum of each gas molecule is determined by its velocity. A molecule possessing a larger velocity will have a greater momentum and hence its change in momentum will correspondingly be larger per collision.



Example:
There are equal numbers of molecules at the same temperature in each of the enclosed containers as shown. As the volume of the gas decreases, the number of molecules per unit volume increases. The number of molecules striking a given area of the walls per unit time also increases. Thus increasing the force on that area. This force per unit area is the pressure and hence the pressure increases.

3.2.1 Derivation of $pV = \frac{1}{3}Nm\langle c^2\rangle$

We shall attempt to relate the macroscopic quantity, pressure, with the microscopic model of gases as a collection of atoms moving in random motion in a vessel.

[*Need to know the derivation of the relationship* $pV = \frac{1}{3}Nm\langle c^2\rangle$ *as stated in LO*]

Step 1
Resolve the motion of molecule along *x*-*y*-*z* directions

Consider a gas of *N* molecules in a cubical container of side L (Fig. F1). Suppose the mass of a molecule is *m*. Consider a molecule moving with a velocity c_1. The velocity c_1 of the molecule can be resolved into *x*, *y* and *z* directions as u_1, v_1 and w_1 respectively.

Using Pythagoras' theorem, $c_1^2 = u_1^2 + v_1^2 + w_1^2$.

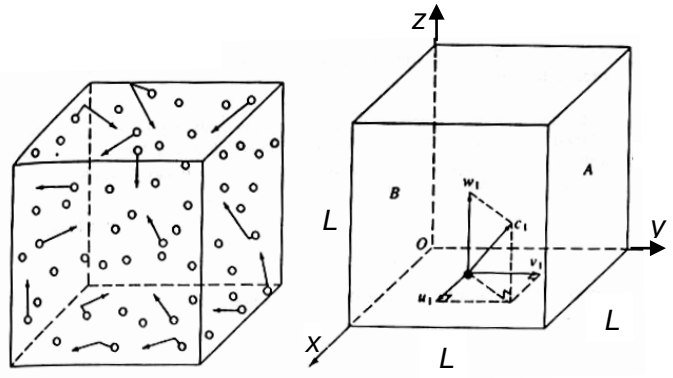
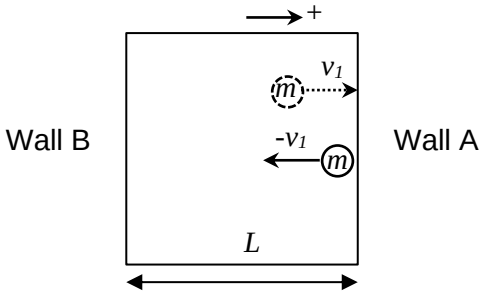


Fig. F1

Step 2
Find the change of momentum upon collision with a wall

Consider the component of velocity v_1 along the positive *y*-axis. After colliding with the wall A, the molecule rebounds with a velocity $-v_1$ since the collision is perfectly elastic.

Momentum before collision = mv_1
Momentum after collision = $-mv_1$
∴ change of momentum = $-mv_1 - mv_1$
= $-2mv_1$



Step 3
Find the average force on the wall due to the bombardment of molecules on the wall

Assuming that the molecule does not collide with other molecules, the time interval *t* between a collision of the molecule with the wall A and its next collision with the same wall is the same as the time the molecule takes to move from wall A to wall B and back and is given by:

$$\Delta t = \frac{2L}{v_1}$$

Hence, the number of collisions made by the molecule with wall A per unit time $= \frac{1}{\Delta t} = \frac{v_1}{2L}$

the average force exerted on the molecule by wall A $= \frac{\Delta(mv)}{\Delta t} = -2mv_1 \times \frac{v_1}{2L} = -\frac{mv_1^2}{L}$

By Newton's 3rd Law, the force exerted on wall A by the molecule = $\frac{mv_1^2}{L}$

During its passage back and forth across the container, the molecule may collide with the top and bottom and sides of the container, but this does not alter its component of momentum and thus does not alter our result. It may also collide elastically with other molecules which may change its v_1. However, any loss (or gain) of momentum is acquired by other molecules, and because we will eventually sum over all the molecules, this effect will be included. So our result above is not altered.

The actual force due to one molecule is a spike, but because a huge number of molecules are striking the wall per second, the force is, on average, nearly constant. To calculate the force due to all the molecules in the container, we have to add the contribution of each.

Since there are N molecules, each with the component of velocity along the *y*-axis as v_1, v_2, v_3, …, v_N respectively, then the total force on wall A due to the collision of these N molecules is

$$F = \frac{mv_1^2}{L} + \frac{mv_2^2}{L} + \frac{mv_3^2}{L} + \cdots + \frac{mv_N^2}{L}$$

Step 4

Find the pressure on the wall

Pressure on wall A,

$$p = \frac{F}{L^2} = \frac{\frac{mv_1^2}{L} + \frac{mv_2^2}{L} + \frac{mv_3^2}{L} + \cdots + \frac{mv_N^2}{L}}{L^2} = \frac{m}{L^3}\left(v_1^2 + v_2^2 + v_3^2 + \cdots + v_N^2\right),$$

where L^3 = volume of container.

Let $V = L^3$

Since $v_1^2 + v_2^2 + v_3^2 + \ldots v_N^2 = N\left\langle v^2 \right\rangle$, $p = \frac{Nm\left\langle v^2 \right\rangle}{V}$

Step 5

Express $\left\langle v^2 \right\rangle$ along the y-axis in terms of mean-square speed $\left\langle c^2 \right\rangle$

Each molecule has a different velocity. Following the square speed equation (using Pythagoras' theorem) for a molecule in Step 1, we denote the square speed equations for N molecules in the container as follows:

$$c_1^2 = u_1^2 + v_1^2 + w_1^2$$
$$c_2^2 = u_2^2 + v_2^2 + w_2^2$$
$$\vdots$$
$$c_N^2 = u_N^2 + v_N^2 + w_N^2$$

Adding all N square speed equations and dividing by N, we have

$$\frac{c_1^2 + c_2^2 + c_3^2 + \cdots + c_N^2}{N} = \frac{u_1^2 + u_2^2 + u_3^2 + \cdots + u_N^2}{N} + \frac{v_1^2 + v_2^2 + v_3^2 + \cdots + v_N^2}{N} + \frac{w_1^2 + w_2^2 + w_3^2 + \cdots + w_N^2}{N}$$

In short, the mean-square speed or average value of square speed of N molecules is:

$$\langle c^2 \rangle = \langle u^2 \rangle + \langle v^2 \rangle + \langle w^2 \rangle$$

Since the motion of the gas molecules is random in nature and the number of molecules is large, there is no preference for one direction or another, and

$$\langle u^2 \rangle = \langle v^2 \rangle = \langle w^2 \rangle$$

This implies

$$\langle c^2 \rangle = \langle v^2 \rangle + \langle v^2 \rangle + \langle v^2 \rangle$$

$$\langle v^2 \rangle = \frac{1}{3}\langle c^2 \rangle$$

Substitute the above into p in Step 4: $p = \frac{1}{3}\frac{Nm\langle c^2 \rangle}{V}$ (provided in A-level Formulae List)

rearranging, we have $\boxed{pV = \frac{1}{3}Nm\langle c^2 \rangle}$ ------ Equation (1)

where N = no. of molecules, m = mass of 1 molecule (also known as molecular mass) and $<c^2>$ is *the mean square speed* of the molecules.

Notes:

1. There are a few ways to rewrite Equation (1).
   - Since N is the total number of molecules and m is the mass of each molecule, then
     $\Rightarrow$ Total mass of the gas = Nm

     If the **molar mass of the gas** is M and there are n moles of gas in the container,
     $\Rightarrow$ Total mass of the gas = $Nm = nM$

     and therefore
     $$\boxed{pV = \frac{1}{3}nM < c^2 >}$$
   - We can also write Equation (1) as
     $\Rightarrow$ $p = \frac{1}{3}\frac{Nm < c^2 >}{V} = \frac{1}{3}\frac{(\text{Mass of gas})}{(\text{Volume of gas})} < c^2 >$
     $\Rightarrow$ $\boxed{p = \frac{1}{3}\rho < c^2 >}$

     where ρ is the density of the gas.

3.3 Molecular Interpretation of Temperature

For ideal gases, $pV = \frac{1}{3}Nm\langle c^2\rangle$ and $pV = NkT$

Equating the two expressions: $\frac{1}{3}Nm\langle c^2\rangle = NkT$

$$\boxed{\frac{1}{2}m\langle c^2\rangle = \frac{3}{2}kT}$$ (*E* = 3/2 *kT* is provided in A-level Formulae List)

Notes

1) $\frac{1}{2}m\langle c^2\rangle$ is the average kinetic energy of a gas molecule and it is proportional to the absolute temperature of the gas T.

2) $\langle c^2\rangle$ is the mean square speed of the molecules. Taking the square root of $\langle c^2\rangle$, i.e. $\sqrt{\langle c^2\rangle}$ gives us the root-mean-square speed of the molecules $c_{r.m.s.}$. The molecules have different speeds; $c_{r.m.s.}$ is one way to represent the "average speed" of these molecules.

3) If we multiply the number of molecules N on both sides of the equation i.e.

$$N\left(\frac{1}{2}m\langle c^2\rangle\right) = N\left(\frac{3}{2}kT\right)$$

$N\left(\frac{1}{2}m\langle c^2\rangle\right)$ is the total kinetic energy of all the ideal gas molecules; i.e. the microscopic kinetic energy of all molecules.

Since ideal gases are assumed to have no intermolecular forces of attraction, implying that they do not have microscopic potential energy, the internal energy U is entirely made up of microscopic kinetic energy. $U = N\left(\frac{1}{2}m\langle c^2\rangle\right) = N\left(\frac{3}{2}kT\right) = \frac{3}{2}nRT = \frac{3}{2}PV$.

For an ideal gas, **the internal energy is a function of its absolute temperature**.

4) If you look back at the derivation, you'd see that the "3" in $\frac{3}{2}nRT$ came from the three directions of motion (*x*, *y* and *z*). Strictly speaking, $U = \frac{3}{2}nRT$ applies only to monatomic gases where we do not have to consider other "types" of kinetic energy such as rotational kinetic energy; we only consider their *x*, *y* and *z* translational kinetic energy. For non-monatomic gases which can rotate in situ there are more than the three translational directions that contribute to kinetic energy, and the factor is strictly speaking not $\frac{3}{2}$. However, in many A-level questions the type of gas is not specified (or what is required is just the translational kinetic energy) and we often just use $U = \frac{3}{2}nRT$.

Example 6

The distribution of the velocities of particles moving along the *x*-axis are as follows:

Velocity *v* / ms^{-1}	– 3.0	– 2.0	– 1.0	0.0	1.0	2.0	3.0
No. of particles	1	3	4	3	2	1	3

What is a) the average velocity of the particles? [0 m s^{-1}]
b) the average speed of the particles? [1.53 m s^{-1}]
c) root-mean-square speed of the particles? [1.85 m s^{-1}]

Example 7

A cylinder contains 2.00 mol of helium gas at 20.0 °C. Assume that helium behaves like an ideal gas.

a) What is the internal energy of the system? [7300 J]
b) What is the average kinetic energy per molecule? [6.07 x 10^{-21} J]
c) How much energy would have to be added to the system to double the r.m.s. speed? [21900 J]

Tutorial 11B

Unless otherwise stated,

molar gas constant	R	$8.31\ J\ K^{-1}\ mol^{-1}$
Avogadro constant	N_A	$6.02 \times 10^{23}\ mol^{-1}$
Boltzmann constant	k	$1.38 \times 10^{-23}\ J\ K^{-1}$
Gravitational field strength	g	$9.81\ N\ kg^{-1}$
density of mercury		$13\ 600\ kg\ m^{-3}$

Self-Review Questions

S1. **N07/P1/19.** 20 g of oxygen at a temperature of 100 °C occupy a cubic container of side 20 cm. The mass of a mole of oxygen is 32 g. What is the pressure of the oxygen in the container?

A 12 kPa **B** 65 kPa **C** 240 kPa **D** 7800 kPa

S2. When ice changes into water at 0 °C,

I *work is done in breaking the molecular structure of ice in solid state.*
II *internal energy is increased.*
III *energy is absorbed to raise the temperature.*

Which of the statement(s) above is/are correct?

A **I** only **B** **II** only **C** **I** and **II** only **D** **I**, **II** and **III**

S3. When a gas of fixed mass is heated under constant pressure,

I *the volume of the gas increases.*
II *the average speed of the gas molecules increases.*
III *the average separation of the gas molecules increases.*

Which of the statement(s) above is/are correct?

A **I** only **B** **II** only **C** **II** and **III** only **D** **I**, **II** and **III**

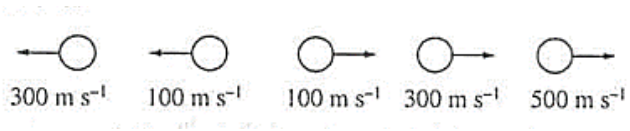
S4. Water in a kettle boils and turns into steam. What is happening to the water molecules?

A They are losing energy. **B** They are beginning to vibrate.
C They are expanding and getting lighter. **D** They are moving further apart.

S5. Use the kinetic theory of matter to explain briefly why

a) a gas exerts pressure on the walls of its container,
b) energy is required to evaporate a liquid.

S6. **(N91/1/25; N97/1/25)** Five molecules are moving with the speeds and directions shown.



What is the root mean square (r.m.s.) speed of these molecules?

A 100 m s^{-1} **B** 224 m s^{-1} **C** 300 m s^{-1} **D** 500 m s^{-1}

S7. A kinetic theory formula relating pressure *p* and the volume *V* of a gas to the root mean square speed of molecules is $p = \frac{1}{3}\frac{Nm}{V}\langle c^2 \rangle$

a) In this formula, what does the product *Nm* represent?

b) In the derivation of this relationship, what are the assumptions of the kinetic model of a gas?

c) From this relationship, show that the expression of the average translational kinetic energy of the molecules of gas is (3/2)*kT* , where *k* is the Boltzmann constant.

d) The density of argon at pressure of 1.00 x 10^5 Pa and at a temperature of 300 K is 1.60 kg m^{-3}. What is the root mean square speed of argon molecules at this temperature?

Discussion Questions

D1. (J82/II/28; J99/I/25) Two vessels X and Y, of volumes V_x and V_y, are kept at temperatures T_x and T_y. They are filled with the same ideal gas and are connected by a narrow tube.

What is the ratio of $\frac{\text{number of molecules in X}}{\text{number of molecules in Y}}$?

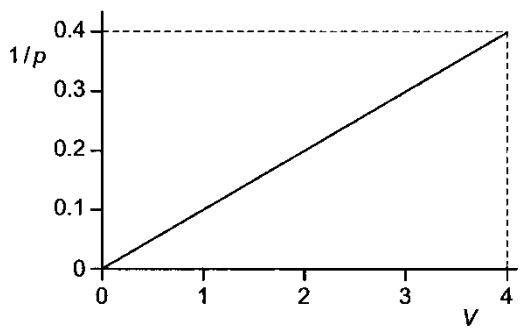
A $\frac{T_X V_X}{T_Y V_Y}$ **B** $\frac{T_X V_Y}{T_X V_X}$ **C** $\frac{T_Y V_Y}{T_Y V_X}$ **D** $\frac{T_Y V_X}{T_X V_Y}$

D2. (N11/I/19) A car tyre, initially at 25 °C, has been inflated to a pressure of 200 kPa as indicated by the pressure gauge. This means that the pressure in the type is 200 kPa above the atmospheric pressure of 100 KPa.

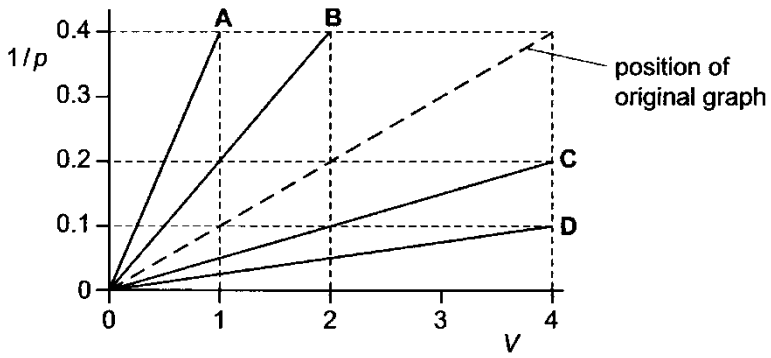
After driving on hot roads, the temperature of the air in the tyre is 50 °C. What is the percentage increase in the pressure gauge reading?

A 8.4% **B** 12.5% **C** 100% **D** 150%

D3. (**N08/I/18)** A fixed amount of an ideal gas has pressure *p* and volume *V*. Graph I shows the variation of 1/*p* with *V* at a constant temperature.



Graph I



The amount of gas and the thermodynamic temperature are both doubled. Which line will be produced?

D4. **(N87/I/22)** A gas is fitted with a safety valve which releases a gas when the pressure inside the cylinder reaches 2.0 x 10^6 Pa. Given that the maximum mass of this gas the cylinder can hold is 10°C is 15 kg, what would be the maximum mass at 30°C?

A 5.0 kg **B** 14 kg **C** 15 kg **D** 16 kg **E** 45 kg

D5. **(N08/III/1(a))** The variation with volume V of the pressure p of a fixed mass of gas at constant temperature is shown in Fig. 1.1.

i. Using values from Fig. 1.1, plot a second graph using the axes drawn in Fig. 1.2, to show that p is inversely proportional to V. [2]

ii. State how inverse proportion is demonstrated by your graph. [1]

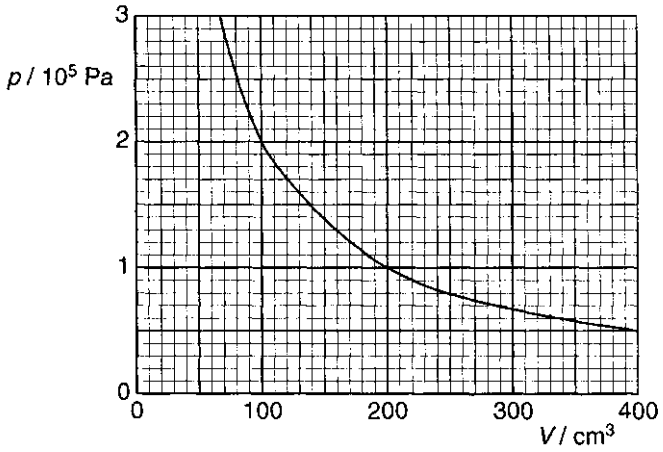


Fig. 1.1

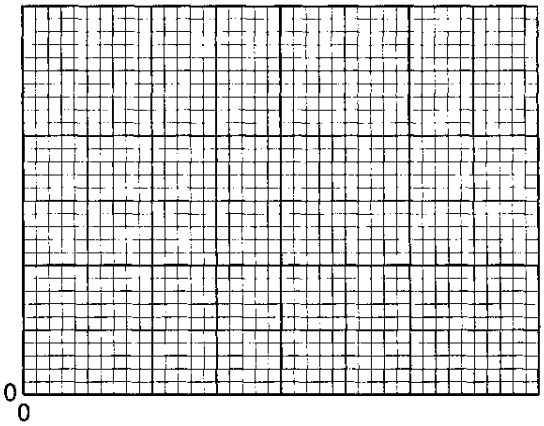


Fig. 1.2

D6. A uniform capillary tube, closed at one end, contained air trapped by a thread of mercury 85 mm long. When the tube was held horizontally, the length of the air column was 50 mm; when it was held vertically with its closed end below its open end, the length was 45 mm. Given that density of mercury is 13600 kg m^3, find the atmospheric pressure.

D7. **(N13/II/4)**

a) Explain what is meant by the internal energy of an ideal gas. [1]

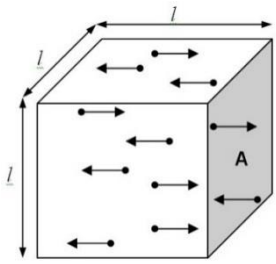
b) An oven contains air at a pressure of 1.0 x 10^5 Pa and a temperature of 25 °C. the volume of the air in the oven is 0.075 m^3. The mass of one mole of air is 0.030 kg. Air may be considered as an ideal gas.

i. Calculate the mass of air in the oven. [3]

ii The oven is heated to 200 °C and the pressure is constant at 1.0 x 10^5 Pa.

Calculate the ratio $\frac{\text{density of air at 25 °C}}{\text{density of air at 200 °C}}$ Explain your working. [2]

D8. **(N94/III/5 (parts), modified)** The figure below shows a cubical box of side *l* which contains *N* molecules, each of mass *m*, all moving horizontally with speed *v* at right angles to wall A.



a) Deduce

i. the momentum of a molecule just before a collision with a wall,

ii. the change in momentum of a particle when it makes one collision with wall,

iii. the time taken by one molecule between collisions with wall A.

iv. the total number of collisions per unit time made with wall A by all the molecules with wall A.

v. the average force exerted by all *N* molecules on wall A. [7]

b) Use your answer to (a) to show that the pressure *p* on the wall A is given by $p = \frac{Mv^2}{V}$ where *M* is the total mass of all the molecules and *V* is the internal volume of the box. [2]

c) The conditions considered in part (a) are highly improbable. Explain briefly how the conditions may be altered to provide a better model of an ideal gas. State, how the equation in (b) might be modified. [4]

D9. **(N20/III/2)** A sample of an ideal monatomic gas a volume of 4.2 x 10^{-3} m^3 at a pressure of 3.6 x 10^5 Pa and a temperature of 70 °C.

(a) Explain why the mean velocity of the atoms of the gas is zero. [2]

(b) Show that the sample of gas contains 3.2 x 10^{23} atoms. [3]

(c) Each atom of the gas has a diameter of approximately 2 x 10^{-10} m.
Estimate the volume of the gas atoms. [1]

(d) Explain why your answer in (c) provides evidence to support the assumption that the gas is ideal. [2]

D10. **(N21/III/3)**

(a) State what is meant by the internal energy of an ideal gas. [2]

(b) A fixed mass of an ideal gas has a volume of 3.2 x10^{-3} m^3 at a pressure of 1.0 x 10^5 Pa and a temperature of 12 °C. The gas is heated at constant pressure so that its volume increases to 3.6 x 10^{-3} m^3 at temperature θ, as shown in Fig. 3.1.

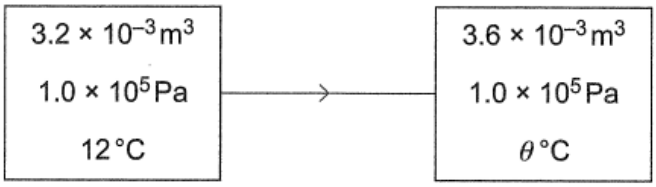


Fig. 3.1

(i) Calculate the final temperature θ, in degrees Celsius, of the gas. [2]

(ii) Determine the work done against the atmosphere during the expansion of the gas. [2]

(c) During the heating process in (b), 101 J of thermal energy is supplied to the gas.

(i) Use your answer in (b)(ii) to determine the increase in internal energy of the gas. [1]

(ii) Calculate the average increase in kinetic energy of a molecule of the gas. [3]

Numerical answers to selected questions

S7
433 m s^{-1}

D6	D7	D9	D10
1.0 x 10^{5} Pa	0.091 kg 1.59	1 x 10^{-6} m^{3}	bi) 47.6 °C bii) 40 J ci) 61 J cii) 7.4 x 10^{-22} J

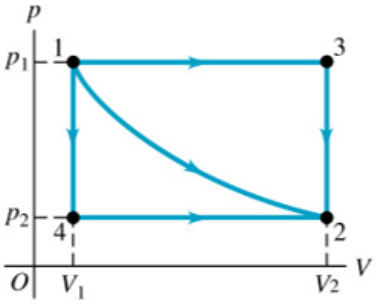
4 State of A System

We started this topic by introducing the concept of a system. If a system is not undergoing any change, the set of all the properties that can be measured or calculated throughout the entire system will specify the condition of this system. We call this the state of the system.

For this course, there are a few properties of a system that we are interested in: pressure p, volume V, temperature T, internal energy U.

pV Diagram

One way to monitor how a system changes its state is to chart the properties on a diagram. One such diagram is known as an indicator diagram. We will use mainly the pressure vs volume graph (pV diagram). A point on the pV diagram indicates a particular state of the system.



Process

A process occurs when a system changes from one state to another state. In our course, we indicate a process using a line with an arrow that connects the initial state to the final state.

4.1 First Law of Thermodynamics

The study of any system is centred on the concept of energy conservation. We will now develop this concept in the context of thermodynamics.

Here are the two forms of energy we have dealt with so far:

1) Heat transfer Q
2) Internal energy U

We have shown that if heat enters or leaves the boundary of a system, the internal energy could change.

So the first link between these two forms of energy could be: $\Delta U = Q$

A positive sign for Q means heat enters the system and the internal energy U increases (leading to a positive sign for ΔU).

This link between ΔU and Q may appear correct. However it is incomplete. Your experience would have told you that the internal energy could change even when there is no heat flow. Can you describe one?

There are certainly other forms of energy transfer (other than heat) that could affect the internal energy. We use the term 'work' (symbol W) to represent these alternative energy transfers.

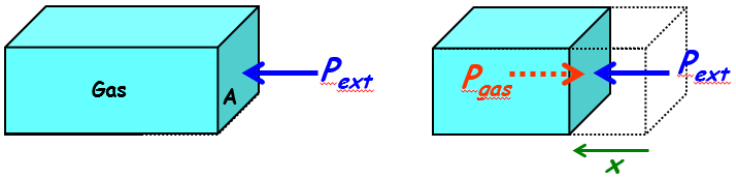
So our energy transformation relationship becomes: $\Delta U = Q + W$

This is known as the ***First Law of Thermodynamics***:
The increase in the internal energy of a system is the sum of the heat supplied to the system and the work done on the system.

We will limit work W mainly to the energy transfer via some form of force or pressure.

Remember that in mechanics, if a force F acting on a system causes the point of application to move by a distance x in the direction of F, we say that positive work Fx is done on the system. So there is an attempt by the agent exerting the force F to transfer energy of magnitude Fx into the system.

This can be applied to a gas system.



Suppose an external pressure P_{ext} is exerted on a gas and the gas is compressed by a tiny distance x at constant cross-sectional area A. In this case, work done by external agent, $W = +P_{ext}(Ax) = +P_{ext}V$.

One immediate advantage of using a pV diagram is that it allows us to visualize the work done W on our system. The magnitude of W for any process is the area between the line due to this process and the volume axis. How about the sign of W? If the system contracts (volume decreases), W is positive and vice versa.

Summary of Each Term in First Law of Thermodynamics & its sign convention

The increase in the internal energy of a system is the sum of the heat supplied to the system and the work done on the system.

$$\Delta U = Q + W$$

	ΔU	Q	W
Definition	Increase in internal energy of system	Heat energy supplied to system	Work done on system
Sign	+: U increases –: U decreases	+: heat enters system –: heat leaves system	+: system contracts –: system expands
magnitude	May be deduced from question	May be deduced from question	Area between pV graph and V-axis

4.2 Four Basic Thermodynamic Processes

A system can change from one state to another via many processes. There are, however, four of them that are particularly well-known:

Process	Characteristic
Isothermal	Constant temperature
Isovolumetric	Constant volume
Adiabatic	No heat transfer (Q = 0)
Isobaric	Constant pressure

Isothermal Process (Constant Temperature Process)

Imagine an ideal gas undergoing isothermal expansion at constant temperature T_1 from state *i* to state *f*.

$$pV = nRT_1 \quad \Rightarrow \quad p = (nRT_1)\left(\frac{1}{V}\right)$$

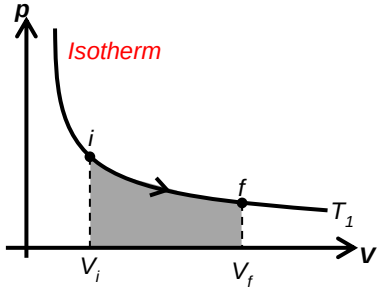
Plotting p against V gives a reciprocal graph.

We have already proven that the internal energy of an ideal gas is a function of its absolute temperature. Hence internal energy of an ideal gas remains constant during an isothermal process; i.e. $\Delta U = 0$.

To maintain constant temperature throughout the process, the expansion or compression should be done in a good conductor, and very slowly, to allow for effective heat exchange with the surroundings and hence for temperature to stay constant.

By the first law of thermodynamics, $Q = -W$

Magnitude of W is represented by the shaded area under the graph.



Example: The p-V graph shows expansion of the gas at constant temperature.
$\Delta V > 0 \Rightarrow$ work done on gas is negative, $W < 0$
$T_i = T_f \Rightarrow \Delta U = 0$

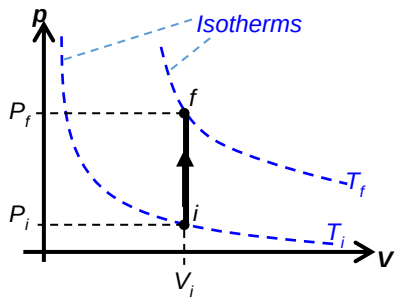
Applying First Law: $Q = -W > 0$
i.e. heat is supplied to the gas to do work against the atmosphere.

Isovolumetric Process (Constant Volume Process)

This process is represented by a vertical line on a PV diagram.
Clearly the area under an infinitesimally thin line is zero.

Hence $W = 0$.

$\Delta U = Q$



Example: The p-V graph shows increase of gas pressure at constant volume.
$\Delta V = 0 \Rightarrow$ work done on gas is zero, $W = 0$
$T_f > T_i \Rightarrow \Delta U > 0$

Applying First Law: $Q = \Delta U > 0$
i.e. heat is supplied to the gas to increase internal energy.

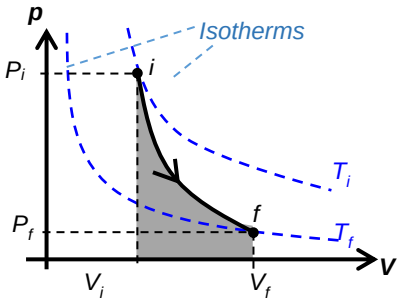
Adiabatic Process (No Heat Transfer Process)

As the name implies, $Q = 0$.

$\Delta U = W$

Magnitude of W is represented by the shaded area under the graph.

Two ways to achieve an adiabatic process is by *insulating* the system or performing the process *rapidly* (contrast this to an isothermal process).



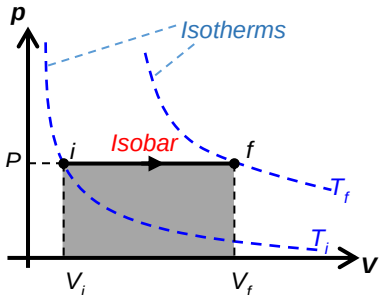
Example: The p-V graph shows gas expansion with no heat exchange ($Q = 0$).
$\Delta V > 0 \Rightarrow$ work done on gas is negative, $W < 0$
$T_f < T_i \Rightarrow \Delta U < 0$
Work is done by the gas and internal energy drops.

Isobaric Process (Constant Pressure Process)

An isobaric process is represented by a horizontal line on a pV diagram.

The magnitude of W can be computed from the rectangular area under the graph.

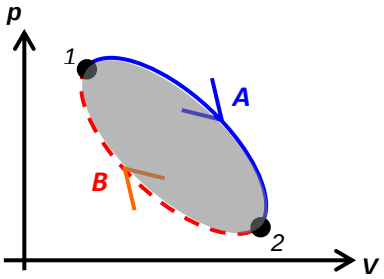
$\Delta U = Q \pm (area\ of\ rectangular\ area)$



Example: The p-V graph shows gas expansion at constant pressure.
$\Delta V > 0 \Rightarrow$ work done on gas is negative, $W < 0$
$T_f > T_i \Rightarrow \Delta U > 0$
Applying First Law: $Q = \Delta U - W$
Heat is supplied to the gas to increase internal energy and allow gas to expand, to do work against the atmosphere.

4.3 Cyclic Process

A series of processes that eventually returns a system to its initial state is called a cyclic process. On the PV diagram on the right, the system starts from state *1*, undergoes a process *A* to state *2* and goes back to state *1* via another process *B*.



Net Change in Internal Energy

Since the internal energy depends only on the state (i.e. the position on a PV diagram), the overall change in internal energy for a cyclic process must be zero; i.e. $(\Delta U)_{net} = (\Delta U)_{A\to B} + (\Delta U)_{B\to A} = 0$

Net Work Done on System

Since work done on the system W is given by the area under the graph, it is easy to see that

Process	Work done on the system, W
A (expansion)	W_A = – |Area under graph *A*|
B (compression)	W_B = + |Area under graph *B*|
A and B	W_{net} = – |Area enclosed in the cycle|

Net Heat Transfer

Applying the first law of thermodynamics to the cyclic process: $(\Delta U)_{net} = Q_{net} + W_{net}$

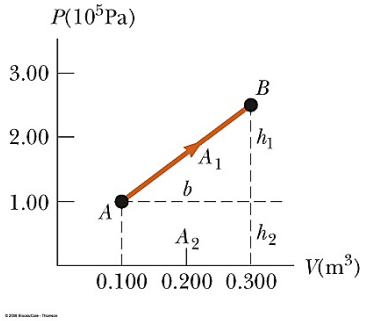
$$Q_{net} = -W_{net}$$

Hence for this cycle, there is a net heat transfer into the system. Deduce for yourself that if the cyclic process were in the opposite direction, there would be a net heat transfter out of the system.

Example 8

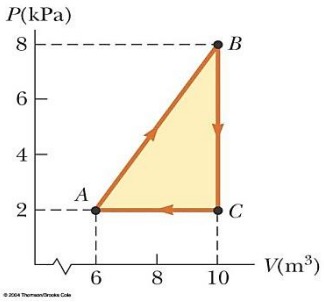
A quantity of 4.00 moles of a monatomic ideal gas expands from an initial volume of 0.100 m^3 and pressure of 1.0×10^5 Pa to a final volume of 0.300 m^3 and pressure of 2.5×10^5 Pa. Compute the

a) work done on the gas, [-35000 J]
b) change in internal energy of the gas, and [97500 J]
c) thermal energy transferred to the gas. [132500 J]



Example 9

A gas is taken through the cyclic process described in the figure. Find the net energy transferred to the system by heat during one complete cycle. [12000 J]



Tutorial 11C

Unless otherwise stated,

molar gas constant	R	$8.31\ J\ K^{-1}\ mol^{-1}$
Avogadro constant	N_A	$6.02 \times 10^{23}\ mol^{-1}$
Boltzmann constant	k	$1.38 \times 10^{-23}\ J\ K^{-1}$
Gravitational field strength	g	$9.81\ N\ kg^{-1}$
density of mercury		$13\ 600\ kg\ m^{-3}$

Self-Review Questions

S1. State the First Law of Thermodynamics. Is the First Law of Thermodynamics only valid on an ideal gas system?

S2. Describe how you would determine work done from a *PV* graph? How do you tell that positive work is done on the system? How do you tell that negative work is done on the system? What is the relationship between work done on the system and work done by the system?

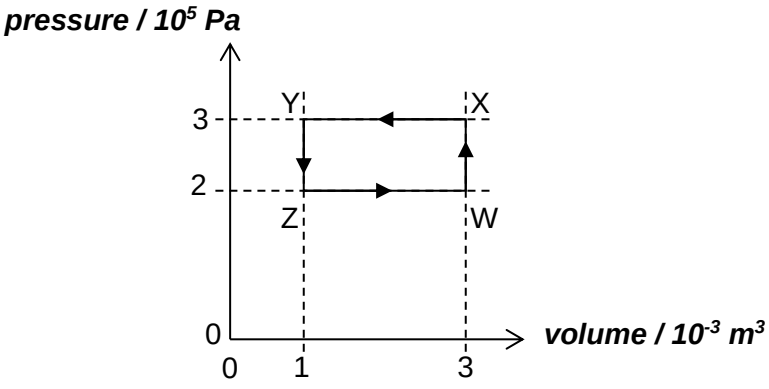
S3. Consider a mass of ideal gas. Of the three quantities stated in the first law of thermodynamics i.e. *U*, W and *Q*, which one cannot be determined directly from the *PV* graph.

S4. **a)** Consider a system of ideal gas,
- **i.** sketch on a graph the variation of *P* vs. *V* for the system at a fixed temperature *T*,
- **ii.** sketch on the graph that you have drawn in (i) two additional curves, to illustrate the new isotherms for (1) a higher temperature T_H, and (2) a lower temperature T_L.

b) Explain what is meant by an
- **i.** isochoric process
- **ii.** isobaric process
- **iii.** isothermal process
- **iv.** adiabatic process.

c) For each of the processes in (b), and by further considering a compression
- **i.** indicate the process clearly on an indicator diagram.
- **ii.** shade the area that indicates the work done on the system (if any). State also whether this work done on the system positive or negative.
- **iii.** for processes (i), (ii) in (b) state the equation or value of the work done on the system.

S5. **(N89/I/3)** A gas undergoes the cycle of pressure and volume changes W→ X → Y → Z → W shown in the diagram.



What is the net work done *by* the gas?

A -600 J **B** -200 J
C 0 J **D** 200 J
E 600 J

Discussion Questions

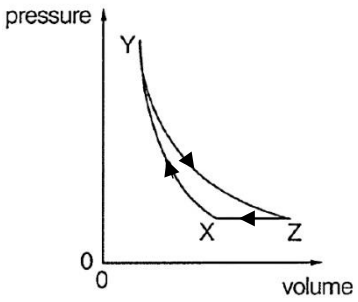
D1. (N09/I/21) The first law of thermodynamics may be expressed as $\Delta U = q + w$

where ΔU is the change in internal energy, q is the heating of the system and w is the work done on the system.

A fixed mass of ideal gas at high pressure is contained in a balloon. The balloon suddenly bursts, causing the gas to expand and cool. In this situation, which row describes the values of ΔU, q and w?

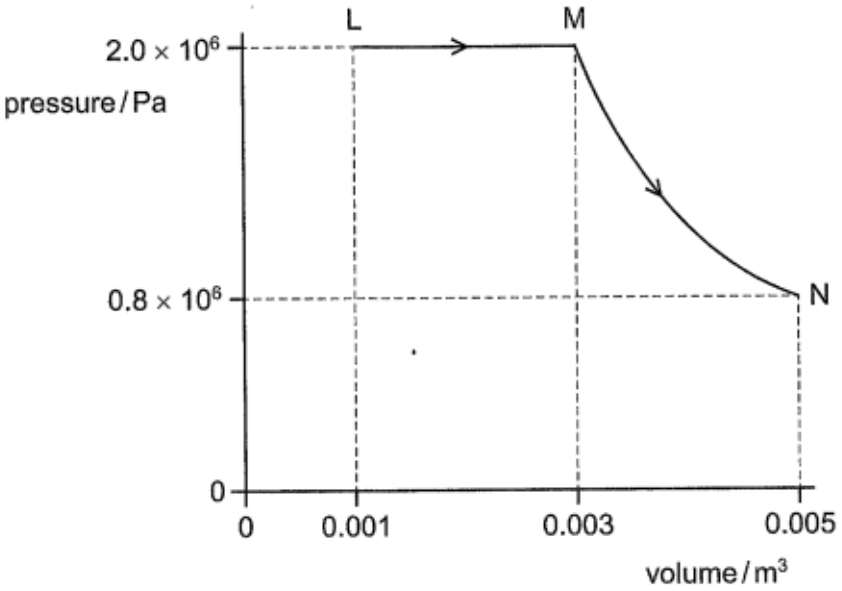
	ΔU	q	w
A	negative	negative	positive
B	negative	zero	negative
C	positive	zero	negative
D	positive	negative	positive

D2. N84/II/28. A fixed mass of an ideal gas undergoes the changes represented by XYZX.
Which one of the following sets could describe this set of changes?



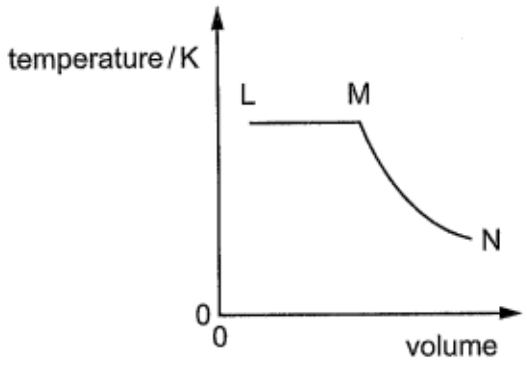
	XY	YZ	ZX
A	Isothermal expansion	Adiabatic compression	Compression at constant pressure
B	Adiabatic expansion	Isothermal compression	Pressure reduction at constant volume
C	Isothermal compression	Adiabatic expansion	Compression at constant pressure
D	Adiabatic compression	Isothermal expansion	Pressure reduction at constant volume
E	Adiabatic compression	Isothermal expansion	Compression at constant pressure

D3. **[N20/I/14]** A fixed mass of ideal gas undergoes changes of pressure and volume starting at L, as shown.

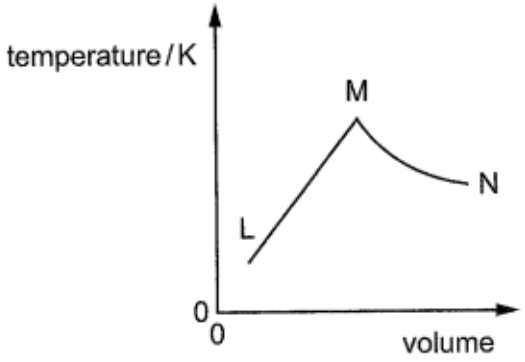


Which graph shows how temperature (measured in kelvin) changes with volume?

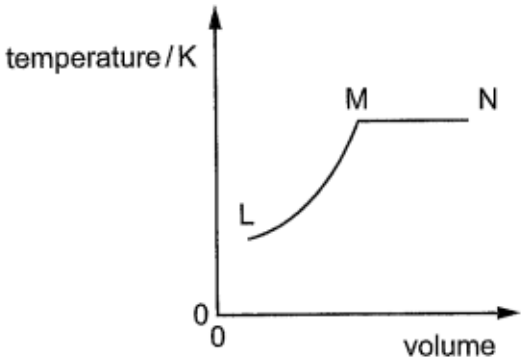
A



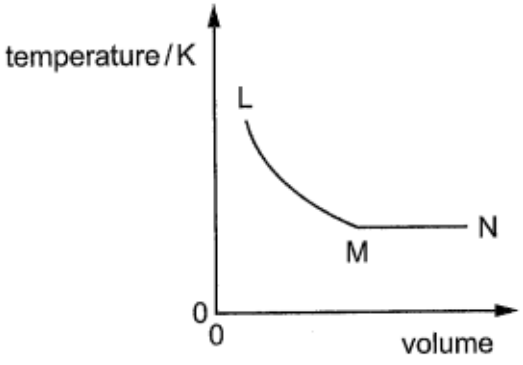
B



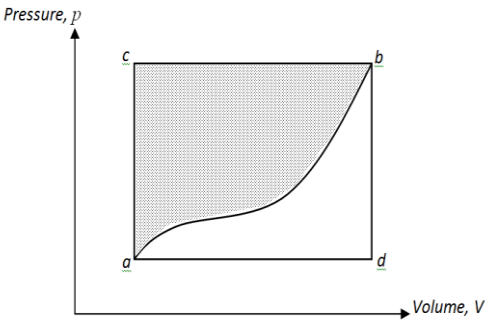
C



D



D4. The *p*-*V* diagram below represents the states of a system of gas. 75 J of heat flows into the system as it goes along path *acb*, while 25 J of work is done by the system. The shaded area represents a value of 12 J.



i. What change is there, if any, in the internal energy of the ideal gas after each cycle acbda? Explain your answer.

ii. Determine the amount of heat that flows into the system as it goes from *b* to *a* along the curved path.

D5. (N93/III/15 (part), N98/III/5(part), modified)

a) The first law of thermodynamics may be expressed in terms of the equation $\Delta U = Q + W$

i. Identify each of the terms in this equation.

ii. Some solids contract and some solids expand when they melt. Complete the following table with the symbols + or – to indicate the *signs* of the thermodynamic quantities for each of the two types of solid when the solids melt at constant pressure. [6]

	Solids which *contracts* on melting	Solids which *expands* on melting
ΔU		
Q		
W		

b) Use the following physical data for ice, water and steam when necessary in answering this question.

	Ice	***Water***	***Water***	***steam***
Temperature / °C	0	0	100	100
Volume occupied by 1 kg at standard pressure / m^3	0.00109	0.00100	0.00104	1.67
Kinetic energy of all the molecules in 1 kg /10^5 J	1.89	1.89	2.58	2.58
Potential energy of all the molecules in 1 kg (referred to ice at 0°C) / 10^5 J	0	3.36	3.41	24.3
Internal energy of 1 kg / 10^5 J	1.89	5.25	5.99	26.9

i. Explain why there is no change in kinetic energy of the molecules when ice at 0 °C changes to water at 0 °C. [2]

ii. What determines the internal energy of 1 kg of substance? [2]

iii. Determine the specific latent heat of fusion of ice. [2]

iv. Calculate how much work has to be done by 1 kg of water in order to change to steam at 100°C and at atmospheric pressure of 1.01 x 10^5 Pa. [3]

v. Use the first law of thermodynamics to calculate the specific latent heat of vaporisation. [3]

vi. Examining your calculation in (v), explain qualitatively why the specific latent heat of vaporisation of an object is higher than the specific latent heat of fusion. [2]

D6. (N08/P3/Q4)

a) State **i.** the meaning of the term *internal energy,* [2]

ii. the first law of thermodynamics. [1]

b) An ideal gas undergoes a cycle of changes A → B → C → A, as shown in Fig.4.1.

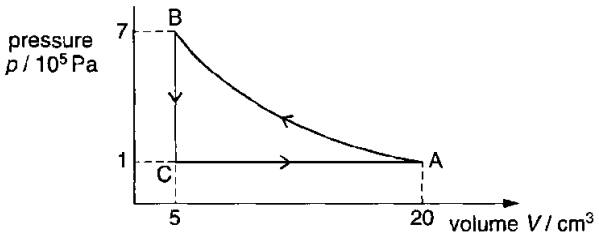


Fig. 4.1

section of cycle	heating supplied to gas / J	work done on gas / J	increase in internal energy of gas / J
A → B	zero	4.2	
B → C	– 8.5		
C → A			

Fig. 4.2

i. Calculate the work done by the gas during the change C → A. [2]

ii. Fig. 4.2 is a table of energy changes during one cycle. Complete Fig. 4.2. [4]

D7. (N17/P3/Q8)

(c)(i) By reference to the energy of molecules, explain why the internal energy of an ideal gas is proportional to its thermodynamic temperature. [4]

(ii) Some neon-20 gas is heated at constant volume so that its pressure and temperature change as shown in Fig. 8.3.

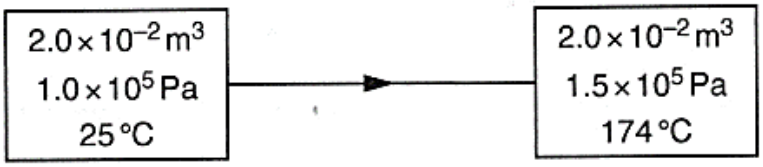


Fig. 8.3

The neon-20 gas is a monatomic gas that may be assumed to behave as an ideal gas. For the change illustrated in Fig. 8.3 to occur, 1220 J of thermal energy is supplied to the gas.

1. Show that the mass of gas remains constant during the heating process. [3]
2. Determine the specific heat capacity of the gas when measured at constant volume. [3]

(iii) Use the first law of thermodynamics to suggest why the specific heat capacity, as calculated in (ii), would be larger when the heating takes place at constant pressure, rather than at constant volume. [3]

D8. (N11/P3/Q2)

A mass of 5.6 g of nitrogen gas occupies a volume of 4.6×10^3 cm^3 at a pressure of 1.0×10^5 Pa and a temperature of 280 K. The specific heat capacity of nitrogen gas, when heated at constant volume, is 0.73 J g^{-1} K^{-1}. Nitrogen may be assumed to be an ideal gas.

(a) Calculate the thermal energy required to raise the temperature of the gas by 1 K at constant volume.

(b) The temperature of the gas is raised from 280 K to 281 K at constant pressure.
Determine, for this temperature change,
(i) the change in volume of the gas,
(ii) the external work done by the gas

(c) **(i)** State the first law of thermodynamics, indicating the directions of all energy changes.
(ii) Use the first law and your answers in (a) and (b) to determine the thermal energy required to raise the temperature of the nitrogen gas from 280 K to 281 K when the gas is heated at constant pressure.

Numerical answers to selected questions

D4	D5	D6	D7	D8
– 63 J	biii) 3.36 x 10^5 J kg^{-1} biv) 1.69 x 10^5 J bv) 2.26 x 10^6 J kg^{-1}	bi) 1.5 J	cii)2. 0.507 J g^{-1} K^{-1} or 507 J kg^{-1} K^{-1}	a) 4.1 J bi) 16 cm^2 bii) 1.6 J cii) 5.7 J

Appendices

(The following content has been covered in the 'O' level syllabus and it no longer appears in the learning objectives of 9749 H2 Physics syllabus)

Appendix A: Temperature Measurement & Temperature Scales

Thermometric Properties

Thermometers are devices for measuring temperature. All thermometers are based on some material's thermometric property. A thermometric property is a physical quantity which varies continuously with temperature in a repeatable manner.

Thermometer	Thermometric Property	Advantages (+) / Disadvantages (–)	
Liquid-in-glass thermometer	Length of liquid in a capillary	+	Portable, cheap, simple to use, direct reading
		–	Not very accurate, fragile
Thermocouple thermometer	E.m.f. between junctions of a wire at different temperatures	+	Wide range, small, robust and compact, quick response, widely used in industry for measuring rapidly changing temperatures and temperatures at a 'point'.
		–	Less accurate than constant volume gas & platinum resistance thermometer. Fairly accurate
Thermistor	Resistance of a semiconductor	+	Can be linked directly to other circuits or computer.
		–	Limited range, not very accurate, not lasting.
Platinum resistance thermometer	Resistance of a coil of wire	+	Wide range, highly accurate, sensitive.
		–	No direct reading, slow in response, not suitable for varying temperatures.
Constant-volume gas thermometer	Pressure of a fixed mass of gas at constant volume	+	Very wide range, very accurate and sensitive, can be used as a standard to calibrate other thermometers.
		–	Very bulky and slow to respond, not suitable to varying temperatures, no direct reading.
Pyrometer	Colour displayed will vary with temperature	+	for temperature above 1500 ^{o}C.
		–	Very judgmental and can only give an approximate range of the temperature.

Other than thermometric property, a thermometer also need two more components to produce meaningful indication of temperature:

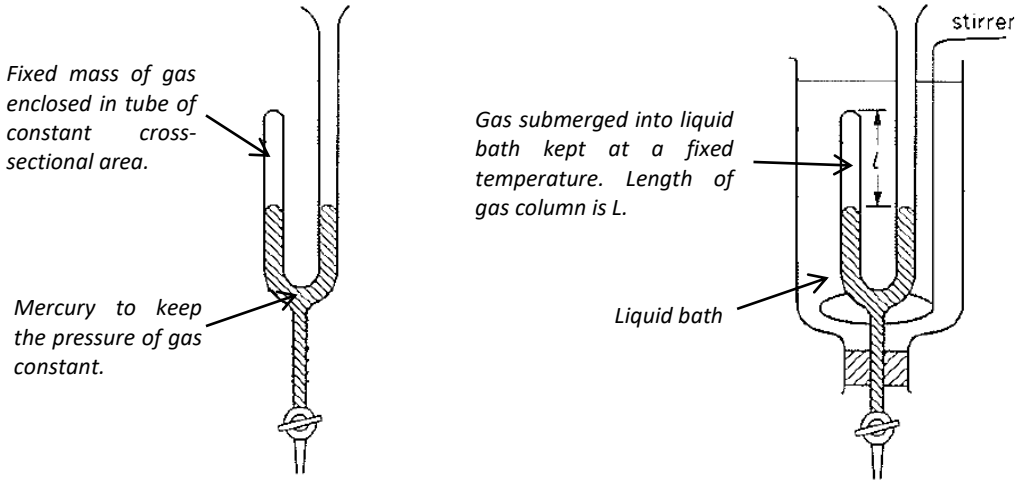
i) some references known as fixed points.
ii) a set of number assigned to the thermometric property

Once these three components are in place, we have established a **Temperature Scale**, a set of number (with appropriate units) that is commonly understood to represent certain temperatures. For example a number 40 (degree centigrade) that appears in your oral digital thermometer after you have taken out of your mouth indicates that your fever has not subsided.

Appendix B: Empirical Scale

Calibrating A Thermometer To Read Temperature On Centigrade Scale

We will first illustrate one simple temperature scale here and explain how it is established.

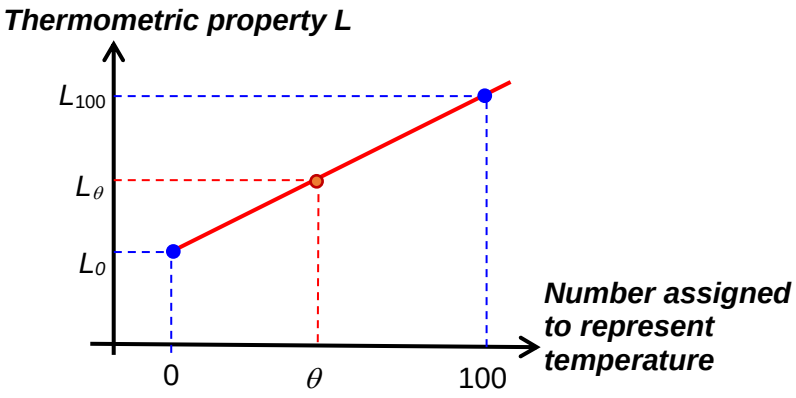


i) In 1781, Jacques Charles discovered that under certain conditions, the volume of a given mass of gas is related to its temperature if its pressure is constant. This is the well-known Charles's law. Hence the volume of such a gas is a thermometric property. If this gas is kept in a tube of constant cross-section area as shown, the length L of the gas is related to the temperature.

ii) Allow this system to reach thermal equilibrium with pure melting ice. Measure the length of gas and called it L_0. We can arbitrarily assign a number to represent the temperature of this pure melting ice but historically people have commonly use the number zero. The temperature represented by this number zero is known as ice point. This is one of the fixed points. So we say that the thermometric property has the length L_0 at 0.

iii) We need another reference (fixed point). Allow this system to attain thermal equilibrium with steam from boiling water at standard atmospheric pressure at sea level (defined as 101.325 kilopascals). Measure the length of gas and let it be L_{100}. For the same reason, we assign number 100 to represent this temperature which is also known as steam point. The thermometric property has length L_{100} at 100.

iv) We now have information on the thermometric properties at two different situations – melting ice and boiling water. How about any other temperature? We can of course use different objects at different temperatures to calibrate the length of the gas. But what if we don't. We can simply ASSUME a linear relationship joining the two references we have.



v) Using the tube of gas together with the calibration graph, we can tell the temperature of any unknown object θ by allowing this gas to attain thermal equilibrium with it, measure its length L_θ and read the corresponding temperature θ from the graph. It is not difficult to show mathematically that

$$\theta = \left(\frac{L_\theta - L_0}{L_{100} - L_0}\right) \times 100$$

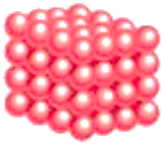
If we calibrate our thermometer in this way, the numbers that represent the temperature will take the unit degree centigrade and we say this thermometer measure temperature in ***Centigrade Scale***.

Other Common Temperature Scales

There are of course other temperature scales depending on how we calibrate the thermometers e.g. The Fahrenheit temperature scale is based on 32 °F for the freezing point of water and 212 °F for the boiling point of water, with the interval between the two being divided into 180 parts.

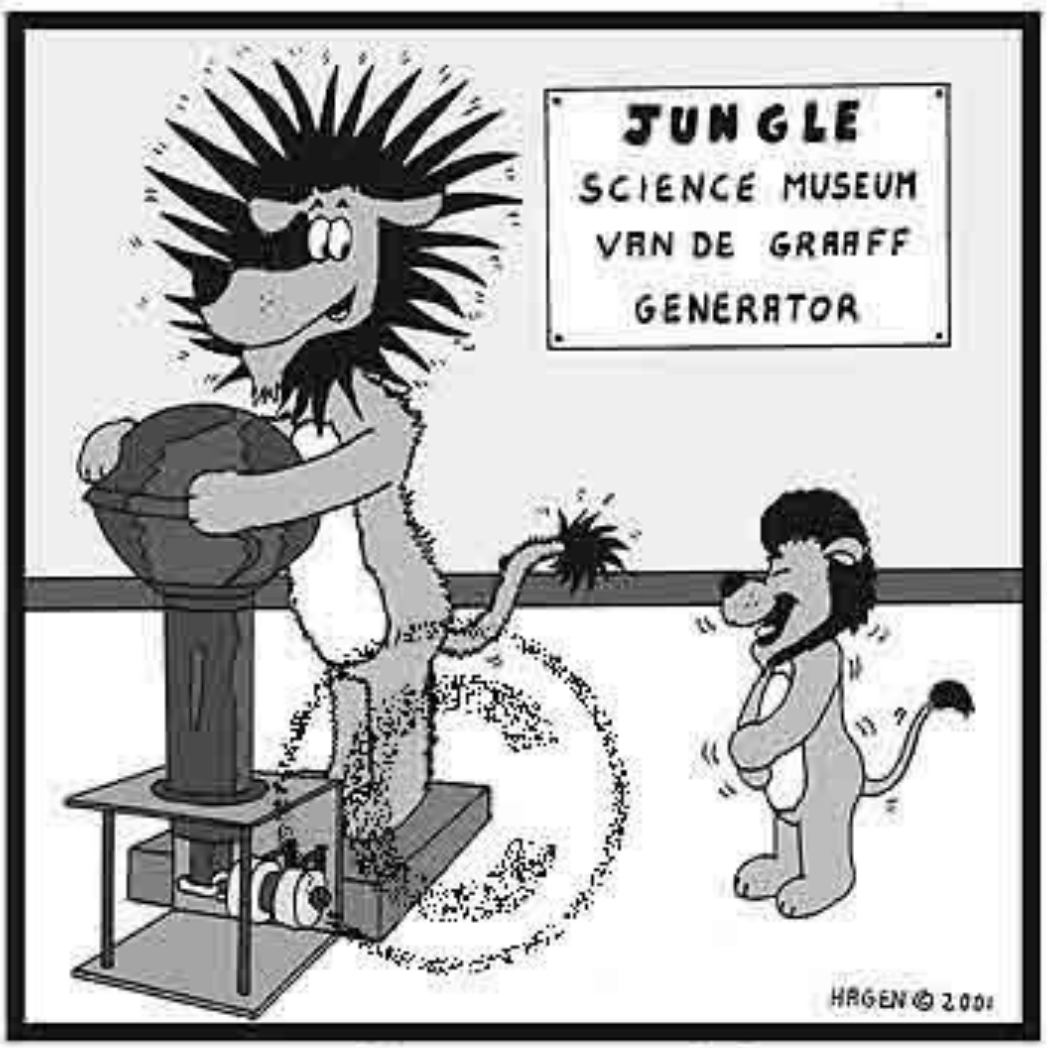
So long as the thermometer is calibrated based on experimental measurement of specific thermometric properties at certain fixed points, the temperature they read will be on the **empirical scales.** The word 'empirical' means 'based on experimental observation'.

Appendix C: Energy and Phases of Matter

PHASES OF MATTER	STRUCTURE AND CHARACTERISTICS	MOVEMENT OF MOLECULES – A KINETIC MODEL
Solid 	Atoms or molecules are closely packed a) Crystalline solid — atoms or molecules are compactly arranged in a regular and orderly lattice structure. Crystals have a high density. b) Non-crystalline solid — atoms or molecules are arranged in a more random way. Density is still high but generally lower than many crystalline solids.	Atoms or molecules are locked in positions constrained by intermolecular forces. They vibrate continually about their equilibrium positions. Molecular motion is minimum at zero kelvin. As temperature rises, molecular motion and kinetic energy increase.
	Has fixed shape and volume. Solids are not easily compressed or stretched by mechanical means.	When solids are stretched, atoms or molecules move further apart but are opposed by their attractive forces. When solids are compressed, atoms/ molecules move closer together but are opposed by their repulsive forces.
Liquid	Atoms or molecules are only slightly further apart than in solids. Density remains high.	Atoms or molecules vibrate, but in addition have random and restricted translational motion, since they occur in clusters.
	Volume is nearly constant but has variable shape. Liquids are not easily compressed by mechanical means.	Attractive forces between atoms or molecules are weaker. This allows for greater mobility which in turn explains why liquids can take the shape of any container. Attractive forces also manifest themselves as surface tension and viscosity of a liquid.
Gas	Atoms or molecules are very far apart. Densities of gases are very low.	Atoms or molecules move randomly at very high speed throughout the space they occupy. The average kinetic energy is directly proportional to the absolute temperature.
	Gases do not have a fixed volume or shape. They are highly compressible.	Atoms or molecules will occupy any available space in a container. Intermolecular forces between them are very weak as they are very far apart.

Chapter 12

Electric Fields



I told you Science was fun!

Playlist of Lecture Examples can be found at https://for.edu.sg/dfwq3m

Content

- Concept of an electric field
- Electric force between point charges
- Electric field of a point charge
- Uniform electric fields
- Electric potential

Learning Outcomes

Candidates should be able to:

(a) show an understanding of the concept of an electric field as an example of a field of force and define electric field strength at a point as the electric force exerted per unit positive charge placed at that point.

(b) represent an electric field by means of field lines.

(c) recognise the analogy between certain qualitative and quantitative aspects of electric field and gravitational field.

(d) recall and use Coulomb's law in the form $F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r^2}$ for the electric force between two point charges in free space or air.

(e) recall and use $E = \frac{Q}{4\pi\varepsilon_0 r^2}$ for the electric field strength of a point charge in free space or air.

(f) calculate the electric field strength of the uniform field between charged parallel plates in terms of the potential difference and plate separation.

(g) calculate the forces on charges in uniform electric fields.

(h) describe the effect of a uniform electric field on the motion of charged particles.

(i) define the electric potential at a point as the work done per unit positive charge in bringing a small test charge from infinity to that point.

(j) state that the field strength of the electric field at a point is numerically equal to the potential gradient at that point.

(k) use the equation $V = \frac{Q}{4\pi\varepsilon_0 r}$ for the electric potential in the field of a point charge, in free space or air.

Contents **Pg**

H2 Physics Syllabus 9749 2

12.1 Review of Basic Concepts 4
12.2 The Electric Field 6
12.2.1 Electric Field Strength 6
12.2.2 Electric Field Lines 8
12.2.3 Force between Two Point Charges (Coulomb's Law) 10
12.2.4 Electric Field around a Point Charge 12
12.3 Electric Potential Energy and Electric Potential 14
12.3.1 Electric Potential Energy 14
12.3.2 Electric Potential 16
12.4 Relation between Electric Potential Energy and Electric Force 19
12.4.1 Relation between Potential Gradient and Electric Field Strength 19
12.4.2 Graphs of Electric Field Strength and Electric Potential between Point Charges 20
12.5 Representation of Electric Fields by Equipotential Lines or Surfaces 23
12.5.1 Charges in Equilibrium on Conductors 24
12.6 Electric Field between Two Charged Parallel Plates 26
12.6.1 Motion of Charged Particles in a Uniform Electric Field 26
12.7 Analogy between Electric fields and Gravitational Fields 28

Appendices 29

Tutorial 31

12.1 Review of Basic Concepts

Electrical Charges

Many of us could have encountered such an experiment in our secondary school years. When a glass rod that has been rubbed with a silk cloth is brought near another rubber rod which has been rubbed with fur, they attract each other. However, when two rubber rods which have been rubbed with fur are brought together, they repel each other (Fig. 1).

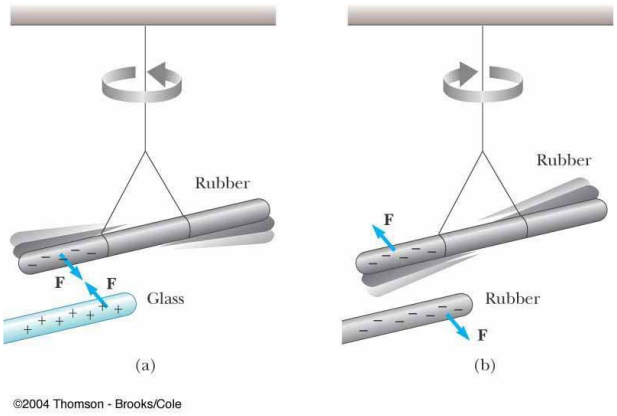


Fig. 1. (a) A negatively charged rubber rod is attracted to a positively charged glass rod. (b) A negatively charged rubber rod is repelled by another negatively charged rubber rod. (Jewett & Serway, 2008)

Using similar experiments like the one described above, Benjamin Franklin (1706 – 1790) showed that:

- **There are two kinds of charges** - **positive charge and negative charge.** These labels are *arbitrary*; i.e. there is no particular reason why positive charges are called positive charges and not negative charges. (It was later discovered that electrons carry negative charge while protons carry positive charge.)
- **Charges of the same sign repel one another. Charges with opposite signs attract one another.**

Atomic Structure

All materials are made up of atoms. The structure of atoms can be described in terms of three particles – the negatively charged electron, the positively charged proton and the neutral neutron.

The protons and neutrons in an atom make up a small, very dense core called the **nucleus**, which is approximately 10^{-15} m in size (Fig. 2). Surrounding the nucleus are electrons, extending out to distances of approximately 10^{-10} m from the nucleus. These negatively charged electrons orbit the nucleus and the attractive electrical force on the electrons by the protons provides the force required to keep the electrons in orbit.

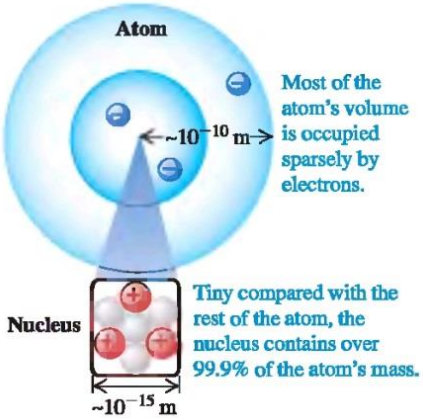


Fig. 2. The structure of an atom. The particular atom depicted here is lithium. (Young & Freedman, 2007)

The **electrical charge** and the **mass** of the electrons, protons and neutrons are given in the table below:

	Charge	Mass
Electron	$-e = -1.60 \times 10^{-19}$ C	9.11×10^{-31} kg
Proton	$+e = 1.60 \times 10^{-19}$ C	1.67×10^{-27} kg
Neutron	No charge (0 C)	1.67×10^{-27} kg

*e is known as the fundamental charge or the elementary charge.

Neutral Atoms, Positive Ions and Negative Ions

A **neutral atom** is one in which the **number of electrons equals the number of protons**. If one or more electrons are removed from this atom, the atom becomes positively charged and is known as a **positive ion**. A **negative ion** is an atom that has *gained* one or more electrons. This process of gaining or losing electrons is called **ionisation**.

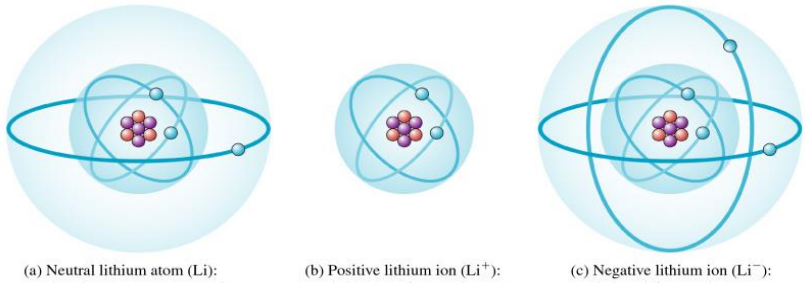


Fig. 3. (a) A neutral lithium atom has equal number of protons and electrons. (b) A positive lithium ion has a deficit of electrons. (c) A negative ion has an excess of electrons. (Young & Freedman, 2007)

Furthermore, when the total number of protons in a body equals the number of electrons in the body, the total charge is zero and the body is on the whole electrically neutral. To charge a body negatively, we add electrons to the originally neutral body so that it has excess negative charges. Similarly, to positively charge a body, we can remove electrons. Hence, when we speak of charge of a body, we always mean its net charge.

Principles of Conservation and Quantisation of Charge

Implicit to the discussion above are two principles:

- Principle of Conservation of Charge
- Principle of Quantisation of Charge

The *principle of conservation of charge* states that charge cannot be created or destroyed. Hence, for any closed system, the algebraic sum of all electric charges must be constant. For example, suppose we rub a plastic rod with a piece of fur, both initially uncharged. The plastic rod acquires a positive charge and the fur a negative charge. The principle of conservation of charge tells us that the positive charge acquired by the rod must be equal in magnitude to the charge on the fur. The rod has acquired a positive charge as it has transferred its electrons to the fur. Hence, in the charging process, charge is not created or destroyed; it is merely *transferred* from one body to another.

The *principle of quantisation of charge* states that electric charge always occurs as integral multiples of a fundamental amount of charge *e* which is also known as the elementary charge. Charges are therefore said to be quantized[1], i.e. any charge q must be equal to $\pm Ne$ where *N* is some positive integer. This was discovered in 1909 by Robert Millikan (1868 – 1953)[2].

[1] "Quantum" means "smallest discrete quantity". Thus, if something is quantised, it comes in (multiples of) *fixed amounts.*
[2] See https://en.wikipedia.org/wiki/Oil_drop_experiment

12.2 The Electric Field

How do objects not in contact with each other affect each other's behaviour? How does a proton exert a pull on an electron at a distance? To explain 'action at a distance', physicists developed the idea of a **field**. You have learnt about the gravitational field that results from the presence of a mass. In this chapter you will learn about the electric field, which comes about due to the presence of charge.

The electric force between charged objects is an example of action at a distance. One charged object exerts a force on another even though they are not in physical contact. Object A sets up a (force) field in the space around it. When a second object B is in the field set up by A, object B experiences a force (Fig. 4). Just as object B experiences a force as it is in the field set up by A, it also generates an electric field around itself and hence A experiences a force by B as it is in the electric field set up by B.

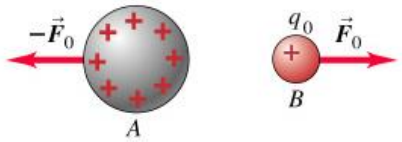


Fig. 4. When the charged object B is placed in the electric field set up by charged object A, it experiences a force. Similarly, A experiences a force as it is in the field of B. (Young & Freedman, 2007)

Electric Field

An electric field is a **region of space** where a **charge** will experience an **electric force**.

12.2.1 Electric Field Strength

Definition: Electric Field Strength, E

The electric field strength E **at a point** is the electric **force per unit *positive* charge** acting on a **small test charge** placed at the point.

$$\boldsymbol{E} = \frac{\boldsymbol{F}}{q}$$

- Electric field strength is a **vector quantity**.
- The *direction of the electric field* at a point is the **direction of the electric force experienced by a positive test charge** if placed at that point.
- The S.I. unit of electric field strength is newton per coulomb (N C^{-1}).

Rearranging the equation above,

$$\boldsymbol{F} = q\boldsymbol{E}$$

Therefore:

- the **magnitude of the force** on the charge placed in the field is given by $F = |qE|$
- The **direction of the force on the charge is the same as that of** the electric field at the point if the charge $\boldsymbol{q}$ **is positive**.
- The **direction of the force on the charge is opposite to that of** the electric field if the charge $\boldsymbol{q}$ **is negative**.

Example 1

An isolated electron (in a vacuum) near the surface of the earth was found to be accelerating directly away from the surface with an acceleration a = 2.63 x 10^6 m s^{-2}.
[elementary charge, e = 1.60 x 10^{-19} C, mass of the electron, m_e = 9.11 x 10^{-31} kg]

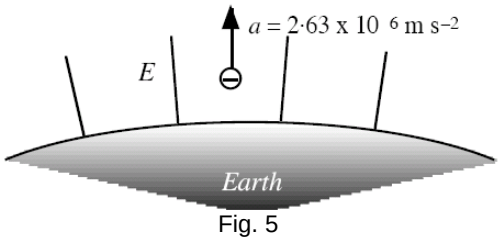


Fig. 5

(a) What is the direction of the electric field near the surface of the earth at the location of the electron?

(b) Calculate the magnitude of the electric field near the surface.

12.2.2 Electric Field Lines

Just as for a gravitational field, field lines are used to represent an electric field.

The electric field lines must be able to represent the variations in the **magnitude** and **direction** of the electric field in a region of space:

- **Magnitude of the Strength of the Electric Field:**
  - The *number of field lines per unit area,* through a surface perpendicular to the lines, is proportional to the *magnitude* of the electric field.
  - Hence, where the lines are close together, the electric field is strong, while where the lines are far apart, the field is weak (Fig. 6).

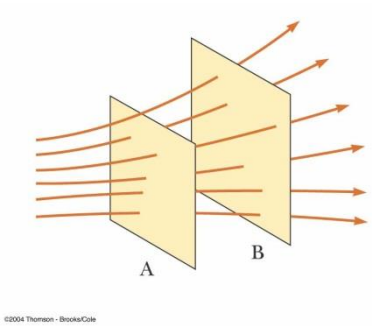


Fig. 6. Electric field lines penetrating two surfaces. The magnitude of the field is greater on surface A than on surface B. (Jewett & Serway, 2008)

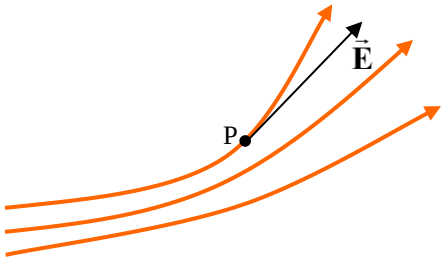


Fig. 7. The electric field at point P is tangent to the field line through P.

- **Direction of the Electric Field:**
  - The **direction of the electric field strength E at a point in an electric field is tangent to the electric field line at that point** (Fig. 7). It is the direction of the force experienced by a positive charge if placed at that point.

<u>Rules for Drawing Electric Field Lines[3]:</u>

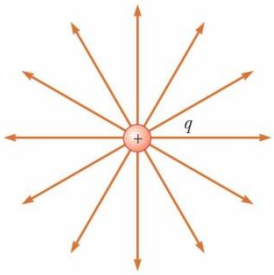
i. Electric field lines are directed from higher potential (more positive) to lower potential (more negative).

ii. The number of lines drawn leaving a positive charge or approaching a negative charge is proportional to the magnitude of the charge.

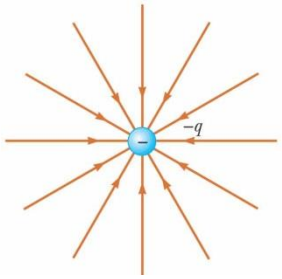
iii. Field lines do not intersect because $\boldsymbol{E}$ at a point can only point in one direction.

[3] There are also rules associated with drawing field lines from equipotential lines. We will cover that in a later section and summarize all the rules there (see pg 25)

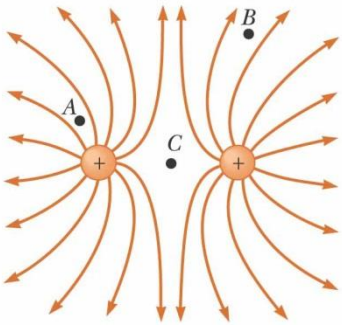
Fig. 8 (a) to (f) illustrate the field patterns of some common charge systems.



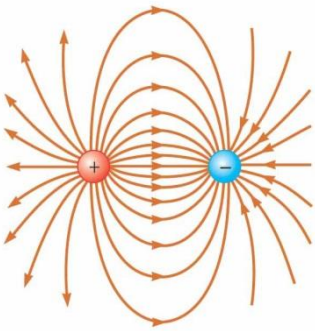
(a) Single positive charge



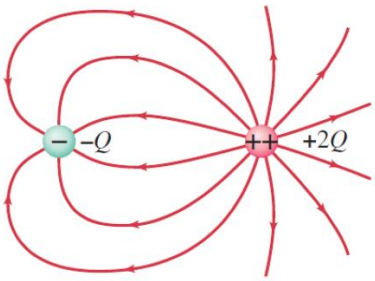
(b) Single negative charge



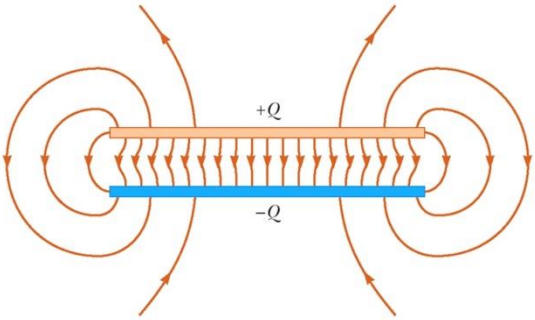
(c) Two positive charges equal in magnitude



(d) Two opposite charges equal in magnitude



(e) Two opposite charges, the magnitude of one charge being twice the magnitude of the other
(note the number of lines at each charge)



(f) A pair of oppositely-charged parallel plates[4]

Fig. 8. Electric field of some common charge systems. (Jewett & Serway, 2008)

Electric field lines are not real. They are only used as a pictorial representation to provide a qualitative description of the electric field. Only a finite number of lines from each charge can be drawn, which makes it appear as if the field only exists in certain parts. The field, in fact, is continuous, existing at every point in space.

[4] The charged parallel plates are said to produce a uniform electric field between them (indicated by the parallel lines between the plates), except near the edges, where there are *fringe effects*. This is because the charges at the edges are not distributed uniformly, unlike the other charges that are nearer the centre of the plates.

12.2.3 Force between Two Point Charges (Coulomb's Law)

As stated earlier, when a charge Q_1 is placed near another charge Q_2, each of them will experience a force as they are in the field set up by the other charge. If Q_1 and Q_2 are point charges, the magnitude of the force can be determined by **Coulomb's Law**.

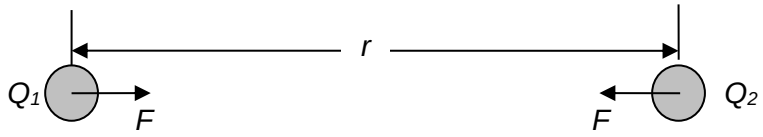


Fig. 9. The force *F* between two point charges is determined by Coulomb's Law.

Coulomb's Law

The magnitude of the electrical force acting between two point charges is proportional to the product of the magnitude of the charges and inversely proportional to the square of the distance between them.

$$F = \frac{1}{4\pi\varepsilon_0}\frac{Q_1Q_2}{r^2}$$

where F : magnitude of the force between charges Q_1 and Q_2

ε_o : permittivity of free space, and

r : the distance between Q_2 and Q_2

Notes:

(i) There is one force acting on Q_1 and another force acting on Q_2. These two forces are an *(Newton's Third Law) action-and-reaction pair*.

(ii) Permittivity:

In the S.I. system of units, the constant of proportionality in Coulomb's Law is usually written as $1/4\pi\varepsilon$. The constant ε is known as *permittivity*. Permittivity is a property of the medium, i.e. its value depends on the medium in which the charges are placed. If the space between the charges is '*free space*' (or *vacuum*), its permittivity (known as permittivity of free space) will be denoted as ε_0.

$$\varepsilon_o = 8.85\times10^{-12}\ \text{F m}^{-1}$$

The S.I. unit for permittivity is farad (F) per meter (m). Farad is the unit for capacitance (which is not in the H2 syllabus). 1 F = 1 C V^{-1}.

For most practical purposes the permittivity of 'air' is the same as that of 'free space'.

When using Coulomb's Law, just **substitute the magnitude of the charges** (without the signs) to find the **magnitude of the force.**

Indicate the direction of the force (or whether it is attractive or repulsive) separately.

What happens if there are more than two charges?

The force between any two charges in a group of charges is independent of the presence of the other charges. Hence, when more than two charges are present, the force between any pair of them is still governed by Coulomb's law.

> **The resultant force on any one of them is equal to the vector sum of the forces exerted by the other charges.**

For example, if four charges are present, the resultant force exerted by particles 2, 3 and 4 on particle 1 is given by

$$\vec{F} = \vec{F}_{12} + \vec{F}_{13} + \vec{F}_{14} .$$

Example 2:

Two small equal charges 2 x 10^{-8} C, one positively and the other negatively charged, are placed at A and B as shown in Fig. 10. AB is 6 cm.

Calculate the force acting on a charge +1 x 10^{-8} C placed at P, where P is 5 cm from both charges.

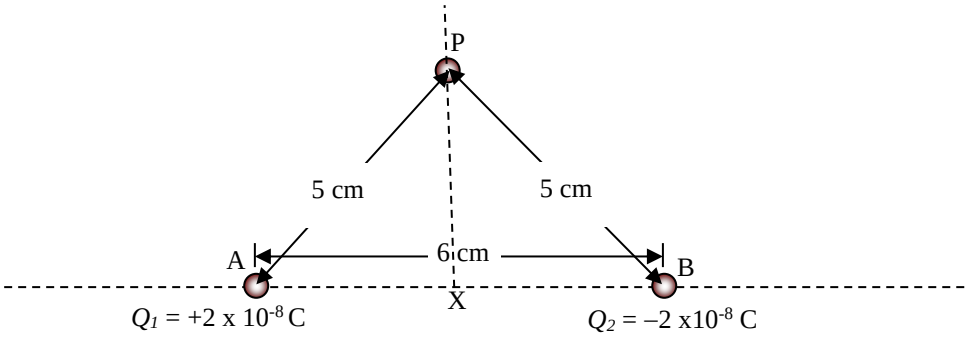
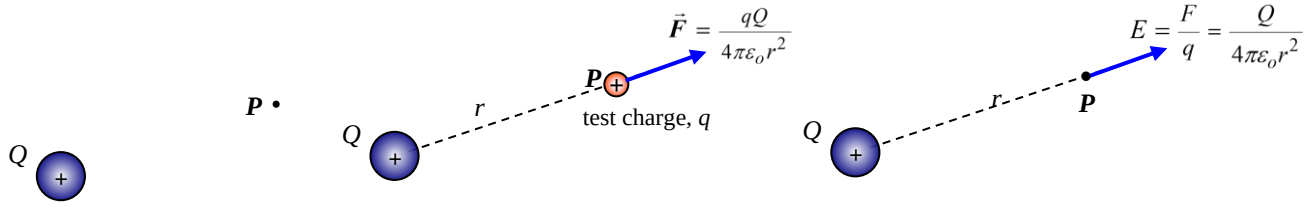


Fig. 10.

12.2.4 Electric Field around a Point Charge

With the electric field strength $\boldsymbol{E}$ defined, and *Coulomb's Law*, we can now quantitatively describe the electric field around a point charge.

Consider an isolated point charge Q that is placed in free space. An electric field is set up by Q, in the space around it. If a small positive charge q is brought into the space and is placed at a distance r from Q, a force F will be exerted by the field of Q on the charge q.



(a) What is the electric field at P due to a point charge Q?

(b) A test charge q is placed at P and the force on the test charge is measured.

(c) The electric field at P is the force per unit positive charge exerted by Q on the test charge at P.

Fig. 11. Determining the electric field of a point charge

According to Coulomb's law, the magnitude of repulsion or attraction between Q and q is

$$F = \frac{1}{4\pi\varepsilon_0}\frac{Qq}{r^2}$$

Hence, the magnitude of the electric field strength at a distance r away from the point charge Q,

$$E = \frac{F}{q}$$

$$\Rightarrow \quad \boxed{E = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}}$$

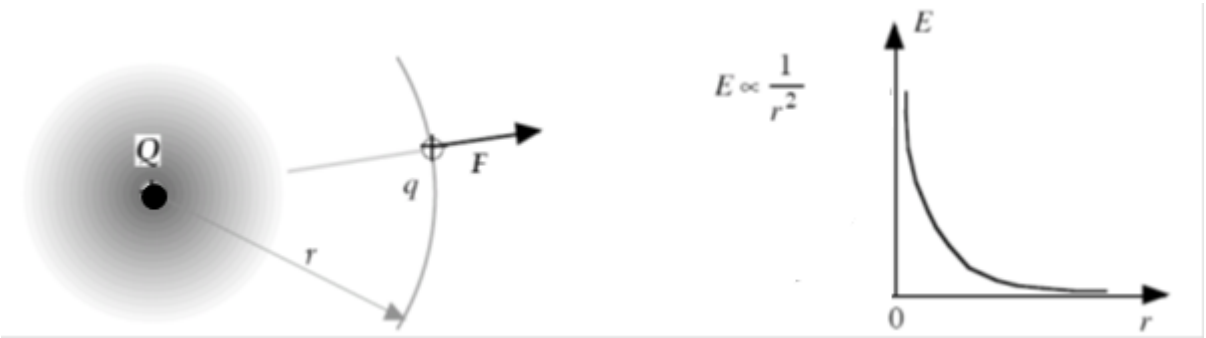


Fig. 12. Electric field of a point charge Q

Notes:

(i) $\boldsymbol{E}$ is a vector quantity, and its direction at a point is given by the direction of the force experienced by a positive charge if it is placed at that point.

(ii) In calculating the electric field strength, **just substitute the magnitude of the charge** (without the sign). **Indicate the direction separately.**

(iii) The field is radial for a point charge. It is directed uniformly in all directions outward from the centre if Q is a positive charge and inward toward the centre if Q is a negative charge. At all points that are an equal distance away from Q the magnitude of E is the same.

What if there are multiple charges?

The resultant electric field $\vec{E}$ at a point *P* in an electric field is the *vector sum* of the fields at *P* due to each point charge in the system.

$$\vec{E} = \vec{E}_1 + \vec{E}_2 + \vec{E}_3 + ...$$

Example 3

A charge +1.4 μC and a charge +2.6 μC are separated at 15 cm from each other.
At what point P, other than a point at infinity, is the resultant electric field zero? *This point is known as the neutral point in the field between the two charges.*

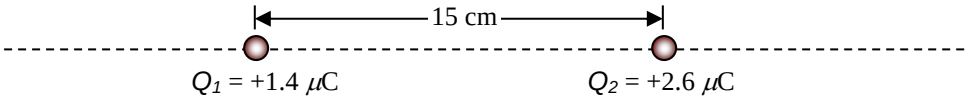


Fig. 13.

12.3 Electric Potential Energy and Electric Potential

12.3.1 Electric Potential Energy, *U*

The electric potential energy, U, of a system of charges is the **work done** by an external force in assembling the system where the charges were originally at infinity, **without any change in the kinetic energy** of the charges.

Potential Energy of a System of Two Point Charges

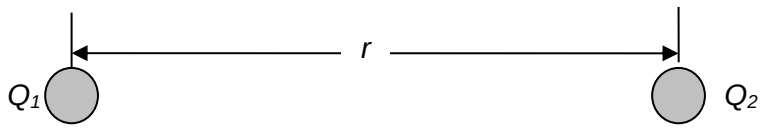


Fig. 14. Two charges Q_1 and Q_2 are separated by a distance *r*.

The potential energy U of a system of two point charges Q_1 and Q_2 separated by a distance *r* is given by

$$U = \frac{Q_1 Q_2}{4\pi\varepsilon_o r}$$

Notes:

(i) U is a scalar quantity. When calculating its value, **substitute the sign of the charges** into the expression (unlike for force and field).

(ii) If Q_1 and Q_2 are both positive (or both negative) the electric force is repulsive and the external force (which is opposite in direction to the electric force) would be in the same direction as the displacement from infinity to the position set and hence does positive work to bring the charges to the position set. *U* is positive.

(iii) If Q_1 and Q_2 are of opposite signs, the electric force is attractive. The external force will do negative work since its direction will now point towards infinity and is opposite to the direction of displacement from infinity. *U* is negative. *Note the similarity of this situation with gravitational potential energy.*

What is the potential energy of a system of more than two point charges?

Consider the system of point charges shown in the diagram below. Let us determine the expression for the potential energy of this system.

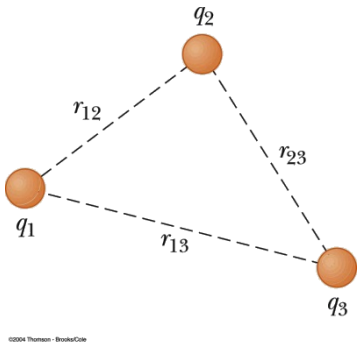


Fig. 15. Potential Energy of the System

Recall that potential energy of the system refers to the total amount of work done needed to be done by an external agent to assemble the charges (initially infinitely far apart) at positions as shown *without a change in the kinetic energy* of the charges. This calculation can be done by taking the sum of the work done to position one charge at a time.

Step One: Bring q_1 from infinity to the point. As there is no electric field initially in the region,

$\Rightarrow$ Work done, $W_1 = 0$ J

Step Two: Bring q_2 from infinity to the point. As q_1 is already in place, it will have set up an electric field in the region and an external force will need to do work against the electrical force experienced by q_2 due to its interaction with the field set up by q_1. Hence,

$$\Rightarrow \quad \text{Work done, } W_2 = \frac{q_1 q_2}{4\pi\varepsilon_o r_{12}}$$

Step Three: We bring q_3 from infinity to its position in the electric field. Since q_1 and q_2 are now in place, the electric field in the region is now due to q_1 and q_2, and the external force will need to do work against the force on q_3 due to its interaction with the field set up by both q_1 and q_2. Hence, the work done in this process is

$$\Rightarrow \quad \text{Work done, } W_3 = \frac{q_1 q_3}{4\pi\varepsilon_o r_{13}} + \frac{q_2 q_3}{4\pi\varepsilon_o r_{23}}$$

The potential energy of the system is the work done by an external agent to assemble the system:

$\Rightarrow$ **Potential energy of this system,** $U = W_1 + W_2 + W_3$

$$= \frac{q_1 q_2}{4\pi\varepsilon_o r_{12}} + \frac{q_1 q_3}{4\pi\varepsilon_o r_{13}} + \frac{q_2 q_3}{4\pi\varepsilon_o r_{23}}$$

12.3.2 Electric Potential

Definition: Electric Potential, *V*

The electric potential at a point in an electric field is the **work done per unit *positive* charge**, by an external force, in bringing a **small test charge** from infinity to that point, **without any change in kinetic energy**.

$$V = \frac{W}{q}$$

The SI unit of potential is J C^{-1} or more commonly, the volt V.

Notes:

(i) Electric potential, *V* is a scalar quantity.

(ii) The work done in this definition refers to the work done by the external force that is needed to overcome the electric force that acts on the test charge and is commonly referred to as the "work done on the system". Some questions in A-levels will ask you to look for "work done by the system" instead – this would refer to the work done by the electric force on the test charge. You can simply find it by using:

Work done on system (by external force) = - Work done by system (by the electric force)

Electric Potential due to a Point Charge

Imagine a field created by a point charge *Q*. If a test charge *q* were brought from infinity to a distance *r* from *Q*, the work done by the external agent is given by $W = \frac{Qq}{4\pi\varepsilon_o r}$.

The electric potential in the field of a point charge[5] *Q* is given by

$$V = \frac{W}{q} = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r}$$

where *V* is the potential at a point in the electric field, at a distance *r* away from the point charge *Q*.

Note:

Substitute for the sign of the charge into the expression. If *Q* is positive, then the potential *V* around it is positive. If *Q* is negative, then the potential *V* around it is negative.

[5] The derivation can be referred to in Appendix A.

If I have two or more point charges, how do I determine the potential at a point in the electric field set up by these point charges?

Since the electric potential is a scalar quantity, the potential at a point in such an electric field is the ***scalar sum*** of the potentials contributed by each charge at the point.

$$V = V_1 + V_2 + V_3 + ...$$

Remember to keep the signs in your substitution.

Example 4

Four charges are placed at the corners of a square as shown. Each side has a length *a*. Find the resultant electric potential at P, the centre of the square.

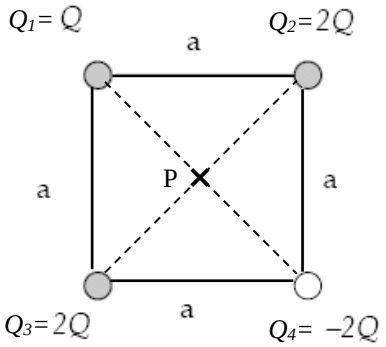


Fig. 16.

Moving between Two Points in an Electric Field

The **change in potential energy in moving a charge *q* from point A to point B** in an electric field is given by

$$\Delta U = U_B - U_A = q(V_B - V_A)$$

where V_A and V_B are the potentials of the field at A and B respectively.

Notes:

(i) The change in potential energy is the work done by the external agent in moving *q* from A to B.

(ii) The change in potential energy may be positive (meaning gain) or negative (meaning loss), depending on the sign of *q* and $(V_B - V_A)$.

(iii) The electrical field is a conservative field (like a gravitational field is). This means that

⇒ work done on a charged particle in moving it from one point to another depends only on the particle's initial and final positions and is **independent of the path taken.**

⇒ the change in potential energy is therefore independent of the path taken.

Example 5

An electron in the vacuum of a cathode ray tube moves from rest at a point where the potential is –1400 V to a point where the potential is zero.

Calculate

(a) the change in its potential energy,

(b) the change in its kinetic energy, and

(c) its final velocity.

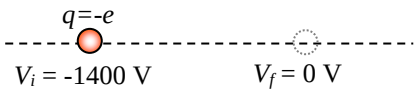


Fig. 17.

12.4 Relation between Electric Potential Energy and Electric Force

The relationship between the electric force *F* and the potential energy of the charge U can be written as

$$\boldsymbol{F} = -\frac{dU}{dr}$$

Note: This equation gives us both the magnitude and direction of the force acting on the charge placed at a point in an electric field.

- The *magnitude* of the force is given by:

  $$F = \left|\frac{dU}{dr}\right| = \left|\text{gradient of the } U \text{ vs. } r \text{ graph}\right|$$

- The direction of the force on the charge is given by the negative (-) sign i.e. **the force on the charge points in the direction of decrease in potential energy**.[6]

12.4.1 Relation between Potential Gradient and Electric Field Strength

From above, $$F = -\frac{dU}{dr}$$

$$\Rightarrow \quad qE = -\frac{d(qV)}{dr}$$

where *q* is the charge placed at the point in the electric field with field strength *E* and potential *V*. Dividing both sides by *q*, we get the relationship

$$\boldsymbol{E} = -\frac{dV}{dr}$$

Note:

- $\frac{dV}{dr}$ is known as the ***potential gradient*** of the field.
- The *magnitude* of $\boldsymbol{E}$ is given by:

  $$E = \left|\frac{dV}{dr}\right| = \left|\text{gradient of the } V \text{ vs. } r \text{ graph}\right|$$

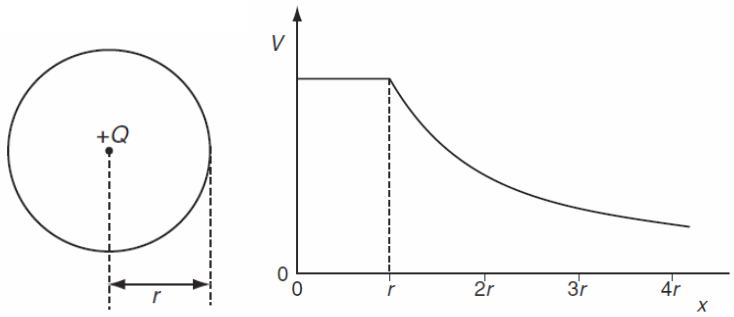
- The *direction of* $\boldsymbol{E}$ is **towards decreasing potential**.

[6] In fact, this is something fundamental in physics. To achieve greater stability charges would try to move to a location such that the system will end up with lower potential energy.

12.4.2 Graphs of Electric Field Strength and Electric Potential between Point Charges

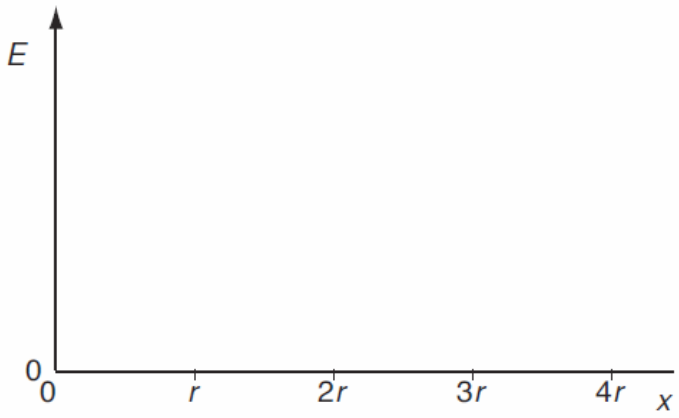
Example 6 J05/P4/Q5

An isolated conducting sphere of radius *r* is given a charge +*Q*. This charge may be assumed to act as a point charge situated at the centre of the sphere. The graph shows the variation with distance *x* from the centre of the sphere of the potential *V* due to the charge +*Q*.



(a) State the relation between electric field and potential. [1]

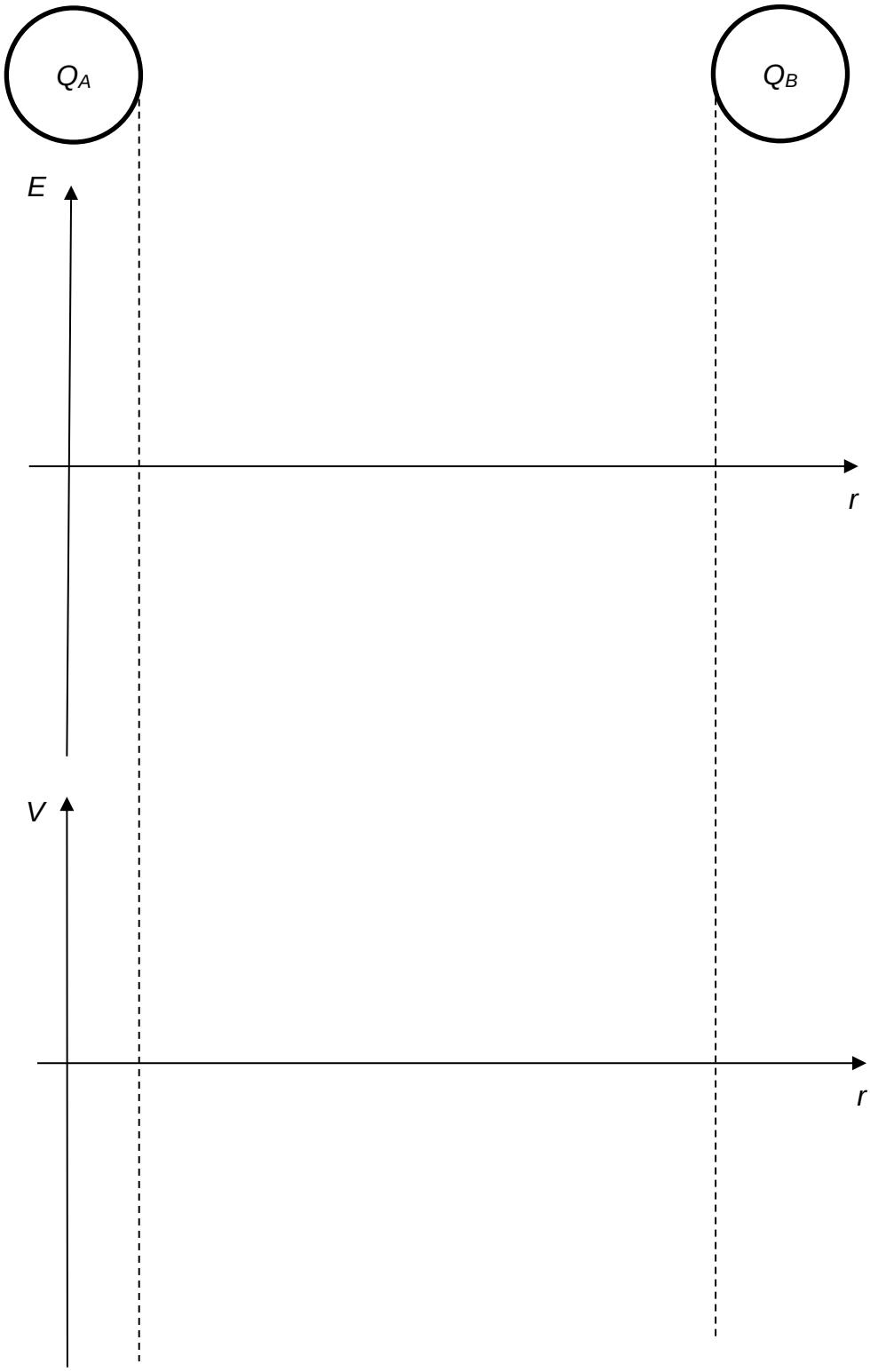
(b) Using the relation in **(a)**, sketch a graph to show the variation with distance *x* of the electric field *E* due to the charge +*Q*. [3]



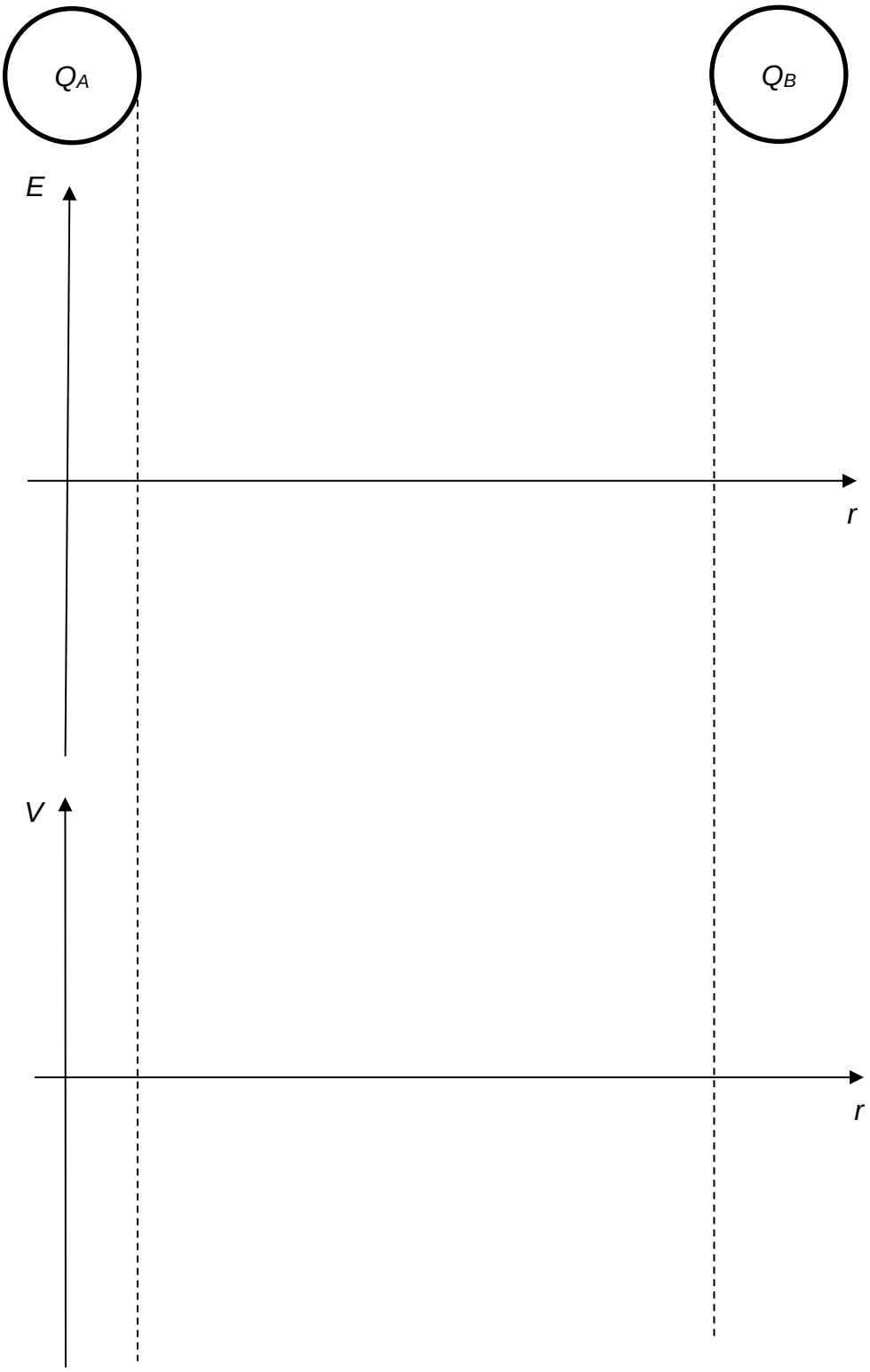
Example 7

Sphere A and sphere B are charged conducting spheres of radius R, with charge Q_A and Q_B respectively. They are sufficiently far apart that they obey Coulomb's Law. Let r be the distance from the centre of sphere A. Sketch the graphs of electric field strength against r and electric potential against r when:

(a) Both spheres are positively charged, and $|Q_B| = 2|Q_A|$.



(b) Sphere A is positively charged and sphere B is negatively charged, and $|Q_B| = 2|Q_A|$.



12.5 Representation of Electric Fields by Equipotential Lines or Surfaces

Another way to represent the electric field in a region is through *lines* or *surfaces* of equal potential.

An equipotential surface is a surface on which every point has the same potential.

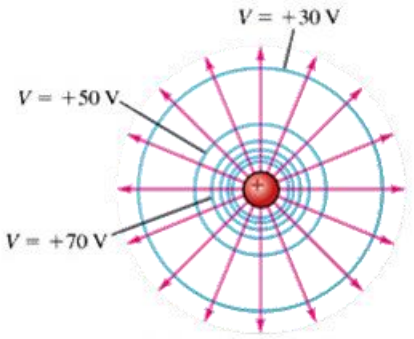
In general:

- Magnitude of the electric field $E = \left|\frac{dV}{dr}\right|$. Hence, **the closer the equipotential lines or surfaces** are to each other, the greater the change in potential for the same distance and hence **the stronger the electric field**.
- Direction of the **electric field points from higher potential to lower potential**.

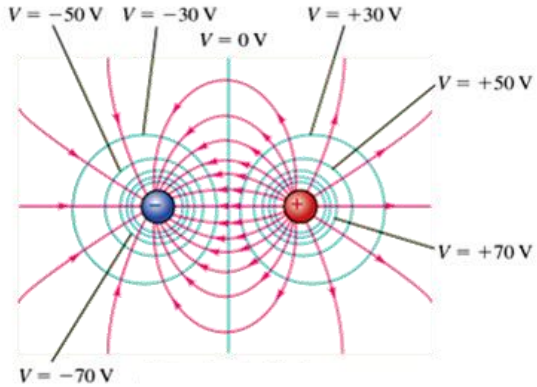
Further, as all points on an equipotential surface have the same potential,

- Potential gradient along an equipotential surface is zero.
- There is no component of electric field along an equipotential surface.
  - **Electric field lines are always perpendicular to equipotential surfaces.**

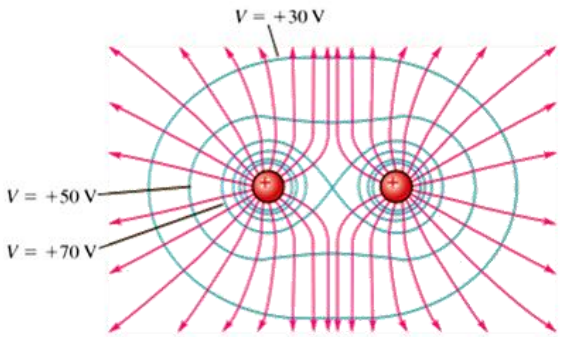
Fig. 18 below shows some examples of equipotential surfaces for different systems. We usually draw these equipotential lines or surfaces the *same number of volts* apart.



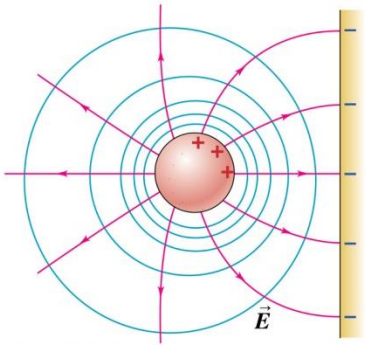
(a) A positive charge.



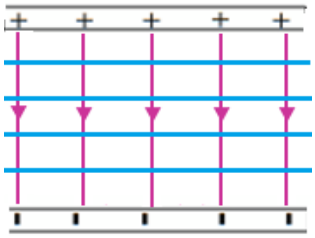
(b) Two opposite charges equal in magnitude.



(c) Two equal positive charges.



(d) A positively charged sphere and a negatively charged plate.



(e)Two parallel plates with opposite charge.

electric field lines equipotential lines (cross-section of equipotential surfaces)

Fig. 18. Equipotential and field lines.

12.5.1 Charges in Equilibrium on Conductors

An electrical conductor contains charges that are free to move. Hence, when a conductor is charged, the charges will always redistribute themselves until the excess charge resides on the surface (they try to move out of the conductor as they repel each other). When the charges are no longer moving about, i.e. they have settled down in equilibrium, they experience no net force. This implies that the electric field in the conductor is zero.

Since the charges at the surface experience no force parallel to the surface (else they will continue to move along the surface),

From $F_{//} = qE_{//} = 0$

$\Rightarrow$ **Electric field along the surface is zero,** $E_{//} = 0$.

From $\boldsymbol{E} = -\dfrac{dV}{dr}$

$\Rightarrow$ $\dfrac{dV}{dr} = 0$ along the surface of the conductor

$\Rightarrow$ **The surface of a conductor is an equipotential surface.**

$\Rightarrow$ **Electric field lines emerging or ending on a conductor are perpendicular to the surface.**

Furthermore, since the electric field within a conductor is zero, there is no potential gradient within the conductor all the way to the surface

$\Rightarrow$ **The potential in the conductor is constant everywhere inside the conductor and is equal to its value at the surface**.

$\Rightarrow$ **No field lines exist within the conductor.**

To be clear, the electric field in a neutral uncharged conductor placed in an external electric field is also zero. The free electrons in the neutral conductor redistribute themselves such that the (net) electric field within the conductor is zero! One can imagine the separated charges within the conductor producing an internal electric field that balances the external electric field. See Fig. 19.

Every point in and on the surface of the conductor is at the same potential; in other words, it is an equipotential region!

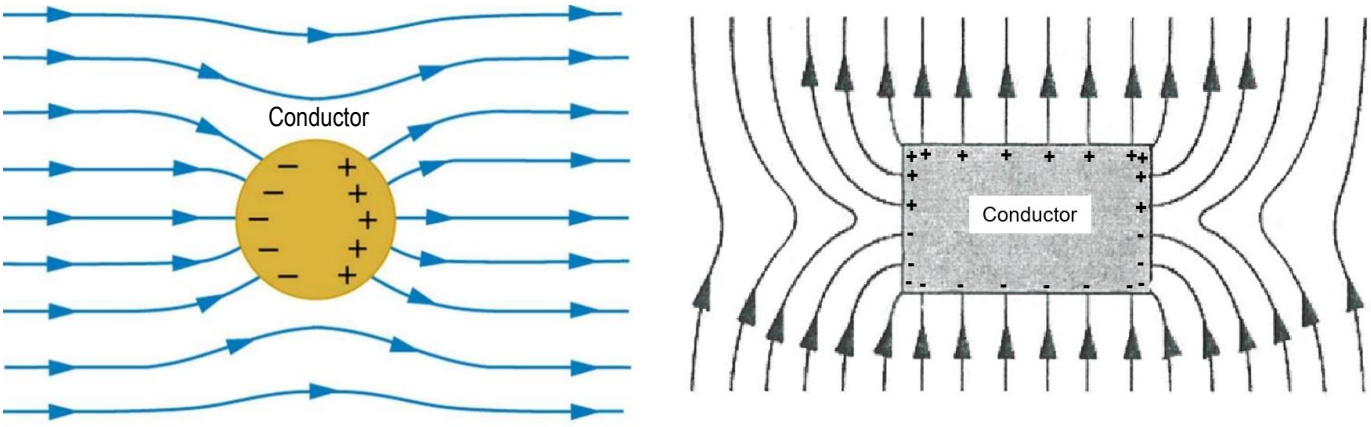


Fig. 19.

We summarise below a more complete set of rules for drawing electric field lines and equipotential lines for charged systems. We will include the previous rules for completeness.

Rules for Drawing Electric Field Lines and Equipotential lines:

i. Electric field lines are directed from higher potential (more positive) to lower potential (more negative).

ii. The number of lines drawn leaving a positive charge or approaching a negative charge is proportional to the magnitude of the charge.

iii. Field lines do not intersect, because ***E*** at a point can only point in one direction.

iv. Equipotential lines/surfaces are lines/surfaces that join all points of equal potential.

v. Equipotential lines/surfaces are drawn at equal intervals of potentials (e.g. 10 V, 20 V, 30 V etc.). Hence, the closer the equipotential lines or surfaces, the stronger the field.

vi. Electric field lines are always perpendicular to lines of equal potential, e.g. to the surface of a conductor.

vii. Electric field lines always point towards the direction in which potential is decreasing.

viii. The electric field in a conductor is zero. No field lines are drawn.

12.6 Electric Field between Two Charged Parallel Plates

Consider two large metal plates that carry equal but opposite charges. This can be done by connecting them to the opposite terminals of a constant voltage supply. It is reasonable to suppose that within each plate, except at the edges, the charges on each plate are spread uniformly over the inside surface of each plate, because of their mutual repulsion and the attraction by the opposite charges on the other plate.

The electric field lines are straight, parallel to each other and equally spaced (except at the fringes – see Fig. 8(f)). **The electric field in the space between the two plates is said to be uniform.**

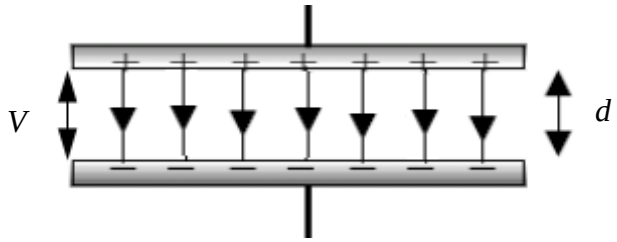


Fig. 20. Electric Field between two parallel, oppositely charged plates.

Suppose that in Fig. 20 the potential difference between two large charged metal plates is V and the plate separation is d. Since the electric field strength between the plates is constant, from $\boldsymbol{E} = -\frac{dV}{dr}$ we have

- **Magnitude of the electric field strength:**

$$\boxed{E = \frac{V}{d}}$$

- **Direction of the electric field:** $\boldsymbol{E}$ is directed from the plate with the higher potential to the plate with the lower potential.

12.6.1 Motion of Charged Particles in a Uniform Electric Field

Consider a particle of mass m, carrying charge q, that is placed in a uniform electric field E.
The field exerts a force qE on it, giving it an acceleration a in the direction of the force.

By Newton's second law, $ma = qE$. The acceleration of the charge is therefore given by $a = \frac{qE}{m}$.

In most questions, the mass is so small that the electric force overwhelms the gravitational force on the mass (weight). Hence the weight can be ignored.

For example, if you are considering an electron:
Its weight is $9.11 \times 10^{-31} \times 9.81 = 8.94 \times 10^{-30}$ N
If it is between parallel plates of potential difference 1.0 V and separation 10 cm, the electric force it experiences is $(1.60 \times 10^{-19}) \times 1.0/0.10 = 1.6 \times 10^{-18}$ N.

However, you should not assume that it is always the case that the weight can be ignored. You might get a question where the mass is large enough to make the weight significant compared to the electric force. (e.g. the Millikan oil drop experiment).

A positively charged particle will accelerate in the direction of E. A negatively charged particle will accelerate in the direction opposite to E.

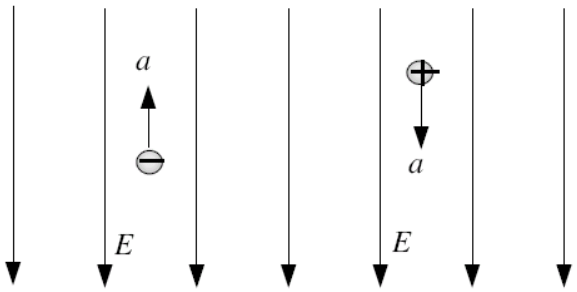


Fig. 21. Acceleration of charged particles subjected to an E-field.

In general, the motion is that of an object moving under constant acceleration. There are two possible paths:

i. If its initial velocity is zero or parallel to the direction of E, the charged particle will move in a straight line parallel to E.

ii. If its initial velocity is inclined to the field E, it will travel in a parabolic path.

Example 8

A beam of electrons, travelling at 1.35 x 10^7 m s^{-1}, enters a uniform electric field between two plates of length 0.060 m, which are separated by a distance of 0.020 m. The top plate is at +50 V while the lower plate is at −50 V. Find the angular deflection of the beam as electrons emerge from between the plates, assuming that the electrons never reach the upper plate.

Fig. 22

12.7 Analogy between Electric fields and Gravitational Fields

The laws governing interaction between point charges and the interaction between point masses are analogous. Mass, m, is analogous to charge, Q. The constant G is analogous to $1/4\pi\varepsilon_0$.

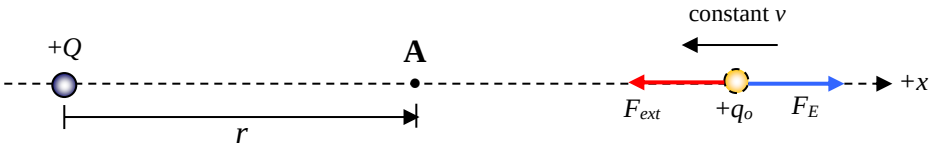
	Electric field	Gravitational field
Field Strength	$\vec{E} = \frac{\vec{F}}{q}$	$\vec{g} = \frac{\vec{F}}{m}$
	Point charge: $E = \frac{Q}{4\pi\varepsilon_0 r^2}$	**Point mass:** $g = \frac{GM}{r^2}$
Force	**Coulomb's Law:** $F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}$ The force between point charges follows an inverse square law, i.e. $F \propto \frac{1}{r^2}$	**Newton's Law of Gravitation:** $F = G\frac{m_1 m_2}{r^2}$ The force between point masses also follows an inverse square law, i.e. $F \propto \frac{1}{r^2}$
	Force is a *vector* quantity. The resultant force at a point must be determined by *vector sum* of the forces.	
Potential	**Due to a point charge:** $V = \frac{Q}{4\pi\varepsilon_0 r}$ Potential can be positive or negative depending on whether Q is positive or negative.	**Due to a point mass:** $\phi = -\frac{GM}{r}$ Potential is always negative.
	Potential is a *scalar* quantity. The resultant potential at a point is determined by the *algebraic sum* of the potential due to each charge or mass.	
Potential Energy	$U = Vq$ **Between point charges :** $U = \frac{Q_1 Q_2}{4\pi\varepsilon_0 r}$ U can be positive or negative. It is positive if Q and q have the same sign. Otherwise, U is negative.	$U = m\phi$ **Between point masses:** $U = -\frac{Gm_1 m_2}{r}$ U is always negative.
Field Strength and potential gradient	$E = -\frac{dV}{dr}$ The negative sign indicates the direction of the field is in the direction of decreasing potential.	$g = -\frac{d\phi}{dr}$ The negative sign indicates the direction of the field is in the direction of decreasing potential.
Force and potential energy gradient	$F = -\frac{dU}{dr}$	$F = -\frac{dU}{dr}$

Appendices

Appendix A: Deriving the Potential Energy and the Potential of an Electric Field by a Point Charge

[Adapted from: Duncan, 2000, p 204]

Consider a point charge $+Q$. To find the potential of the electric field at **A** a distance r away from $+Q$, we bring a very small test charge $+q_o$ from infinity to the point **A** (allowing for no gain in kinetic energy) and find the work done by the external force to carry out this action.



We note that when the test charge is moving in the field set up by the source, it experiences a varying electrical force F_E. An external force needs to move it from infinity to the point.

Now, the work done dW by the external force to move a small distance of dx towards the source charge is

$$dW = F_{ext}dx$$

Since the test charge is moved at a constant velocity, $F_{ext} = -F_E$. Therefore,

$$dW = -F_E dx$$

The electric force experienced by the test charge a distance x away from the source charge is $F_E = \frac{Qq_o}{4\pi\varepsilon_o x^2}$. Therefore, we have the work done by the external force,

$$W = -\int_{\infty}^{r} F_E dx = -\int_{\infty}^{r} \frac{Qq_o}{4\pi\varepsilon_o x^2} dx = -\frac{Qq_o}{4\pi\varepsilon_o}[-\frac{1}{r}] = \frac{Qq_o}{4\pi\varepsilon_o r}$$

By definition of the electrical potential, the potential at point A is therefore,

$$V = \frac{W}{q_o} = \frac{Q}{4\pi\varepsilon_o r}$$

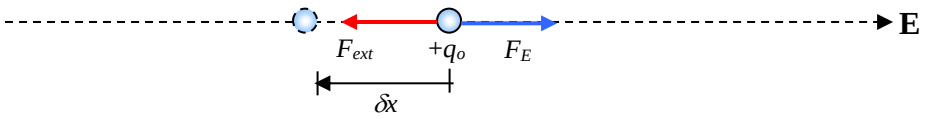
Appendix B: Deriving the Relationship between Force and Potential Energy

The relationship between a (conservative) force F and the associated potential energy is given by

$$F = -\frac{dU}{dx}$$

This is a general relationship that applies to all conservative forces, of which the gravitational force and the electric force are examples.

To get this relationship, it is best to look at a specific case. In this case we will look at the electrical force F_E experienced by a charge in an electrical field.



By conservation of (mechanical) energy, assuming that no dissipative forces act on the system,

Work done by an external force on the system W = Increase in P.E. + Increase in K.E. of system

Assuming we move the object in the field at a constant velocity

$\Rightarrow$ Increase in K.E. of system = 0

Now, let us move the object (a small charge in this case) by a small distance of δx in the direction of the external force, the small amount of work done δW by the external force is this case can be written as

$$\delta W = F_{ext}\, \delta x$$

Since the object is moving with constant velocity, $F_{ext} = -F_E$

Hence, we have

$$\delta W = F_{ext}\, \delta x$$

$$\Rightarrow \quad \delta W = -F_E\, \delta x$$

Now, as stated above, the work done by the external force results in the gain of potential energy of the system, hence we have

$$\Rightarrow \quad \delta U = \delta W = -F_E\, \delta x \qquad \text{where } U \text{ is the potential energy}$$

which can be written as $\Rightarrow \quad F_E = -\frac{dU}{dx}$ (derived)

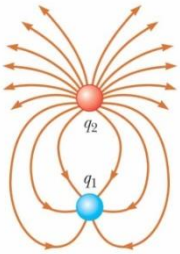
Tutorial 12 Electric Fields

Self-Review Questions

Electric Field and Electric Forces

S1 (Jewett & Serway, P23.32)

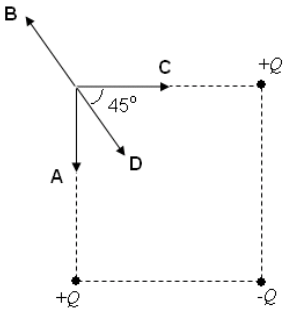
The figure on the right shows the electric field lines for two charged particles separated by a small distance.



(a) Determine the ratio $\frac{q_1}{q_2}$.

(b) What are the signs of q_1 and q_2?

S2 Point charges, each of magnitude Q, are placed at three corners of a square as shown in the diagram. What is the direction of the resultant electric field at the fourth corner? (Diagram on the right)



S3 (Jewett & Serway, P23.4)

A charged particle A exerts a force of 2.62 μN to the right on a charged particle B when the particles are 13.7 mm apart. Particle B moves straight away from A to make the distance between 17.7 mm. What vector force does it then exert on A?

S4 Two *small* spheres with their centres 20 cm apart *in vacuum* are charged oppositely with charges of $4{\cdot}0\text{x}10^{-8}$ C and $-9{\cdot}0\text{x}10^{-8}$ C respectively.

(a) Determine the position of the neutral point in the field along the line where the two charges lie.

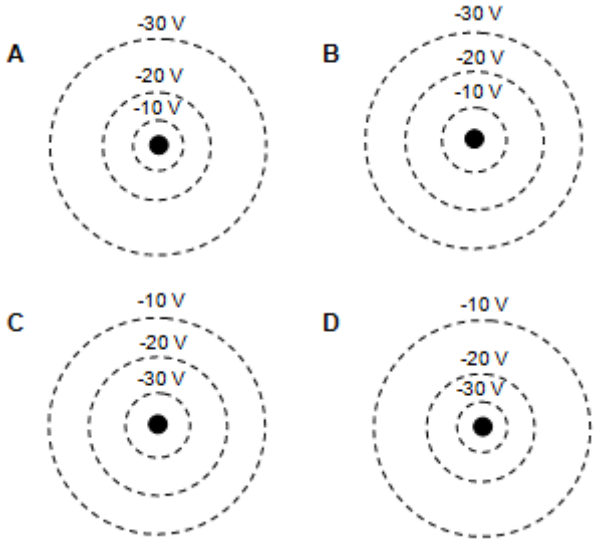
(b) A small positive charge is now placed at this neutral point. What is the force experienced by this charge?

(c) Will the force in (b) change if the charge in (b) is now changed to a negative charge? Explain your answer.

Electric Potential and Potential Energy

S5 HCI/2008/BT1/P1/Q4

Which of the following diagrams may represent the pattern of equipotential lines surrounding a negative point charge?



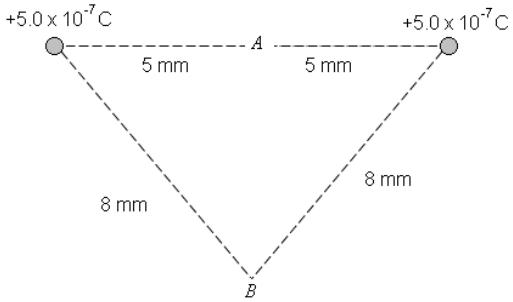
S6 The work done by an external agent to bring a positive charge of 2.0×10^{-7}C from a place of zero potential to a point P in an electric field is 6.0×10^{-4} J.

(a) Calculate the potential at P.

(b) If 4.0×10^{-4} J of work is done by the electric field when the charge at P moves to another point Q, what is the potential of Q?

(c) What is the potential difference between P and Q?

S7 Two ions, each $+5.0 \times 10^{-7}$ C, are 10.0 mm apart as shown. Point A is 5.0 mm away from the ions while point B is 8.0 mm away.



(a) Calculate the work done by an external agent in moving an electron from point A to point B.

(b) Comment if there is an increase or decrease in electric potential energy when the electron moves from A to B.

(c) How would your answer to (b) change, if at all, if one of the ions had been -5.0×10^{-7} C instead?

Relationship between Electric Force and Potential Energy

S8
Is it possible for the electric field strength to be zero where the electric potential is not zero? Give an example.

S9
Refer your answer to Lecture Example 7 in this chapter, and that of Lecture Example 11 in the Gravitational Field lecture notes that you studied last year.

Compare the gravitational field strength and gravitational potential of two spherical masses with the electric field strength and electric potential of:
(a) two spheres with like charges
(b) two spheres with unlike charges

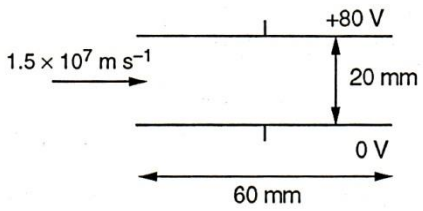
Uniform Electric Field

S10 A proton starts from rest and gains 8.35 x 10^{-14} J of kinetic energy as it accelerates between points A and B in a uniform electric field. Calculate the electric field strength between A and B if they are separated by a distance of 0.500 m.

A 5.22 x 10^{5} V m^{-1}
B 1.34 x 10^{5} V m^{-1}
C 5.99 x 10^{6} V m^{-1}
D 1.04 x 10^{6} V m^{-1}

S11 Cambridge, UCLES, N88/III/11(part)

The figure below shows part of the deflection system of a cathode-ray tube. An electron moving with a speed of 1.5 x 10^{7} m/s approaches the region between two parallel metal plates, which are 20 mm apart and 60 mm long. The upper plate is at a steady positive potential of 80 V with respect to the lower



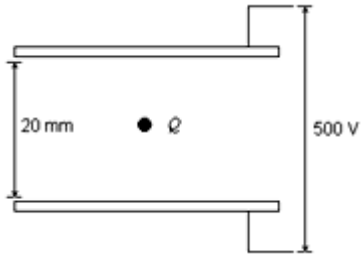
(a) Copy Fig. 11 on to your answer paper and sketch the path of the electron as it passes between and beyond the plates. [1]

(b) Find the magnitude and direction of the acceleration of the electron in the region between the plates. [3]

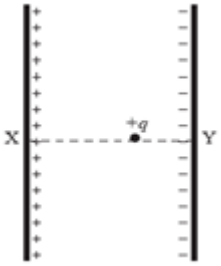
(c) Hence find the vertical and horizontal components of the velocity of the electron when it emerges from the plates. [3]

(d) Use your answer to *(c)* above to find the angle through which the electron beam has been deflected as a result of passing between the plates. [3]

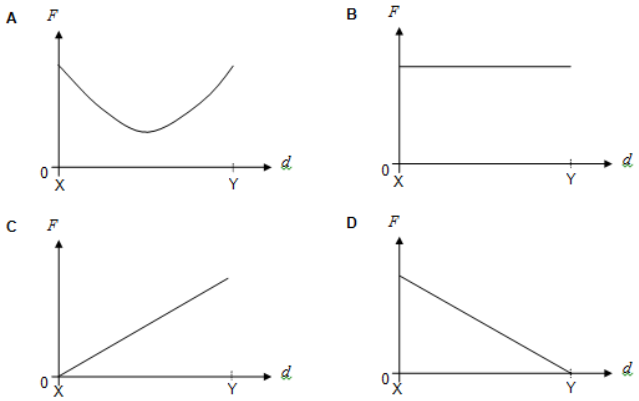
S12 An oil drop of mass 2.0 x 10^{-14} kg carries a charge Q. The drop is stationary between two parallel plates 20 mm apart with a potential difference of 500 V between them. Calculate Q.



S13 HCI/2009/BT2/P1/Q22
A positive test charge +*q* is placed along dotted line XY between two long parallel plates that are oppositely charged as shown.



The charge +*q* is moved from X to Y. Which one of the following graphs shows the variation of the magnitude *F* of the force on +*q* with distance *d* from X?

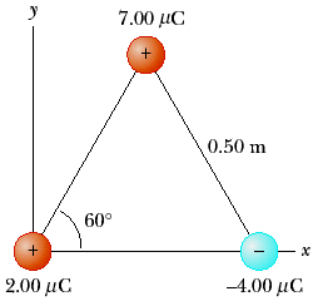


Discussion Questions

Electric Fields and Electric Forces

D1 (Jewett & Serway, P23.9)

Three point charges are located at the corners of an equilateral triangle as shown in the figure on the right. Calculate the resultant electric force on the 7.00 μC charge.



D2 (modified from Jewett & Serway, P23.57)

Two small metal spheres of mass m are suspended from strings of length ℓ that are connected at a common point. One sphere has charge Q; the other has charge $2Q$. The strings make angles θ_1 and θ_2 with the vertical. Assume θ_1 and θ_2 are small.

(a) How are θ_1 and θ_2 related?

(b) Show that the distance r between the spheres is given by

$$r \approx \left(\frac{4k_e Q^2 \ell}{mg} \right)^{1/3}$$

where $k_e = 1/(4\pi\varepsilon_o)$

D3 (Cambridge UCLES, N91/II/6)

(a) Two ions **A** and **B** are separated by a distance of 0.72 nm in a vacuum, as shown in Fig. 7. **A** has a charge of +3.2 x 10^{-19} C and **B** has a charge of -1.6 x 10^{-19} C. What force does **A** exert on **B**?

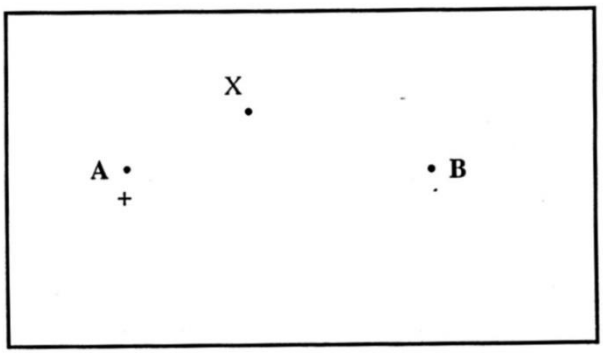


Fig. 7

(b) Without making detailed calculations, draw labelled arrows on your diagram to represent

(i) the field $\mathbf{E_A}$ at the point X due to the charge **A** only,

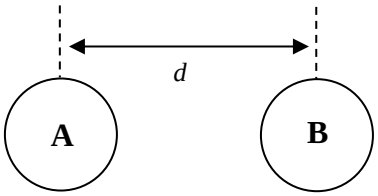
(ii) the field $\mathbf{E_B}$ at X due to charge at **B** only,

(iii) the resultant field **E** at X due to both charges.

(c) Sketch on your diagram above, lines representing the electric field caused by the two ions in the region within the rectangle. Include the field line passing through X.

D4 (Cambridge A-level Nov 1987, modified)

A and **B** are two identical conducting spheres, each carrying a charge of +Q. They are placed in vacuum with their centres distance d apart, as shown in the diagram below.



(i) Explain why the force F between them is **not** given by the expression $F = \dfrac{Q^2}{4\pi\varepsilon_o d^2}$

(ii) If **A** and **B** were uniformly charged spheres that were made of insulating materials instead, would the expression hold? Explain.

D5 (modified from N22/P2/Q3b)

The variations with separation of the gravitational force F_G and of the electric force F_E between two protons are shown in Fig 3.2.

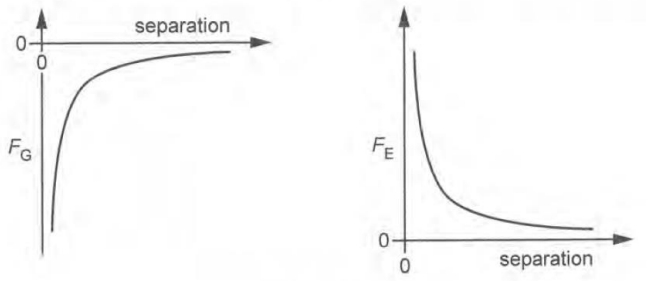


Fig. 3.2

(i) Explain why the gravitational force and the electric force have opposite signs.

(ii) Calculate the magnitude of the gravitational and electric forces on the protons when they are separated by a distance of 2 × 10^{-15} m. Thus, state and explain whether the axes on both graphs are drawn to the same scale.

(iii) Explain what your answers to (i) and (ii) imply for how the protons are able to remain this close together in the atomic nucleus.

Electric Potential and Potential Energy

D6 NYJC/2009/Prelim/P1/Q24
Two charges, - 9.0 × 10^{-6} C and + 2.5 × 10^{-3} C are placed 10.0 m apart. When their separation is changed to 5.0 m, the electrostatic potential energy of the system

A decreases by 20 J
B decreases by 6.0 J
C increases by 6.0 J
D increases by 20 J

D7 (Jewett & Serway, P25.19)

(i) Show that the amount of work required to assemble four identical charges of magnitude *Q* at the corners of a square of sides, *s* is $5.41\frac{Q^2}{4\pi\varepsilon_o s}$.

(ii) Is this positive work done by the system or on the system?

(iii) What is therefore the potential energy of the system?

(iv) Hence or otherwise, what is the work done by an external force in bringing one of the charges to infinity?

D8 (Jewett & Serway, P25.40)

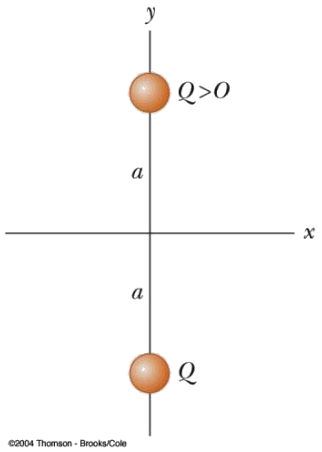
Electric charge can accumulate on an airplane in flight. You may have observed needle-shaped metal extensions on the wing tips and tail of an aeroplane. Their purpose is to allow charge to leak off before much of it accumulates. The electric field around the needle is much larger than around the body of the airplane and can become large enough to produce dielectric breakdown of the air, discharging the airplane.

To model this process, assume that two spherical conductors are connected by a long thin conducting wire and a charge of 1.20 μC is placed on the combination. One sphere, representing the body of the airplane, has a radius of 6.00 cm, and the other, representing the tip of the needle, has a radius of 2.00 cm. You may assume that no charges reside on the wire at equilibrium.

(i) What is the electric potential of each sphere?

(ii) What is the electric field at the surface of each sphere?

D9 Two positively charged particles of equal magnitude are located along the *y* axis equal distances above and below the *x* axis as shown in the figure.



(a) Are there any points in the region at which the electric field due to these two charges is zero? Are there any points in this region at which the potential due to these two charges is zero? Explain your answer.

(b) If the bottom charge were replaced by a negative charge of –*Q*, would there be any points in the region at which the electric field due to these two charges would be zero? Would there be any points in the region at which the potential due to these two charges would be zero? Explain.

Relationship between Electric Force and Potential Energy

D10 J07/P4/Q3
Two charged points A and B are separated by a distance of 6.0 cm, as shown in Fig. 3.1.

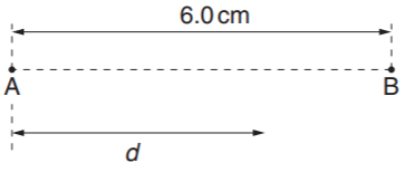


Fig. 3.1

The variation with distance *d* from A of the *magnitude of* electric field strength *E* along the line AB is shown in Fig. 3.2. (*Fig. 3.2 has been shrunk for printing here*)

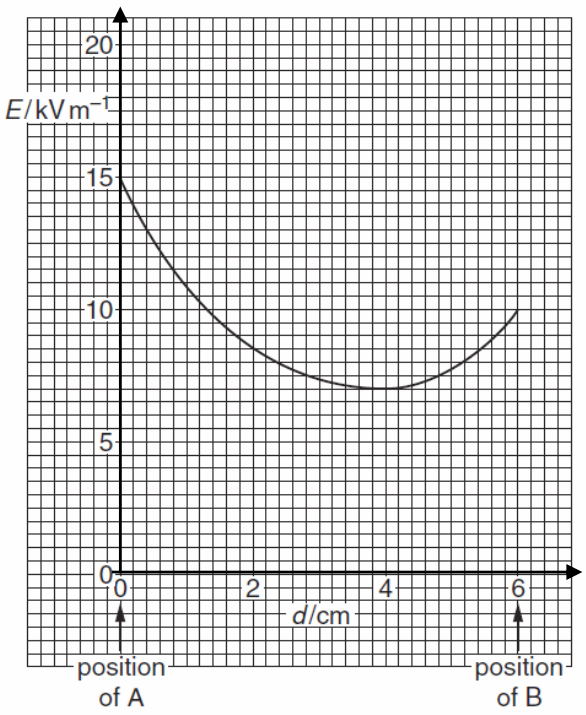


Fig. 3.2

An electron is emitted with negligible speed from A and travels along AB *as the result of the electric field*.

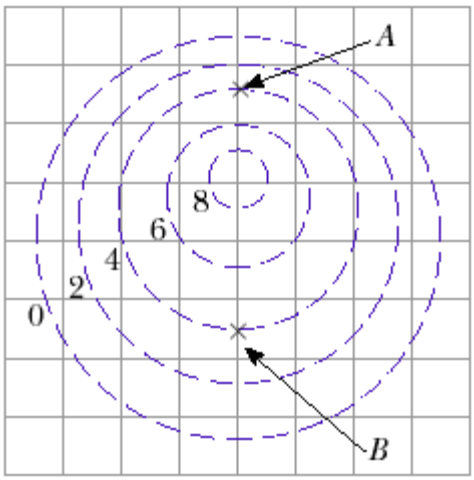
(a) State the relation between electric field strength *E* and potential *V*. [2]

(b) The area below the line of the graph of Fig. 3.2 represents the potential difference between A and B. Use Fig. 3.2 to determine the potential difference between A and B. [4]

(c) Use your answer to (b) to calculate the speed of the electron as it reaches point B. [2]

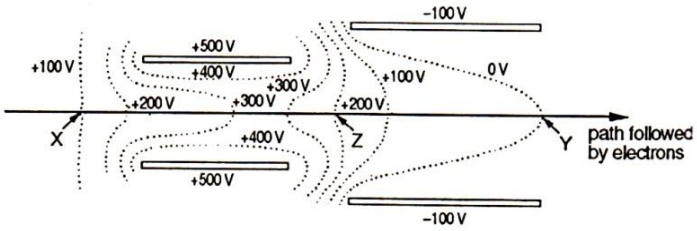
(d)(i) Use Fig. 3.2 to determine the value of *d* at which the electron has maximum acceleration. [1]

(ii) Without any further calculation, describe the variation with distance *d* of the acceleration of the electron. [2]



D11 (Cambridge, CIE, J93/III/4)

In an electron microscope, an electron lens has two cylinders which are at potentials of +500 V and -100 V respectively. An electron beam passes at high speed into the lens from the left. A cross-section of the two cylinders is shown full scale below, together with dotted lines showing points of same potential at different places within the cylinders.



Use the diagram where necessary measuring distances on it with a ruler, to

(i) find the potential difference between X and Y, [1]

(ii) find the loss in kinetic energy of an electron moving in a vacuum between X and Y. [2]

(iii) find the approximate value of the electric field strength at Z. [3]

(iv) sketch a graph showing how the potential varies along the axis of the lens. [2]

(v) sketch on the same *x*-axis drawn in (iv), a graph showing how the speed of the electron varies as it passes from X to Y.

D12 (Serway 's Physics for Scientists and Engineers Pg 790 Problem 40)

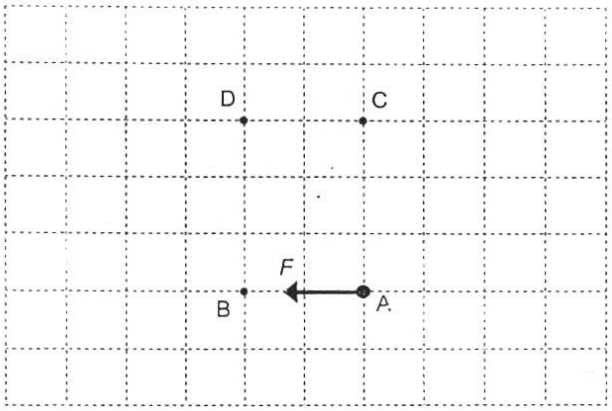
The figure shows several equipotential lines each labelled by its potential in volts. The distance between the lines of the square grid represents 1.00 cm.

(i) Is the magnitude of the field larger at ***A*** or at ***B***? Explain.
(ii) Determine the electric field *E* at ***B***.
(iii) Represent what the electric-field looks like by drawing at least eight field lines.

Electric Field of a Parallel Plate System

D13 (RJC Prelim 2001 Paper 2)

The figure below shows a small charged particle at a point A in a *uniform* electric field. The particle experiences an electrical force *F* (as shown) of 5.00 x 10^{-7} N. The grid lines in the figure are at intervals of 10.0 mm.



(a) Calculate the work done by the electrical force if the particle is moved from

(i) from A to B

(ii) from A to C

(iii) from A to D [3]

(b) If the particle carries a charge of -2.50 x 10^{-11} C, and the point A is at a potential of 200 V, what are the potentials of B, C and D due to the uniform field? [4]

(c) Draw an arrow in the figure above to indicate the direction of the uniform electric field. [1]

(d) If the uniform electric field is produced by a pair of flat metal plates, one of which is earthed and the other is at a potential of 1 000 V, draw on the figure above, the position of the two plates. [2]

D14 (Cambridge, UCLES, N99/III/6 (part))

(a) Fig. 6.1 illustrates two plates A and B, a distance 30 mm apart in a vacuum, with A at a potential of -4.2 V and B at zero potential. Electrons are emitted from B and move directly towards A.

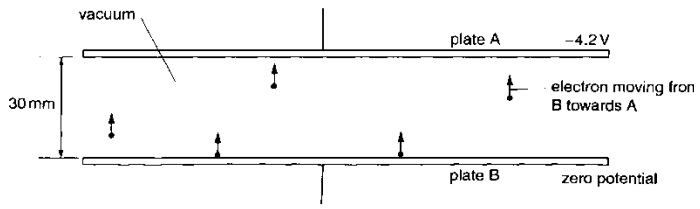


Fig. 6.1

Calculate

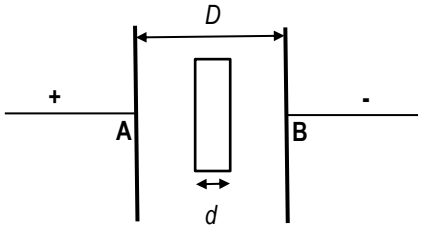
(i) the electric field, assumed uniform, between the plates,

(ii) the velocity with which the electrons need to be emitted in order to reach plate A.

(b) State, with a reason, what your answer to (a) (ii) would have been if the distance between the plates had been doubled to 60 mm, while keeping the potentials the same. [2]

D15 (Cambridge, J87/III/6)

Two large metal plates are oppositely charged and placed a distance *D* apart. A conductor of thickness *d* is situated centrally between the plates.



Sketch graphs to show the variation from A to B of

(a) the electric potential *V*,

(b) the electric field strength *E*.

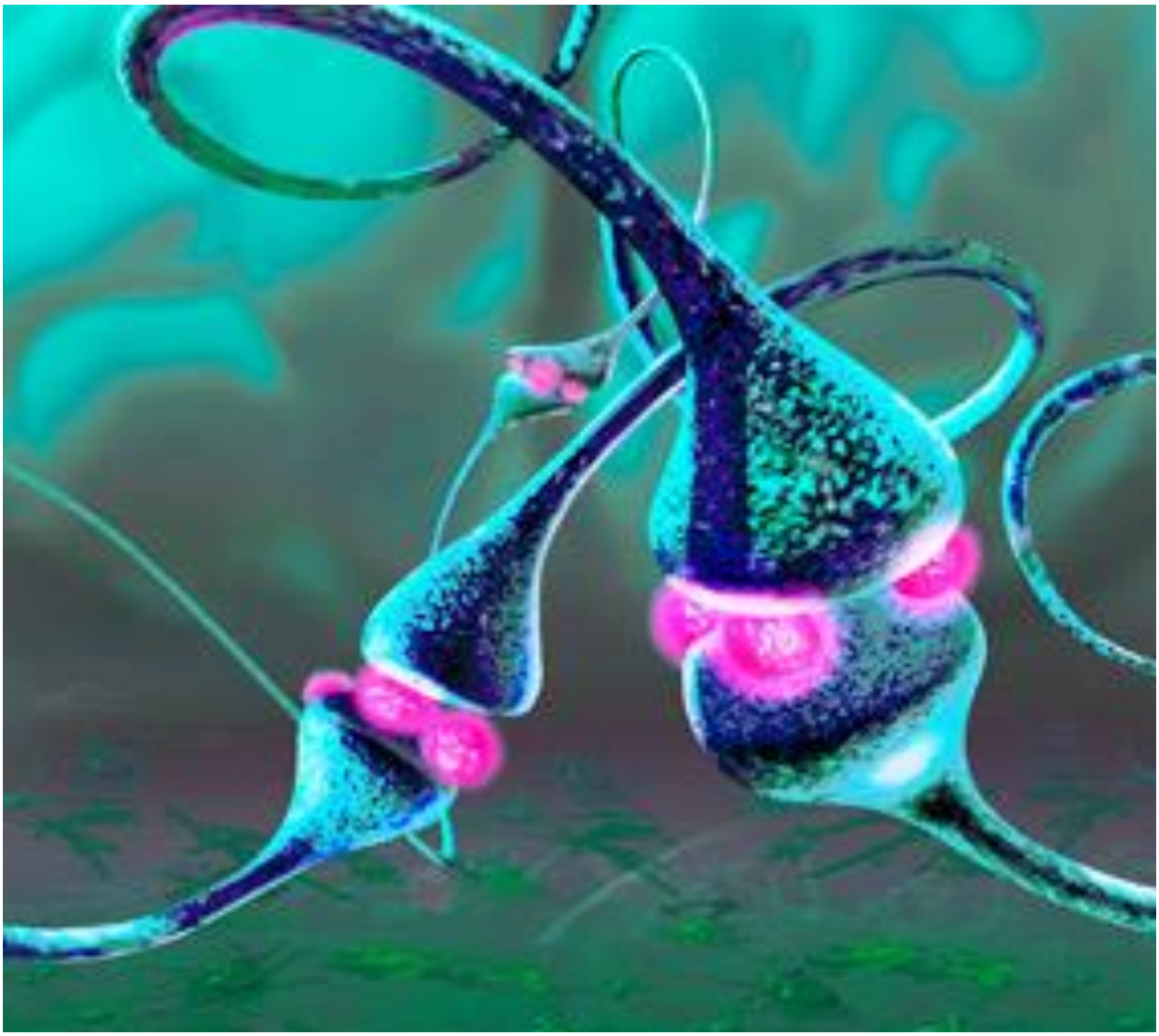
Numerical Answers to Tutorial

S1 (a) 0.33
S2 B
S3 1.57 μN to the left.
S4 (a) 40 cm from the positive charge, (b) 0 N, (c) 0 N
S5 D
S6 (a) 3000V, (b) 1000V, (c) 2000V
S7 1.08 x 10^{-13} J
S10 1.04 x 10^{6} V m^{-1}
S11 (a) 7.03 x 10^{14} m s^{-2}, (b) 2.8 x 10^{6} m s^{-1},
(c) 1.5 x 10^{7} m s^{-1}, (d) 10.6° from the horizontal
S12 7.85 x 10^{-18} C
S13 B

D1 0.872N at 30° below the horizontal
D2 (a) $\theta_1 = \theta_2$
D3 (a) 8.88x 10^{-10} N
D5 (ii) F_G = 4.65 × 10^{-35} N, F_E = 57.5 N
D7 (iii) $5.41\frac{Q^2}{4\pi\varepsilon_o s}$, (iv) $-2.71\frac{Q^2}{4\pi\varepsilon_o s}$
D8 (i) 1.35 × 10^{5} V, (ii) 2.25 × 10^{6} V m^{-1}, 6.74 × 10^{6} V m^{-1}
D10 (b) 540 V, (c) 1.38 x 10^{7} m s^{-1}, (d)(i) d = 0
D11 (i) 100 V; (ii) 1.6 x 10^{-17} J; (iii) 15400 V m^{-1}
D12 (ii) 2 V cm^{-1} vertically downwards
D13 (a)(i) 1.00 x 10^{-8} J, (ii) 0 J, (iii) 1.00 x 10^{-8} J,
(b) V_B = 600 V, V_C = 200 V, V_D = 600 V
(c) towards the right (d) 1000 V plate is 20.0 mm left of BD, 0 V plate is 10.0 mm right of CA
D14 140 V m^{-1}, 1.21 x 10^{6} m s^{-1}, no change

Chapter 13

CURRENT OF ELECTRICITY & D.C. CIRCUITS



The electricity produced by our bodies is what allows synapses, signals and even heartbeats to occur. (Source: health.howstuffworks.com)

H2 Physics (Syllabus 9749)

Current of Electricity

Content

- Electric Current
- Potential Difference
- Resistance and resistivity
- Electromotive force

Learning Outcomes

Candidates should be able to:

(a) show an understanding that electric current is the rate of flow of charge

(b) derive and use the equation $I = nAvq$ for a current-carrying conductor, where n is the number density of charge carriers and v is the drift velocity

(c) recall and solve problems using the equation $Q = It$

(d) recall and solve problems using the equation $V = W/Q$

(e) recall and solve problems using the equations $P = VI$, $P = I^2R$ and $P = V^2/R$

(f) define the resistance of a circuit component as the ratio of the potential difference across the component to the current passing through it and solve problems using the equation $V = IR$.

(g) sketch and explain the $I - V$ characteristics of various electrical components such as an ohmic resistor, a semiconductor diode, a filament lamp and a negative temperature coefficient (NTC) thermistor.

(h) sketch the resistance-temperature characteristic of an NTC thermistor.

(i) recall and solve problems using the equation $R = \frac{\rho l}{A}$.

(j) distinguish between electromotive force (e.m.f.) and potential difference (p.d.) using energy considerations.

(k) Show an understanding of the effects of the internal resistance of a source of e.m.f. on the terminal potential difference and output power.

Playlist for Lecture Examples can be found here!

https://for.edu.sg/ozchoo

D.C. Circuits

Content:

- Circuit symbols and diagrams
- Series and parallel arrangements
- Potential Divider
- Balanced potentials

Learning Outcomes:

Candidates should be able to:

(a) recall and use appropriate circuit symbols as set out in the ASE publication *Signs, Symbols and Systematics (The ASE Companion to 16-19 Science, 2000)*

(b) draw and interpret circuit diagrams containing sources, switches, resistors, ammeters, voltmeters and/or any other type of component referred to in the syllabus.

(c) solve problems using the formula for the combined resistance or two or more resistors in series.

(d) solve problems using the formula for the combined resistance or two or more resistors in parallel.

(e) solve problems involving series and parallel circuits for one source of e.m.f.

(f) show an understanding of the use of a potential divider circuit as a source of variable p.d..

(g) explain the use of thermistors and light-dependent resistors in potential divider circuits to provide a potential difference which is dependent on temperature and illumination respectively.

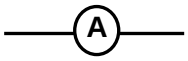
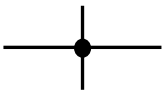
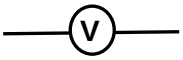
(h) recall and solve problems using the principle of the potentiometer as a means of comparing potential differences.

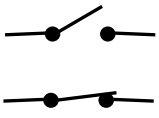
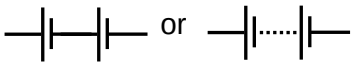
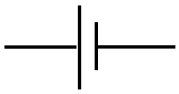
Table of Contents

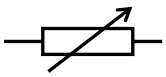
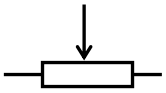
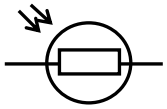
13.1 Practical Circuits 5
13.2 Electric Current 6
13.2.1 Units of Current & Charge 6
13.2.2 Direction of Electric Current 7
13.2.3 Drift Velocity of Charge Carriers 7
13.3 Potential Difference and Electromotive Force 9
13.3.1 Energy considerations of an electric current 9
13.3.2 Electromotive Force (e.m.f.) 9
13.3.3 Potential Difference 9
13.3.4 Distinction between e.m.f. and Potential Difference 10
13.4 Electrical Power 11
13.5 Resistance and Resistivity 12
13.5.1 Resistance 12
13.5.2 Ohm's Law (Review of High School Physics) 12
13.5.3 Resistivity 13
13.5.4 Determination of resistance using an ammeter and a voltmeter 14
13.5.5 Effective Resistance for Combination of Resistors 14
13.5.6 Power dissipation in resistors (Joule heating) 17
13.6 I-V Characteristics of Various Electrical Components 18
13.7 Internal resistance of a source of e.m.f. 21
13.7.1 Terminal potential difference, e.m.f and internal resistance 21
13.8 The Potential Divider Principle 23
13.8.1 Use of thermistors and light dependent resistors in potential dividers 24
13.9 The Potentiometer 26
13.9.1 Applications of the potentiometer 27
APPENDIX A – Two different arrangements of voltmeter and ammeter to find a more accurate value of the resistance of an unknown resistor 31
APPENDIX B – The Wheatstone bridge (Good to know) 32
Tutorial 13A: Current of Electricity 33
Tutorial 13B: DC Circuits 41

13.1 Practical Circuits

Symbols for various circuit elements

Conductor		Ammeter	
Junction of conductors		Voltmeter	
Earth		Galvanometer	

Switch		Battery (combination of cells)	
Cell		Lamp	

Fixed resistor		Diode	
Variable resistor		Thermistor	
Potential divider		Light dependent resistor	

Circuit Essentials

Most complete electric circuits contain 6 parts:

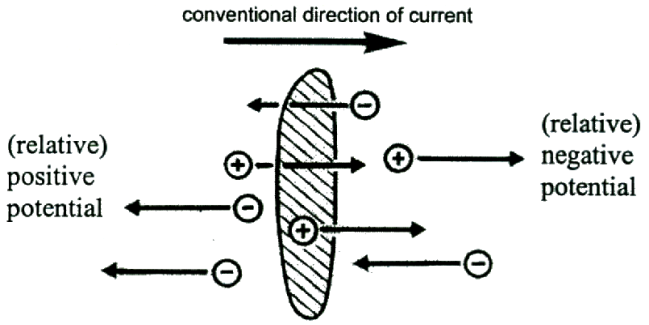
1. An *energy source* to provide the voltage needed to drive current through the circuit.
2. *Conductors* through which the current can travel.
3. *Insulators* to confine the current to the desired paths.
4. A *load* to control the amount of current and convert the electrical energy supplied by the energy source.
5. A *control device*, often a switch, to start and stop the flow of current.
6. A *protection device* to interrupt the circuit in case of a circuit malfunction

The first 4 of the above 6 parts are essential parts. All complete circuits use them. The control device is occasionally omitted. Protection devices are often omitted from circuits.

13.2 Electric Current

Electric current is the rate of flow of electric charge. It can be due to the flow of positive charges or negative charges or both.

Consider some charges moving through a cross-sectional area, which is perpendicular to the flow of charges. Suppose the electric charge δQ flows through the area in time δt.



The electric current I through the area is the **rate of flow** of the electric charges through the cross-sectional area of the surface. $I = \dfrac{\delta Q}{\delta t}$

If the charge flows at an irregular or non-linear rate, then the current I_i due to this flow of charge is

$$\boxed{I_i = \frac{dQ}{dt}}$$ (*Formula for instantaneous current*)

If the charges flow such that in a time interval t, the total charge which crosses that area is Q, the average current I_{av} due to this flow of charge is

$$\boxed{I_{av} = \frac{Q}{t}}$$ (*Formula for average or constant current*)

13.2.1 Units of Current & Charge

Current

SI unit for electric current is **ampere (A),** named after the French scientist who established the laws of electromagnetism. It was previously defined in terms of the magnetic forces that act between current-carrying conductors.[1] This magnetic force is explained in the topic Electromagnetism.

Charge

SI unit for charge is **coulomb (C)**, after the French scientist who first investigated the forces acting between electric charges. The unit is defined in terms of SI base units, ampere (A) and second (s).

One **coulomb (C)** is the quantity of electric charge that pass a given point in a circuit when a steady current of one ampere flows through that point for one second.

1 coulomb = 1 ampere-second

1 C = 1 A s

[1] From May 20, 2019, the ampere has been redefined to be based on the elementary charge (*e*). See this link for more info: **https://www.nist.gov/si-redefinition/ampere-introduction**

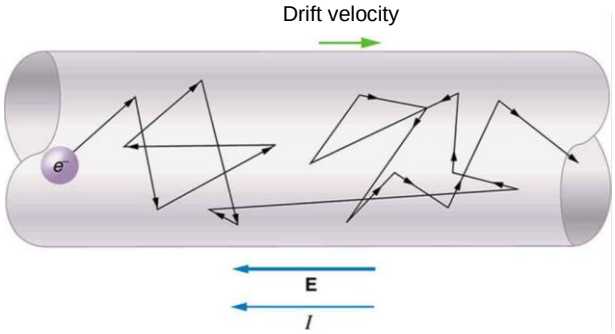
13.2.2 Direction of Electric Current

By convention, the direction of electric current (also known as conventional current) is chosen to be in the **same direction as the flow of positive charges**.

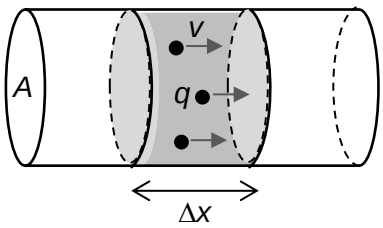
If the conductor is a metal (e.g. copper), the current is due to the motion of electrons. By convention, the current due to the flow of negative charges (i.e. mobile electrons) to the left is treated as the conventional current towards the right. ***Conventional current is in the opposite direction to the flow of electrons***[2]***.***

13.2.3 Drift Velocity of Charge Carriers

Since the current is given by the rate of flow of charges, let us now investigate the motion of the charges. When a battery is connected across a metallic conductor, the mobile charges, in this case mobile electrons, are accelerated due to electric forces acting on them. The electrons gain velocity and energy. When they collide with the atoms vibrating about its fixed mean position, they give up some of their energy to it. After slowing down or stopping on collision, the electrons are again accelerated by the electric force. Although the motion of the electrons are erratic, on the average, they drift in one direction with a mean or average velocity called the **drift velocity**. This one way drift of the electrons is an electric current.



Consider the current in a **conductor of cross-sectional area** *A*.



Suppose there are *n* **mobile charge carriers per unit volume**. If the charge carriers move with **drift velocity** *v*, the **distance moved in time** Δt is given by $\Delta x = v\Delta t$.
The charge carriers that flow out of the right shaded face of the shaded cylinder are the carriers that are within the volume of shaded section during the time Δt.
The total number of mobile charge carriers in that section would be given by $nA\Delta x$ and if **each carrier is of charge** *q*, then **the total charge** ΔQ in this section is given by

$$\Delta Q = \text{(number of charge carriers) x (charge per carrier)} = (nA\Delta x)q = (nAv\Delta t)q$$

By dividing both sides with Δt, we see that the current in a conductor is given by

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

[2] There is an interesting story on why the direction of current is always interpreted based on the flow of positive charges and not that of negative charges. Do find out more by investigating the work of Benjamin Franklin (1706 - 1790) who was the first person to study the electric current.

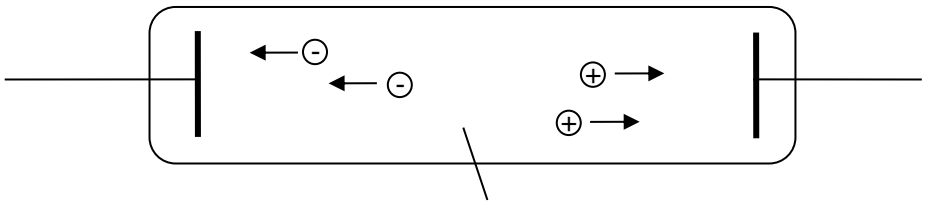
For a given current I, a good conducting material is one for which *n* is large and *v* therefore small. Copper is one of the best conductors, having in the region of 10^{29} free electrons per m^3.

Using the equation, we can estimate a typical value for *v* when a current of 1 A flows through a copper wire of cross sectional area $2.0\times10^{-7}\,m^2$ (about 0.5 mm diameter).

$$v = \frac{I}{nAe}$$
$$= \frac{1}{10^{29}\times2\times10^{-7}\times1.6\times10^{-19}}$$
$$= 3.1\times10^{-4}\ \text{m s}^{-1}$$

The actual values of the velocity *v* are amazingly small! However, it is worth noting that this slow drift starts almost **instantaneously in all parts** of an electrical circuit completed when, for instance, a switch is closed. (Hence there is clearly no delay between closing a switch in one part of a room and the light in another part coming on.)

Example 1



An external electric field is applied to the gas in the tube, resulting in negative particles flowing in one way and positive particles flowing in the opposite direction as illustrated above. In 1 s, 1.56 x 10^{15} negative particles pass through a perpendicular cross-section in the tube and the same amount of positive particles pass through it in the opposite direction. If each positive particle carries a charge +2*e*, and each negative particle carries a charge of –*e*, find

(a) the current due to the flow of positive charges,

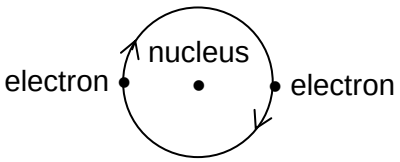
(b) the current due to the flow of negative charges,

(c) the net current through the gas.

Example 2 (J99/I/5)

The diagram shows a model of an atom in which two electrons move round a nucleus in a circular orbit. The electrons complete one full orbit in 1.0 x 10^{-15} s.

What is the current caused by the motion of the electrons in the orbit?



13.3 Potential Difference and Electromotive Force

13.3.1 Energy considerations of an electric current

The circulation of an electric current in a complete circuit leads to the continual dissipation of energy throughout the circuit in the form of heat and other forms of energy (e.g. light energy in a filament lamp, mechanical energy in an electric motor, etc.). All this energy must be continuously supplied by an e.m.f. source e.g., battery, dynamo, etc.

13.3.2 Electromotive Force (e.m.f.)

The sources of e.m.f. (stands for electromotive force) required in any complete circuit to generate electric current are therefore primarily means of <u>*converting other forms of energy to electrical form*</u>. E.g. a dry cell converts chemical energy to electrical energy, a dynamo converts kinetic energy to electrical energy, a microphone converts vibrational energy of a sound wave to electrical energy, etc.

The term e.m.f. has historical origins, but it is ***NOT a force***, despite its name. It is more a measure of energy (i.e., electrical potential energy) rather than force.

The **electromotive force (e.m.f.), *E,*** of a source is defined as the ***energy converted per unit electric charge to electrical energy in driving charge round a complete circuit*.**

The SI unit for e.m.f. is volt (V), after the Italian scientist Alessandro Volta, who first investigated the action of electric batteries.

If charge *Q* flows in a circuit containing a source of e.m.f. *E,* the electrical energy supplied by the source to the whole circuit is

$$W = QE$$

The power *P* of the source varies with the current it drives and hence, we normally specify the power of the source with reference to the current, i.e. its power to current ratio. This ratio which is a constant of the electrical source is the e.m.f. of the source.

$$P = \frac{dW}{dt} = \frac{dQ}{dt}E = IE$$

13.3.3 Potential Difference

A source of e.m.f. produces an *electric field* between its terminals. The electric force will act on charge carriers in the circuit and a complete conducting path between the terminals allows the charge carriers to flow through the circuit.

Inside the conductors, the electric force does work on the moving charged particles. As a result, electrical potential energy is converted into heat and other forms of energy.

We can define the electric potential at a point in the circuit. The **conventional current will flow from a point of higher electric potential to a point of lower electric potential in the circuit.** The potential difference (p.d.), between two points in the circuit is the work done by the electric field to drive a unit charge from one point to the other.

The p.d. across a conductor is sometimes described as “voltage” or “potential drop” across the conductor.

Most of the time, **it is only differences of potential (i.e., p.d.) and NOT the potentials of points in a circuit that are of any significance in electric circuits**.

To determine the electric potential at a point in a circuit, it would be necessary to first identify a point to be at *conventional zero potential*. This point is used as a reference to which the potentials of other points in the circuit can be determined. Many electrical circuits have a point in the circuit that is connected to Earth. By convention, *Earth is taken to be at zero potential*. Electrical apparatus that is connected to the earth is said to be “earthed".

Circuit symbol to represent earthing of a point of a circuit (i.e. the point is at 0 V potential).

The ***potential difference V between two points*** in a circuit is defined as the **energy converted *per unit electric charge* from electrical energy to other forms of energy when charge moved between the two points.**

If the electrical energy W converted is for charge Q moved between the two points, then, **potential difference** V between two points is

$$V = \frac{W}{Q}$$

Thus, if a charge Q flows in a part of a circuit across which there is a p.d. V, *the electrical energy supplied* is given by $W = QV$.

The SI unit for potential difference is volt (V).

13.3.4 Distinction between e.m.f. and Potential Difference

Energy conversion

e.m.f. is a measure of the ***energy converted per unit electric charge to electrical energy in driving charge round a complete circuit*.**

Potential difference is a measure of **energy converted *per unit electric charge* from electrical energy to other forms of energy when charge moved between the two points.**

There will be no potential difference across any part of the circuit without an e.m.f. in the circuit.

An e.m.f. exists, regardless of whether current flows or not; e.g. a left-alone cell has e.m.f.

13.4 Electrical Power

If a charge Q flows in a part of a circuit across which there is a p.d. V, the electrical energy supplied by the source to that part of the circuit is given by

$$W = QV \quad \text{..... (1)}$$

The flow of current is accompanied by the conversion of electrical energy into other forms of energy. *The rate at which energy is converted from one form to another in a device* is called its *power*.

Hence, a device of power P operating for time t converts a quantity of energy W, given by

$$W = Pt \quad \text{..... (2)}$$

From (1) & (2)

$$Pt = QV$$

$$P = \frac{Q}{t}V$$

$$\boxed{P = IV}$$

The unit of power is the watt (W), after the Scottish engineer James Watt who developed the steam engine. Units of power in regular use are the *kilowatt (kW)* and the *megawatt (MW)*.

The equation above gives rise to an alternative definition of potential difference across any electrical device or in electrical circuits.

The *potential difference between two points* in a circuit or across any electrical device can also be defined as the electrical *power delivered per unit current* that flows through it.

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

Alternative units of electrical energy

Recall: The SI unit of energy is the joule (J), named after the British scientist James Prescott Joule who established the relationship between heat and mechanical forms of energy.

1 J = 1 N m (newton meter)

The kilowatt-hour (kW h)

The calculation of quantities of electrical energy for commercial purposes is conveniently carried out by means of this unit.

Since 1 kW = 10^3 J s^{-1} and 1 hour = 3.6 x 10^3 s,
Therefore, 1 kW h = 3.6 x 10^6 J

The electron-volt (eV)

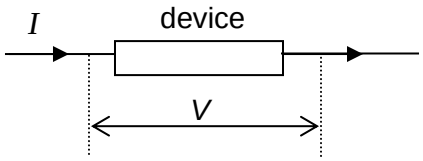
This is a convenient unit for the calculation of the energy of particles in atomic and nuclear physics (frequently used in the topics of Electric Field and Nuclear Physics). It is the quantity of energy gained by an electron in accelerating through a p.d. of 1 V.

Since charge of one electron is -1.60 x 10^{-19} C,
1 eV = 1.60 x 10^{-19} J

13.5 Resistance and Resistivity

If a potential difference is applied across a conductor, then a current I will flow through the conductor. The potential difference required to produce a given current depends on a property of the conductor called its resistance. Resistivity is a property of the material of the conductor. Example: *resistance* of copper wire, *resistivity* of copper.

13.5.1 Resistance



The resistance, R, of a device is the **ratio** of the potential difference V across the device to the current I flowing through it.

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

The electrical resistance of a conductor is due to its atoms obstructing the drift of electrons past them. Resistance is thus a measure of the obstruction to the movement of charge carriers.

Unit for resistance R is **ohm (Ω)**.

13.5.2 Ohm's Law (Review of High School Physics)

Georg Simon Ohm, a German physicist, was the first person to distinguish clearly between the two concepts, p.d. and current, and to investigate the relation between them. Ohm determined that there is a *direct proportionality* between the potential difference applied across a metallic conductor and the resultant electric current through the conductor (known as Ohm's Law). From the results of his experiment, Ohm was able to define the fundamental relationship between voltage, current and resistance.

Thus, **Ohm's Law states that**:

A *steady current* flowing through *a metallic conductor* is *proportional to the **potential difference** across it*, provided the *temperature and other physical conditions are constant.*

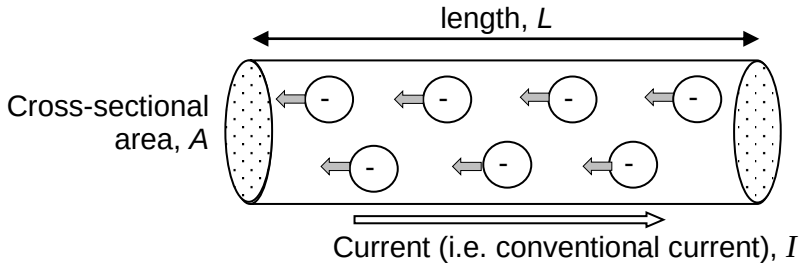
As a consequence of Ohm's Law, the resistance for a given metallic conductor under steady physical conditions is a constant. Such conductors are also known as **ohmic conductors.**

The current in a metallic conductor depends on a number of factors beside the p.d. across it. The most important of these is the temperature. But other physical conditions will also affect the current through it.

In addition, for electrolytic conduction, the relation between p.d. and current is complicated by the e.m.f that arises at the electrodes, accompanying the conversion of electrical energy to chemical energy. The same situation arises in other cases where the conductor is itself the site of an e.m.f. (e.g. the armature coil of a motor, or a coil in which a rapidly changing current induces an e.m.f. by self-induction).

13.5.3 Resistivity

The resistance of a uniform conductor is
(i) proportional to its length *L*
(ii) inversely proportional to its cross-sectional area *A*
(iii) dependent on a property of the material of the specimen known as its resistivity, ρ.



The relationship between resistance and the above factors is given by

$$R = \rho \frac{L}{A}$$

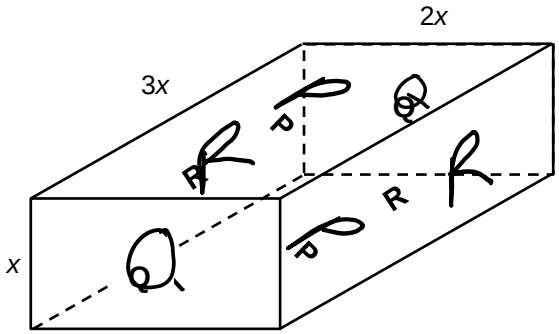
Unit for resistivity is **Ω m**.

The table shows the resistivity of some common materials.

Material	Resistivity at 20°C / Ω m
silver	1.6×10^{-8}
copper	1.7×10^{-8}
gold	2.2×10^{-8}
iron	10.1×10^{-8}
nichrome	110×10^{-8}
skin (dry)	3×10^{4}

Example 3 (J96/I/14)

The diagram shows a rectangular block with dimensions x, 2x and 3x. Electrical contact can be made to the block between opposite pair of faces (P, Q and R). Between which two faces would the maximum electrical resistance be obtained?



13.5.4 Determination of resistance using an ammeter and a voltmeter

In principle, the resistance R of a resistor can be found by measuring the current I which flows through it as a result of there being a potential difference, V, across it, and substituting the measured values of I and V into the equation $R = V/I$.

Whenever **an ammeter** is put into a circuit, it is put **in series** with the circuit components, and therefore it must have **zero resistance** if it is not to affect the current in the circuit.

A voltmeter, on the other hand, is always placed **in parallel** with a circuit component, and therefore must have **infinite resistance** if it is not to draw current through itself and alter the current in the circuit.

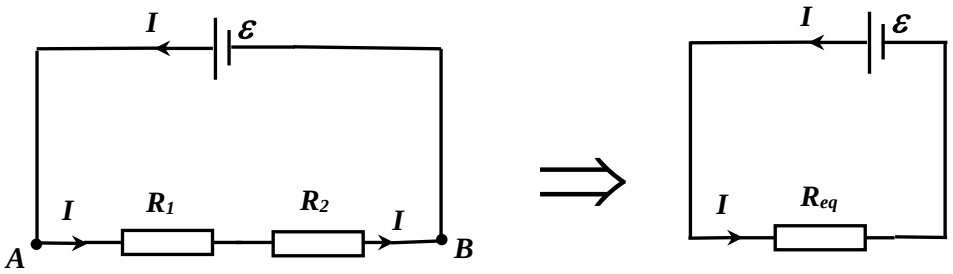
An "ideal" ammeter has zero resistance; an "ideal" voltmeter has infinite resistance.

A real ammeter has a resistance which can range from near zero to 1 Ω; a real voltmeter has a resistance, which can range from 20 kΩ to 2 MΩ.

13.5.5 Effective Resistance for Combination of Resistors

Resistors in series

Consider two resistors R_1 and R_2 which are connected in series across A and B.



$$\varepsilon = IR_1 + IR_2$$
$$\varepsilon = I(R_1 + R_2)$$
$$\frac{\varepsilon}{I} = (R_1 + R_2)$$
$$R_{eq} = R_1 + R_2$$

For n resistors connected in series:

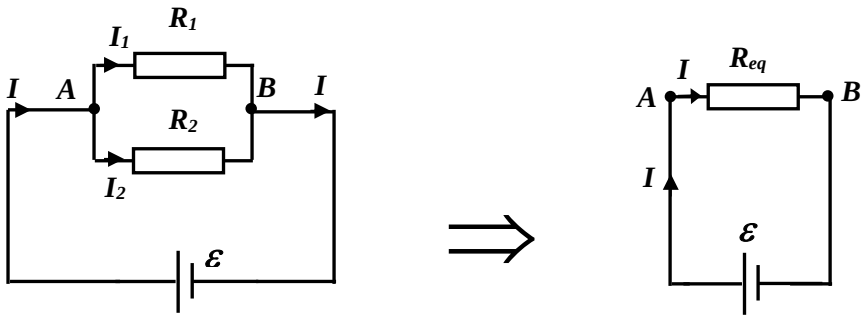
$$\boxed{R_{eq} = \sum R_n}$$

Characteristics of circuits consisting of ***resistors in series*** are;

(i) the **same current** flows through each of the resistors.
(ii) the total p.d across all resistors in series. = sum of individual p.d. across each resistor.
(iii) individual p.d. is directly proportional to individual resistors.
(iv) **equivalent resistance = sum of individual resistances**.

Resistors in parallel

Consider two resistors R_1 and R_2 which are connected in parallel across A and B. A current I enters at A and leaves at B, setting up a potential difference V between them. The current branches into I_1 and I_2 through the two resistors.



$$I = I_1 + I_2$$

$$\frac{\varepsilon}{R_{eq}} = \frac{\varepsilon}{R_1} + \frac{\varepsilon}{R_2}$$

The reciprocal of the equivalent resistance R_{eq} across AB is

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2}$$

Hence, for n resistors connected in parallel:

$$\boxed{\frac{1}{R_{eq}} = \sum\left(\frac{1}{R_n}\right)}$$

Characteristics of circuits consisting of **resistors *in parallel combination*** are;

(i) the potential difference is the same across each of the resistors.
(ii) total current across parallel combination of resistors = sum of individual currents across each resistor.
(iii) the larger a resistor the smaller the current through it.
(iv) equivalent resistance is smaller than the least of the individual resistances.
(v) **reciprocal of the equivalent resistance = sum of the reciprocals of individual resistances**.

For two particular cases, the formula can be simplified, making for greater speed in calculation:

(a) The combined equivalent resistance of ***n*** **EQUAL resistances *R* in parallel**:

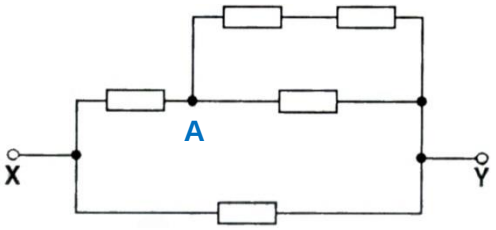
$$R_{eq} = R / n$$

(b) For **two resistors in parallel**:

$$R_{eq} = \frac{R_1 R_2}{R_1 + R_2}$$

Example 4 (N89/I/14)

The circuit diagram shows a network of resistors each of resistance R.
What is the effective resistance between the points X and Y?



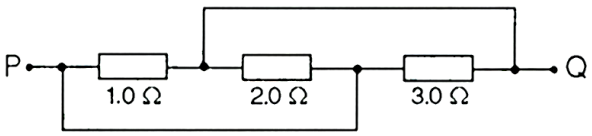
(A) $\frac{2R}{7}$ (D) $\frac{2R}{3}$

(B) $\frac{R}{2}$ (E) $\frac{3R}{4}$

(C) $\frac{5R}{8}$

Example 5 (J89/I/15)

Three resistors are connected as shown in the diagram using connecting wires of negligible resistance. What is the approximate resistance between points P and Q?



Helpful Tips

The connecting wires are taken as ideal conductors (i.e., R = 0 Ω), unless otherwise stated. Thus two points joined by a connecting wire would be at the same potential.

13.5.6 Power dissipation in resistors (Joule heating)

From the definitions of resistance and potential difference, we have learnt:

For an electric device of resistance R carrying a steady current I:
The potential difference V across the device is: $V = IR$
The power dissipated by the device is: $P = IV$

When an electric current passes through a resistor, electrical energy is irreversibly transformed into thermal energy (Joule heating). The following two equations describe the **rate of transfer of electrical energy** to heat generated in a resistor:

$$P = I(IR) = I^2R$$

$$P = \frac{V}{R}V = \frac{V^2}{R}$$

Therefore, we have three expressions for the power dissipated by the device, P:

$$\boxed{P = VI = I^2R = V^2/R}$$

The last two of these formulae ($P = I^2R = V^2/R$) can apply *only to the production of heat* in resistors. The formula ($P = VI$) can be applied to *determine the rate of production of energy in any form* in a circuit component in which a current I flows at a potential difference V.

Example 6

The power output at a power station is 500 MW at a p.d. of 132 kV. If the total resistance of the complete power line is 4.0 Ω, calculate

(i) the current flowing in the power line,
(ii) the voltage drop caused by the power line, and
(iii) the percentage loss of power in transmission.

Before solving, sketch a schematic representation of this circuit and label all given values in your schematic.

Students often apply the wrong formula to compute the required power dissipated by a particular component in a circuit as they fail to identify the correct V, I or R in the circuit.

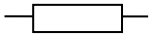
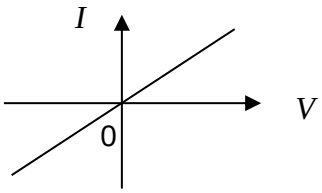
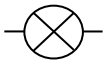
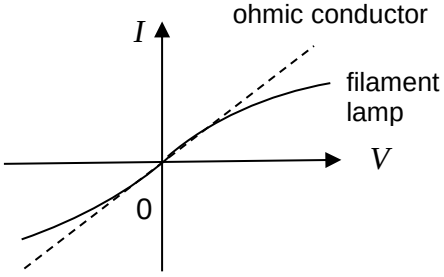
Note that V is the p.d. across the identified component. In questions asking to determine the power loss across the transmission cables, the power loss (through heating) in cables is usually determined using $P = I^2R$ (where I is the current flowing in the cables and R is the resistance of the cables) as it is often not possible to determine V across the transmission cables.

If the value of V across the transmission cables can be measured, the power loss in the cable can then also be determined using $P = V^2/R$.

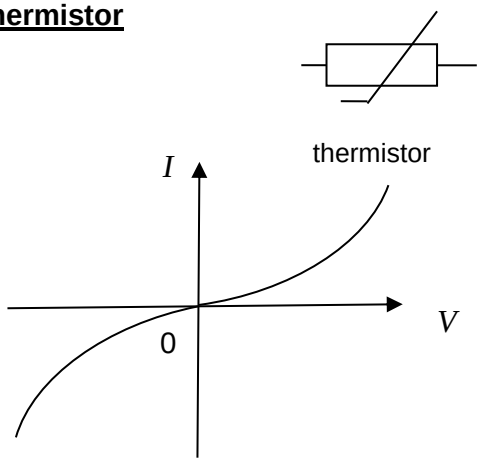
13.6 *I-V* Characteristics of Various Electrical Components

It is often useful to understand how an electrical component (e.g. resistor, diode, etc) would behave or function in an electrical circuit. This information is often obtained from the I-V characteristics graphs of the required electrical component. Thus it is often useful and critical (for GCE A Level exams) to be able to understand and interpret the I-V characteristics graphs of different electrical components.

It is important to note that for any I-V characteristics graph, the I represents the current passing through the component and V represents the p.d. across the component. Thus, the resistance of the component (for a particular set of I and V values) is obtained by directly dividing V by I.

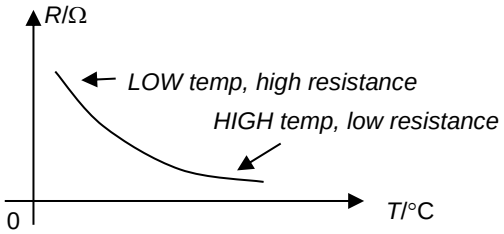
Device	Characteristics
Metallic Conductor  	Metal conductors are ohmic conductors if their physical conditions remain constant. Experimentally, it can be shown that if the temperature of the conductor is kept constant, its resistance will be constant. Any increase in the p.d. produces a proportionate increase in the current. The **ratio of V to I is constant and the conductor has a single value of resistance**. The I-V graph is a **straight line passing through the origin**. Microscopically, the resistance of the metal conductor is due to collisions between the free electrons and the metal ions in the lattice. When the temperature of the conductor is kept constant, the amplitude of lattice ion vibrations will remain the same. This means that the hindrance to the flow of electrons is fixed and the resistance will remain constant.
Filament Lamp  	When a larger p.d. is applied, the temperature of the filament increases due to the heating effect of the current. The **resistance of the filament increases with temperature** at larger values of p.d. and current. The collisions between the electrons and lattice ions cause energy to be transferred to the lattice ions, enabling the ions to vibrate with larger amplitudes – this implies an increase in the temperature of the filament, and beyond a certain point, the filament glows. The increased vibration of the lattice ions presents a greater hindrance to the flow of electrons, resulting in an increase in the resistance of the filament. Hence at any particular value of p.d. (except at low values of p.d. where there is no significant increase in temperature – any energy gained by the filament is dissipated to the surroundings quickly enough), the current in a filament is lower than that in an ohmic conductor.

Thermistor



Usually, a thermistor is a semi-conductor device. Unlike the filament lamp, its **resistance decreases with increasing temperature**.[2] The graph below shows the resistance-temperature characteristic of a thermistor.

R-T graph for a thermistor

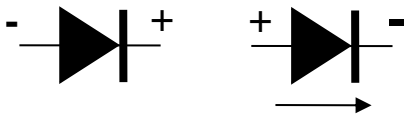


As a larger p.d is applied, the current through the thermistor increases, causing the thermistor to heat up. Even though there is increased lattice vibrations which increases the hindrance to the flow of the charge carriers (like in the case of the filament), the extra thermal energy supplied to a thermistor *also* causes it to release more charge carriers. The outcome is **a decrease in the ratio of V to I – a fall in its resistance at higher values of p.d.**

The thermistor's response is rapid and most sensitive to changes to temperature in the region of 200 K to 600K; it is a more limited range compared to thermocouples or resistance thermometers.

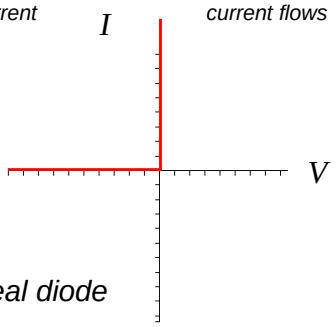
The thermistor is commonly used in very low temperature measurements as the high sensitivity (dR/dT) at low temperature enables more accurate measurements to be made. It can also be used to operate fire alarms and thermostats.

Semiconductor Diode

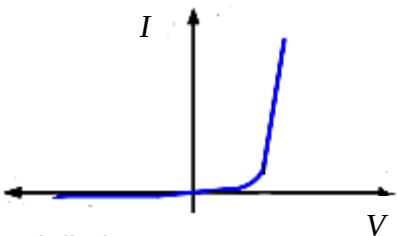


Reverse bias – no current

Forward bias – current flows



An ideal diode



A real diode

A *p-n junction diode* is a semiconductor which conducts well in one direction of p.d. (**forward-bias** direction) but conducts poorly when the p.d. is reversed (**reverse-bias** direction). In an ideal diode (which does not exist in real life), if the diode is reverse biased, the current flowing through it is zero. The ideal diode starts conducting at 0 V and for any positive voltage an infinite current flows and the diode acts like a short circuit (a "perfect" wire).

A real diode needs a minimum positive voltage, called the turn-on voltage, before it conducts any significant current. The diode has very small resistance. If the p.d. is applied in the opposite direction (*reverse biased*), almost zero current flows, and the diode has very high resistance.

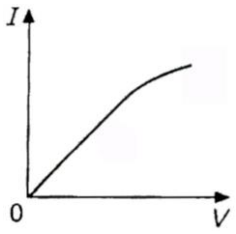
Note that if the reverse-biased p.d. is too high, the diode may break down, resulting in a short circuit.

A diode is used as a rectifier for changing an alternating current to a direct current (greater detail is in the topic Alternating Current).

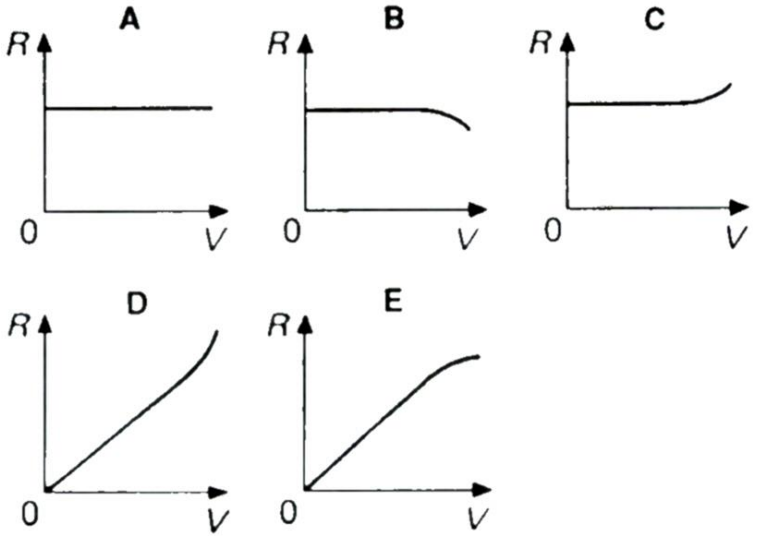
[2] This is what we call a negative temperature coefficient (NTC) thermistor. There are also positive temperature coefficient thermistors, whose resistance increases with temperature.

Example 7 (N91/I/16)

The current *I* flowing through a component vary with the potential difference *V* across it as shown.



Which graph best represents how the resistance *R* varies with *V*?



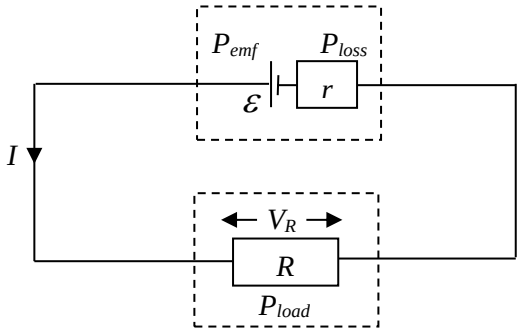
13.7 Internal resistance of a source of e.m.f.

All practical sources of e.m.f. have internal resistance. The chemicals within a cell offer a resistance to current flow, known as the internal resistance of the cell.

13.7.1 Terminal potential difference, e.m.f and internal resistance

Consider a simple circuit with a cell of e.m.f. $\mathcal{E}$ and a load of resistance R. From the energy point of view, the cell is the supplier while the load is the consumer of electrical energy. However, the cell has internal resistance which would consume part of the energy. Hence, less energy is available to drive each coulomb round the external circuit.

A cell can be represented as an ideal source of e.m.f. in series with a resistor representing the internal resistance of the cell. This internal resistance *cannot be* separated out or measured directly.



P_{emf} = *power supplied by e.m.f*
P_{loss}= *power loss due to internal resistance*
P_{load} = power dissipated in the load

By the law of conservation of energy,

Energy supplied by the e.m.f. = total energy consumed by the load and internal resistance of the cell

Since energy is supplied continuously, it follows that:

Power supplied by the e.m.f. = total power consumed by the load and internal resistance of the cell

$$P_{emf} = P_{load} + P_{loss}$$
$$I\mathcal{E} = IV_R + IV_r$$
$$I\mathcal{E} = IV_R + I^2 r$$
$$\mathcal{E} = V_R + Ir$$
$$V_R = \mathcal{E} - Ir$$

Thus, the <u>*potential difference across the terminals of a cell*</u> (***terminal p.d***.) of a source supplying current to an external circuit is always ***smaller than its e.m.f.*** due to the presence of the internal resistance, *r*.

If the circuit is open, no current will flow through the circuit and the terminal p.d. of the source will be equal to the e.m.f. of the source.

The output efficiency of the above circuit is given by:

$$Efficiency = \frac{\text{useful power output}}{\text{power input}}$$

$$Efficiency = \frac{P_{load}}{P_{emf}} = \frac{I^2 R}{IE} = \frac{IR}{E} = \frac{\left(\frac{E}{R+r}\right)R}{E} = \frac{R}{R+r}$$

Example 8

A complete circuit consists of a 24.0 V battery, a 5.60 Ω resistor and a switch. The internal resistance of the battery is 0.28 Ω. The switch is opened. What does a voltmeter read when placed

(a) across the terminals of the battery?
(b) across the resistor?
(c) across the switch?

Repeat (a), (b) and (c) for the case when the switch is closed.

13.8 The Potential Divider Principle

A potential divider is an arrangement of resistors which is used to supply varying levels of p.d. to a circuit from fixed e.m.f. sources.

Consider two resistors connected in series.

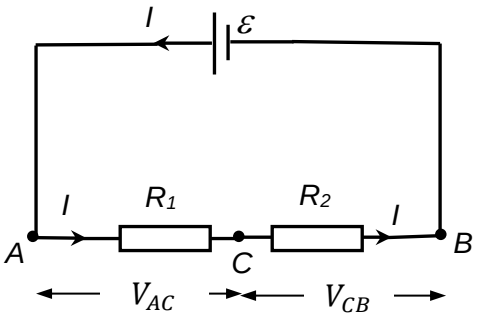


Figure 13.8.1

The current I flowing in the circuit is given by

$$I = \frac{\varepsilon}{R_1 + R_2}$$

Therefore, the potential difference across R_1, V_{AC} is

$$\begin{aligned} V_{AC} &= IR_1 \\ &= \left(\frac{\varepsilon}{R_1 + R_2}\right)R_1 \\ &= \frac{R_1}{R_1 + R_2}\varepsilon \end{aligned}$$

The potential difference across R_2, V_{CB} is

$$\begin{aligned} V_{CB} &= IR_2 \\ &= \left(\frac{\varepsilon}{R_1 + R_2}\right)R_2 \\ &= \frac{R_2}{R_1 + R_2}\varepsilon \end{aligned}$$

Hence, by adjusting the ratio of R_1 to R_2, we can obtain the desired potential difference across AC.

Note: It is assumed that this cell has no internal resistance. If the cell has internal resistance, a proportionate amount of voltage will drop across the internal resistance.

Instead of two resistors, a **rheostat** can be used as in Figure 13.8.2 as a source of variable potential difference.

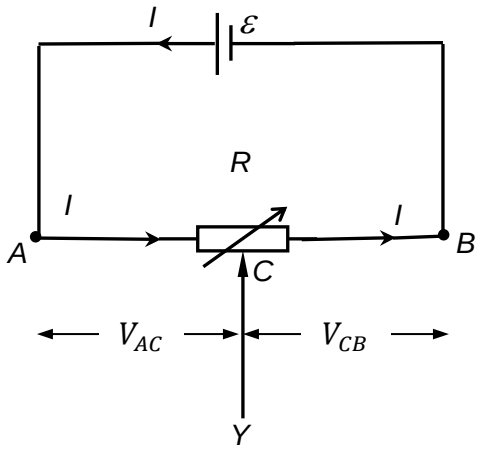


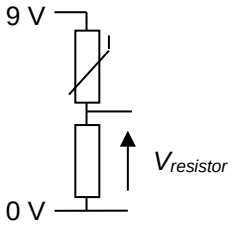
Figure 13.8.2

By sliding the contact/slider Y on the rheostat R, V_{AC} can be made to vary from 0 (when Y is tapped at the point of zero resistance) to ε (when Y is tapped at the point of maximum resistance), assuming the driver cell has negligible internal resistance.

13.8.1 Use of thermistors and light dependent resistors in potential dividers

Thermistors

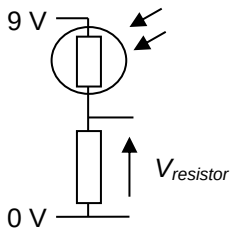
One of the resistors in Figure 13.8.1 can be substituted with a thermistor. When the temperature increases, the resistance of the thermistor decreases, which in turn causes a decrease in the p.d. across the thermistor ($V_{thermistor} = \frac{R_{thermistor}}{R_{thermistor} + R_{resistor}} \varepsilon$) and an increase in the p.d. across the other resistor ($V_{resistor} = \frac{R_{resistor}}{R_{thermistor} + R_{resistor}} \varepsilon$).



One application of a thermistor is its use as a temperature sensor in a fire alarm. When hot conditions are detected, $V_{resistor}$ increases, causing a fire alarm circuit to be triggered.

Light Dependent Resistors

One of the resistors in Figure 13.8.1 can be substituted with a light dependent resistor (LDR). A LDR is a device whose resistance decreases as the intensity of light incident on it increases. Under bright conditions, the p.d. across the LDR ($V_{LDR} = \frac{R_{LDR}}{R_{LDR} + R_{resistor}} \varepsilon$) decreases, while the p.d. across the other resistor in the circuit increases ($V_{resistor} = \frac{R_{resistor}}{R_{LDR} + R_{resistor}} \varepsilon$).

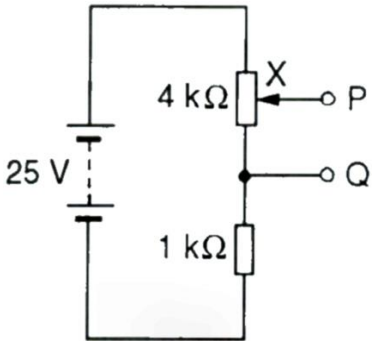


One could use a LDR to control lighting circuits which are switched on automatically in the evening when illumination falls below a certain intensity.

Example 9 (J86/I/15)

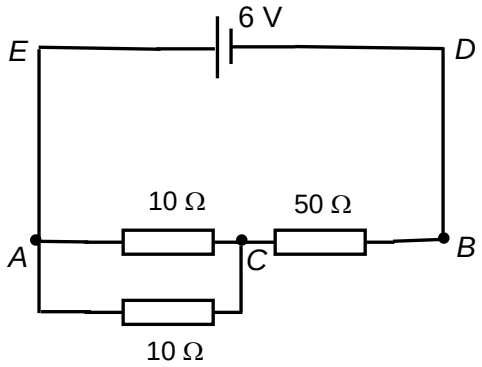
The diagram on the right shows a potential divider circuit which, by adjusting the position of the contact X, can be used to provide a variable potential difference between the terminals P and Q. What are the limits of this potential difference?

A 0 and 25 mV
B 5 mV and 25 mV
C 0 and 20 V
D 0 and 25 V
E 5 V and 25 V



Example 10

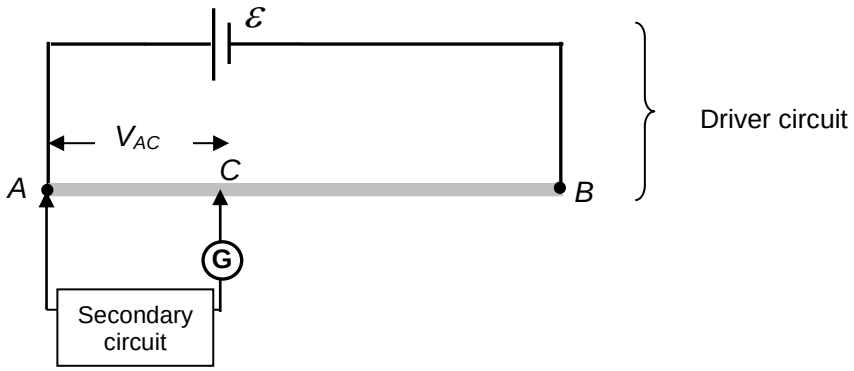
(a) Determine the potential difference across the junction AC in the following circuit.



(b) Determine the reading of a voltmeter of resistance 200 Ω placed across AC.
(c) Determine the reading of an ammeter of resistance 10 Ω placed in the path BD.

13.9 The Potentiometer

The potentiometer comprises a resistance wire AB and a jockey that is connected to the wire which can be shifted up and down the wire. It is usually connected to a driver circuit and a secondary circuit (the secondary circuit will be elaborated in a later section). In its most basic form, the driver circuit consists of a driver cell, $\mathcal{E}$ and a long resistance wire, shown as AB below.



A voltmeter connected across AC will measure a fraction of the potential difference $\mathcal{E}$. The fraction of $\mathcal{E}$ measured out depends on where point C – a metallic contact called a jockey – is placed along AB.

The principle on which a potentiometer works is exactly the same as for a potential divider. Assuming that the resistance wire AB has a uniform cross-sectional area and a length l_{AB}, the resistance R_{AC} will be proportional to the length l_{AC}, since $R = \frac{\rho l}{A}$, or $R \propto l$. Therefore,

$$V_{AC} = \frac{R_{AC}}{R_{AC} + R_{CB}} \varepsilon = \frac{l_{AC}}{l_{AB}} \varepsilon$$

A Balanced Potentiometer

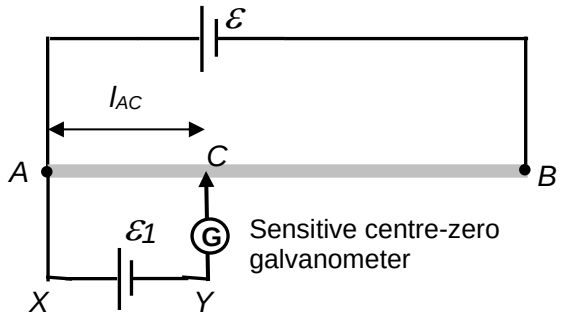
The potentiometer is said to be **balanced** when the jockey is at a position on AB such that there is **NO current** through the galvanometer (essentially an ammeter). This means that there is no current or charge flow between the driver circuit and the secondary circuit, i.e. the driver circuit and secondary circuit are electrically isolated. The applications of the potentiometer that we will discuss in the next section rest on finding this balanced point.

13.9.1 Applications of the potentiometer

The secondary circuit can take on many forms, depending on the application of the potentiometer.

Application 1: To determine the unknown e.m.f. of a cell

The secondary circuit is just composed of the unknown e.m.f. source $\mathcal{E}_1$.



By adjusting the length l_{AC} , we change the p.d. V_{AC} .

At the point where the galvanometer shows null deflection (no current flows through it and the potentiometer is balanced),

$$\mathcal{E}_1 = V_{XY} = V_{AC}$$

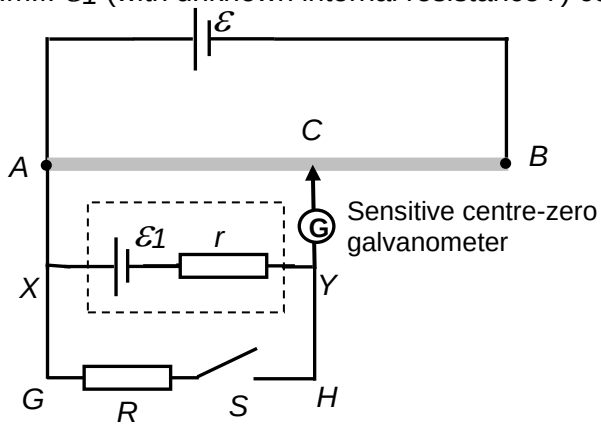
$$\mathcal{E}_1 = \frac{l_{AC}}{l_{AB}}\mathcal{E}$$

The length l_{AC} is known as the balanced length. By measuring l_{AC}, with known l_{AB} and $\mathcal{E}$, $\mathcal{E}_1$ can be found.

Since there is no current flowing through the galvanometer, the potentiometer essentially functions as an ideal voltmeter!

Application 2: To determine the unknown internal resistance of a cell

A driver cell with e.m.f. $\mathcal{E}$ has negligible internal resistance. The secondary circuit consists of an unknown e.m.f. $\mathcal{E}_1$ (with unknown internal resistance *r*) connected in series with an external resistor R.



When **switch *S* is open**, there is no current flowing in XGHY.
When the galvanometer shows null deflection (no current flowing through AXYC),

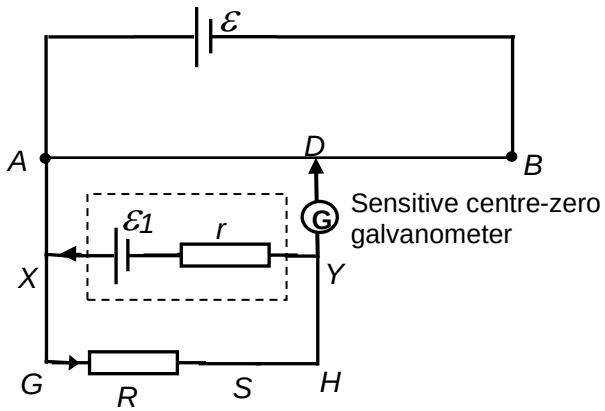
$$V_{XY} = \mathcal{E}_1 \text{ (since there is no p.d. across } r\text{)}$$

Since $V_{AC} = V_{XY}$, $V_{AC} = \mathcal{E}_1$

From the driver circuit, $V_{AC} = \frac{l_{AC}}{l_{AB}} \mathcal{E}$

Therefore, $\mathcal{E}_1 = \frac{l_{AC}}{l_{AB}} \mathcal{E}$ [1]

When we **close the switch** S, a current flows in the circuit XGHY.



At a new balanced length AD (no current flows **between** the driver and secondary circuits, even though a current flows in the driver circuit and a (possibly different) current flows in the secondary circuit),

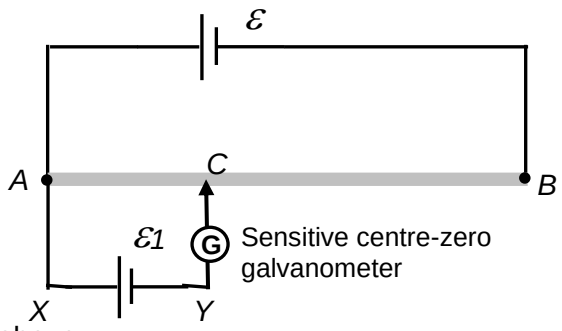
$$V_{XY} = V_{GH} = \frac{R}{R+r}\mathcal{E}_1$$

Since $V_{AD} = V_{XY}$, $\frac{l_{AD}}{l_{AB}} \mathcal{E} = \frac{R}{R+r} \mathcal{E}_1$ [2]

Since $\mathcal{E}$, R, l_{AC}, l_{AD} and l_{AB} are known, $\mathcal{E}_1$ and *r* can be calculated using equations [1] and [2].

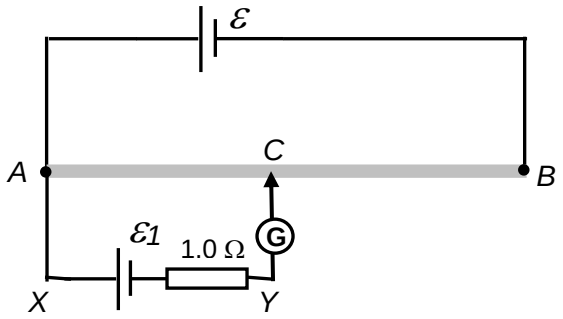
Example 11

In the circuit shown below, cell $\mathcal{E}$ has an e.m.f. of 2.0 V and negligible internal resistance. AB is a uniform wire of length 1.00 m and resistance 5.0 Ω. Cell $\mathcal{E}_1$ has an e.m.f. of 1.5 V and internal resistance of 0.80 Ω. Calculate the length AC required to produce zero deflection in the galvanometer

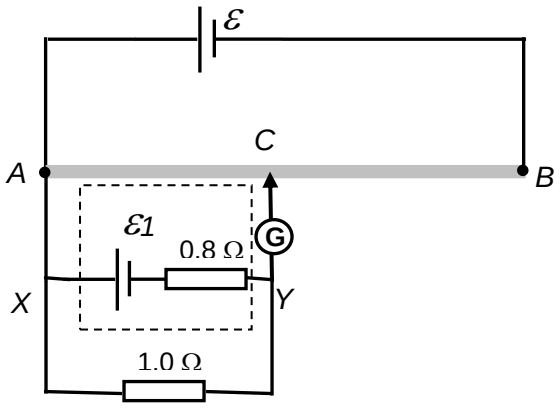


(i) in the circuit as shown above,

(ii) when a resistor of 1.0 Ω is placed *in series* with $\mathcal{E}_1$ as shown below.



(iii) when a resistor of 1.0 Ω is connected with $\mathcal{E}_1$ as shown below.

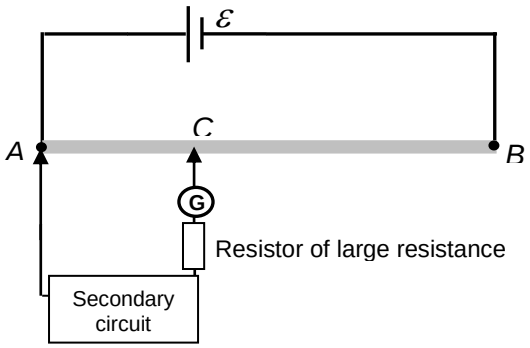


Practical pointers on the potentiometer

- The e.m.f.s of both the driver cell and the unknown cell should be constant throughout. If not, the balance length will fluctuate.

- The jockey should not be scraped along wire AB to determine the balance point, because doing so could damage the uniformity of the wire. Instead gently taps the jockey along the wire AB.

- A very sensitive galvanometer is needed if precise locations of the balance points are to be made. Such a galvanometer should be protected, by a large (typically 1 MΩ) series resistance, from the relatively high current which would flow through it when the jockey is far away from the balanced point. The presence of this resistor will not affect the balance length since at balanced point, no current flows through the galvanometer (and the resistor of large resistance).

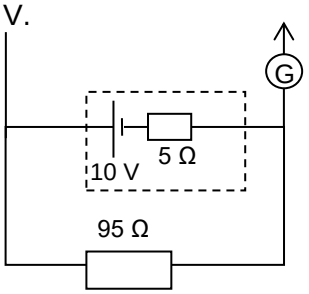


Advantages of using a potentiometer to measure voltage

- It draws no current. Therefore, it functions as an ideal voltmeter.

  When used to measure an unknown e.m.f., the value obtained at balanced point is the true e.m.f. If a practical voltmeter of resistance 1000 Ω is used across an e.m.f. source of 10 V and internal resistance 5 Ω, the practical voltmeter will read only $V = \frac{1000}{1005} \times 10 =$ 9.95 V.

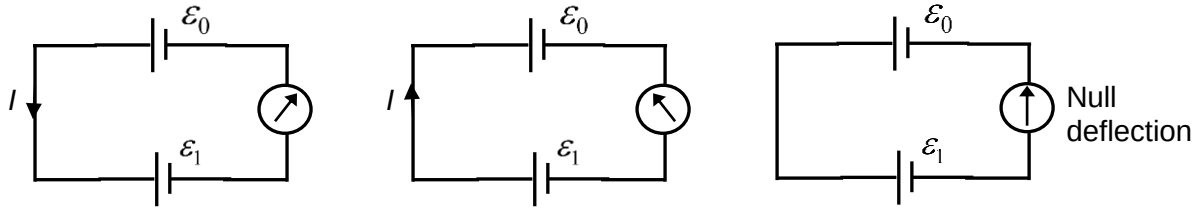
- When used to measure the terminal p.d., the value obtained at balanced point is the true terminal p.d. – which would be 9.5 V for the circuit below. If a practical voltmeter of resistance 1000 Ω is used instead to measure the terminal p.d., the effective external resistance would be $(\frac{1}{95} + \frac{1}{1000})^{-1}$ and the voltmeter would read only 9.46 V.



Disadvantages of using a potentiometer to measure voltage

- It is slow in operation. It cannot be used to monitor rapidly changing voltages.
- The resistance wire may not be uniform due to wear and tear.

It should be appreciated that the measurement using a potentiometer is basically a method of comparison.



$\varepsilon_0 > \varepsilon_1$ $\varepsilon_0 < \varepsilon_1$ $\varepsilon_0 = \varepsilon_1$

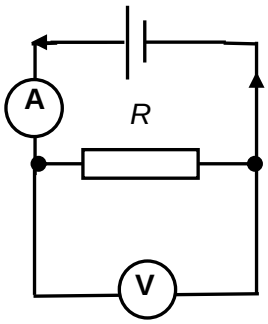
APPENDIX A – Two different arrangements of voltmeter and ammeter to find a more accurate value of the resistance of an unknown resistor.

If the meters are not ideal, the significance of the error in the meter readings can be reduced by using the following circuit arrangements:

<u>If the unknown R is expected to be low</u>

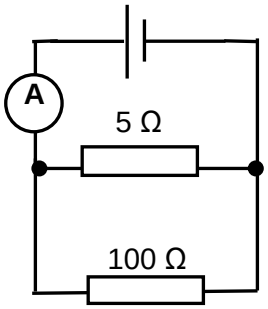
The voltmeter reads the correct value of p.d. across R.
While the ammeter reads the sum of the current flowing through both R and the voltmeter, because R is low, most of the current actually passes through R rather than the voltmeter – hence the error in the ammeter reading will not be too great.

If this circuit were used for a high R (perhaps comparable to the resistance of the voltmeter) instead, the ammeter will not be measuring a value of current close to the current actually flowing through R – the error in the ammeter reading will be more significant (the voltmeter still reads the p.d. across R).

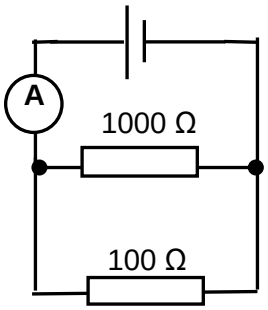


No current passing through ideal voltmeter

Ammeter reading deviates from the true value by about 5%



Voltmeter with resistance



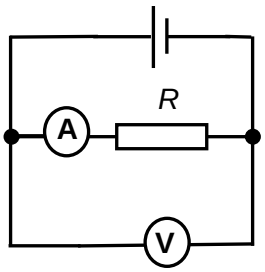
Ammeter reading deviates from the true value by about 91%

Voltmeter with resistance

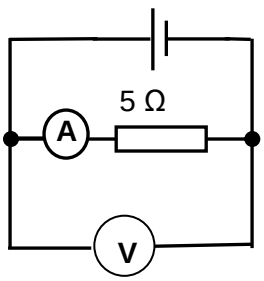
<u>If the unknown R is expected to be high</u>

The ammeter reads the correct value of current flowing through R.
While the voltmeter reads the sum of the p.d. across both R and the ammeter, because R is high, most of the p.d. actually falls across R – hence the error in the voltmeter reading will not be great.

If this circuit were used for a low R (perhaps comparable to the resistance of the ammeter) instead, the error in the voltmeter reading will be more significant (the ammeter still reads the current through R).

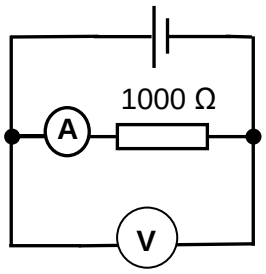


Ammeter with 1 Ω resistance.



Voltage reading of 5 Ω resistor would deviate by 20% from the true value.

Ammeter with 1 Ω resistance .

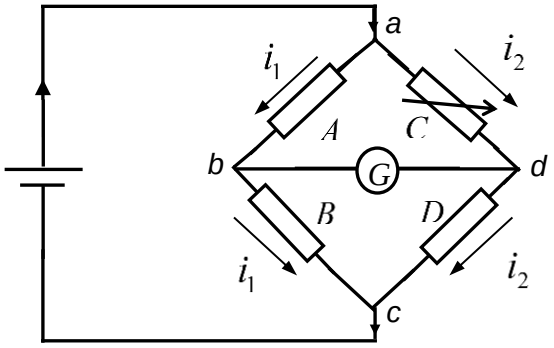


Voltage reading of 1000 Ω resistor would deviate only by 0.5% from the true value.

APPENDIX B – The Wheatstone bridge (good to know)

Unknown resistances can be accurately measured using a circuit known as a Wheatstone bridge.

This circuit consists of the unknown resistance A, three known resistors B, C and D (where C is a variable resistor), a galvanometer, and a source of e.m.f.



Resistor C is varied until the galvanometer reading is zero (null deflection), i.e. there is no current flowing through the galvanometer between points b and d. Under this condition of null deflection, the bridge is said to be balanced, and the current which flows in A flows in B as well; the current which flows in C flows in D as well.

Since point a is common and the potential at b = the potential at d, the potential difference across ab = potential difference across ad. Likewise, the potential difference across bc = potential difference across dc .

$$i_1 A = i_2 C \qquad (1)$$

$$i_1 B = i_2 D \qquad (2)$$

(1)/(2)
$$\frac{A}{B} = \frac{C}{D}$$

Since B, C and D are known, A can be found.

Advantages

- Variation in the e.m.f. of the supply does not affect the balance point.

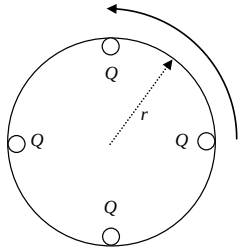
Disadvantages

- This method is unreliable for finding small resistances (< 1 Ω) because the unknown resistances of the (real) wires used to connect the resistors into the bridge are significant.
- This method is insensitive for high resistances (~10^6 Ω) unless a highly sensitive galvanometer is used. This is because the presence of high resistances in the circuit causes the current through the galvanometer to be near zero even when the potentials of points *b* and *d* are far from each other.

Tutorial 13A: Current of Electricity

Self-Review Questions

S1. **J01/I/14** Four small conductors, on the edge of an insulating disc of radius *r* are each given a charge of *Q*. The frequency of rotation of the disc is *f*. What is the equivalent electric current at the edge of the disc?



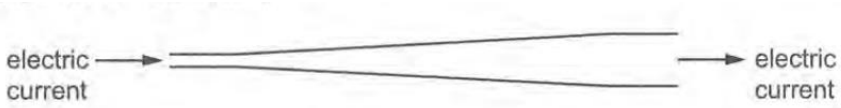
A $4Qf$ **B** $\frac{4Q}{f}$ **C** $8\pi rQf$ **D** $\frac{2Qf}{\pi r}$

S2. **Calculations for A-Level Physics, T.L. Lowe & J.F. Rounce**.
Calculate the mean velocity of electron flow (i.e. the drift velocity) in a wire where the mobile electron density is 5 x 10^{28} m^{-3} if the current is 1 A and the wire has uniform cross-sectional area of 1 mm^2. (Electron charge, *e* = -1.6 x 10^{-19} C)

S3. **N02/I/22** The electron beam current in a cathode-ray oscilloscope is 50 μA. The time-base causes the beam to sweep horizontally across the screen at a speed of 1.0 x 10^4 cm s^{-1}. What is the number of electrons arriving at the screen in one centimetre length of the horizontal trace?

A 3.1 x 10^{10} **B** 3.1 x 10^{13} **C** 1.3 x 10^{13} **D** 8.0 x 10^{14}

S4. **N22/I/20** A length of metal wire with a circular cross-section is connected into an electric circuit. The wire carries a constant electric current.



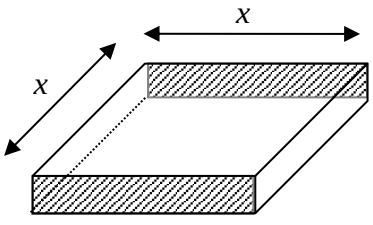
Part of the wire has a diameter that increases uniformly with distance along the wire. The drift velocity v of the charge carriers is related by a constant K to the diameter d of the wire.

Which equation for v is correct?

A $v = K/d^2$ **B** $v = K/d$ **C** $v = Kd$ **D** $v = Kd^2$

S5 A sample of resistive material is prepared in the form of a thin square slab of side x. For a given thickness, the resistance between the opposite edge faces of the sample is

A proportional to x^2.
B proportional to x.
C independent of x.
D inversely proportional to x.
E inversely proportional to x^2.



S6. **N11/III/8a** Distinguish between electrical resistance and resistivity [4]

S7. **Halliday, Resnick & Walker** A wire with a resistance of 6.0 Ω is drawn out through a die so that its new length is three times its original length. Find the resistance of the longer wire, assuming that the resistivity and density of the material are unchanged.

S8. A generator produces 100 kW of power at a potential difference of 10 kV. The power is transmitted through cables of total resistance 5.0 Ω. How much power is dissipated in the cables?

A 50 W **B** 250 W **C** 500 W **D** 50 000 W

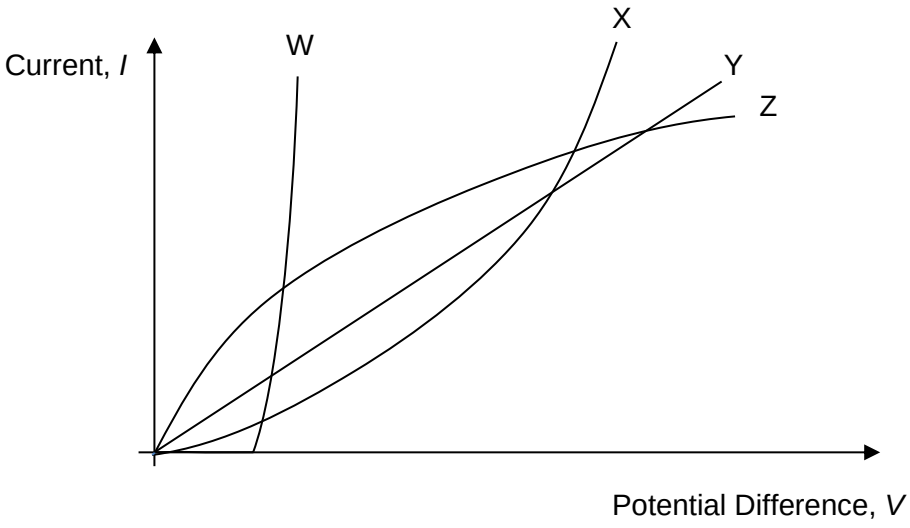
S9. **N00/III/4d&e** A water heater is marked 230 V, 3000 W. It is switched on for 5000 seconds. For this heater, calculate

(i) the current through the heater,
(ii) the resistance of the heater,
(iii) the energy supplied by the heater during this time

S10. A cell of e.m.f. E delivers a charge Q to an external circuit. Which statement is correct?

A The energy dissipation in the external circuit is EQ.
B The energy dissipation within the cell is EQ.
C The external resistance is EQ.
D The total energy dissipation in the cell and the external circuit is EQ.

S11. **H1 2015/P1/19 (modified):** A sketch graph of *I*-*V* characteristics is shown for four different conductors, namely: metallic conductor at a constant temperature, a semiconductor diode, a filament lamp, and a NTC thermistor. Identify which graph corresponds to each conductor.

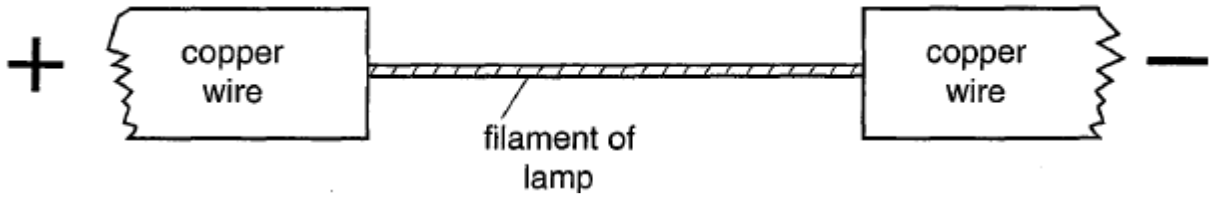


S12 Household electric power in Singapore is supplied at 240 V, rather than the 120 V that is standard in USA and Canada. What are the advantages and disadvantages of each system?

Discussion Questions

D1. **H1/2013/P2/8(part)**

(c) Fig. 8.1 illustrates the connection between the copper leads to the filament and the filament itself.



Describe, giving reasons where appropriate, and without using numerical values, the following features of the current in these three components:

(i) the direction of electron flow [1]
(ii) the rate of electron flow [1]
(iii) the average speed of electrons [1]

(d) Use your answers to (c) to explain why the filament of the lamp gets hot but the copper leads stay relatively cold. [2]

D2. **H2 Phy/2018/P2/Q5**

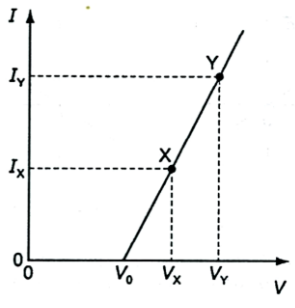
(a) Copper has one conduction electron per atom.
The density of copper is 8960 kg m^{-3}.
The mass of 1.00 mole of copper is 63.5 g.
Show that the number density of charge carriers in copper is 8.49 x 10^{28} m^{-3}. [3]

(b) A tungsten wire of diameter 0.36 mm has resistance 30 Ω. The power dissipated in the wire is 5.0 W.
The number density of charge carriers in tungsten is 3.4 x 10^{28} m^{-3}.
Calculate the average drift velocity *v* of the conduction electrons in tungsten. [4]

(c) The wire in (b) is now replaced with another tungsten wire of the same diameter but twice the length. The potential difference (p.d.) across the wire is unchanged. The temperature of both wires is the same.
State and explain the change to the drift velocity of the electrons in the second wire. [3]

D3. **N09/I/27** The graph shows the variation with potential difference V of the current I in an electrical component.

The resistance is measured for current I_Y and for current I_X.
What is the change in the resistance of the component?



A zero

B $\dfrac{V_Y - V_X}{I_Y - I_X}$

C $\dfrac{V_Y}{I_Y} - \dfrac{V_X}{I_X}$

D $\dfrac{V_Y - V_0}{I_Y} - \dfrac{V_X - V_0}{I_X}$

D4. **J89/I/14** An electrical source with internal resistance *r* is used to operate a lamp of resistance *R*. What fraction of the total power is delivered to the lamp?

A $\dfrac{R+r}{R}$ **B** $\dfrac{R-r}{R}$ **C** $\dfrac{R}{R+r}$

D $\dfrac{R}{r}$ **E** $\dfrac{r}{R}$

D5. **J96/III/4(part)**

(a) Use energy considerations to distinguish between electromotive force (e.m.f.) and potential difference (p.d.). [3]

(b) A cell of e.m.f. *E* and internal resistance *r* is connected to a resistor of resistance *R*, as shown in Fig. 10.

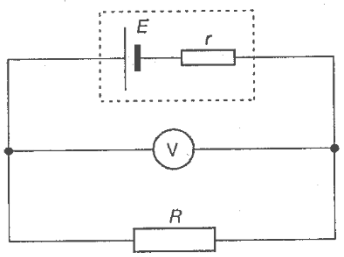


Fig. 10

A voltmeter of infinite resistance is connected in parallel with the resistor.

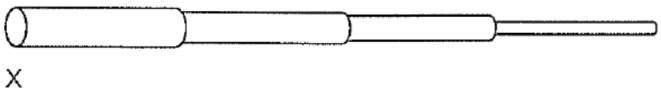
(i) Copy Fig. 10 and include in the circuit a switch so that the voltmeter may be used to measure either the e.m.f. *E* of the cell or the terminal potential difference *V*.

(ii) State whether the switch should be open or closed when measuring
1. the e.m.f.
2. the terminal p.d.

(iii) Derive the relation between *E*, *V*, *r* and the current I in the circuit.

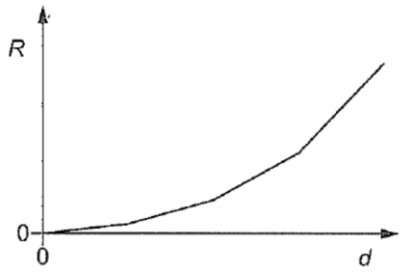
[5]

D6. **N17/I/21** A composite wire is made by connecting in series four uniform wires made of same materials but having different diameters.

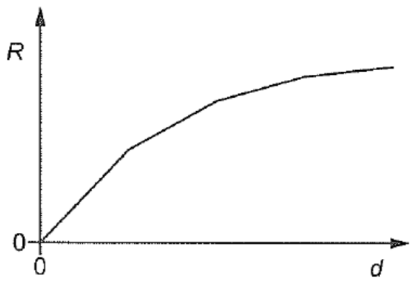


The resistance R of this composite wire is measured between X and other points on the wire at a distance d from X. Which graph best represents the relationship between R and d?

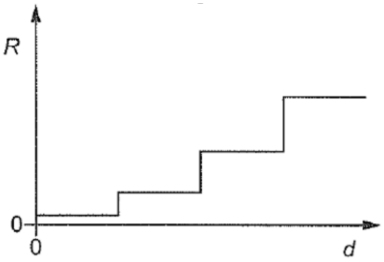
A



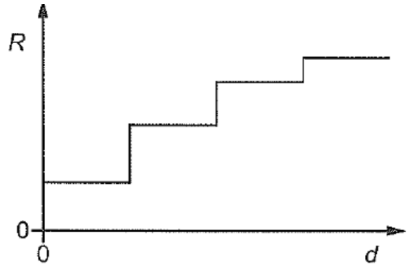
B



C

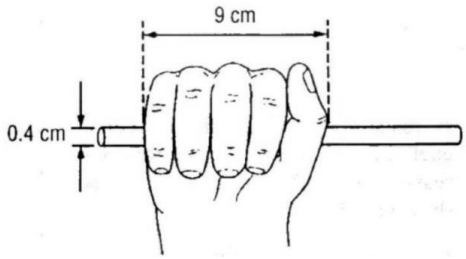


D



D7. The resistivity of the human body is low compared with the resistivity of skin, which is about 3 x 10^4 Ω m for dry skin.

A person, who is well earthed, accidentally grabs a wire of diameter 0.4 cm at a potential of 50 V. His hand makes contact with the whole circumference of the wire over a distance of 9 cm as shown in the figure below.



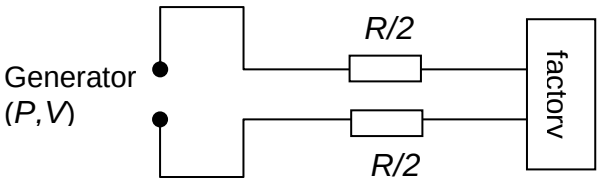
The average thickness of the skin of his hand is 1 mm. Estimate the current through the person.

D8. **J02/I/32** The filament of a 240 V, 100 W electric lamp heats up from room temperature to its operating temperature. As it heats up, its resistance increases by a factor of 16. What is the resistance of this lamp at room temperature?

A 36 Ω **B** 580 Ω **C** 1.5 kΩ **D** 9.2 kΩ

D9. An electrical kettle is rated at 2.0 kW when operating on a 240 V supply. Calculate the power that the kettle will draw if the supply is reduced to 220 V.

D10. **N92/I/14** A generator with output power *P* and output voltage *V* is connected to a factory by cables of total resistance *R*. What is the power input to the factory?



A P **B** $P-\left(\frac{P}{V}\right)\frac{R}{2}$

C $P-\left(\frac{P}{V}\right)R$ **D** $P-\left(\frac{P}{V}\right)^2\frac{R}{2}$ **E** $P-\left(\frac{P}{V}\right)^2 R$

D11. **N01/III/4c** The figure below shows the circuit used to produce the I-V characteristic of the lamp.

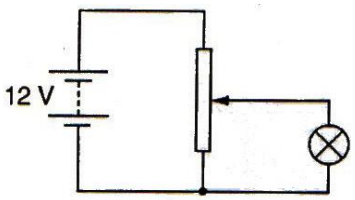
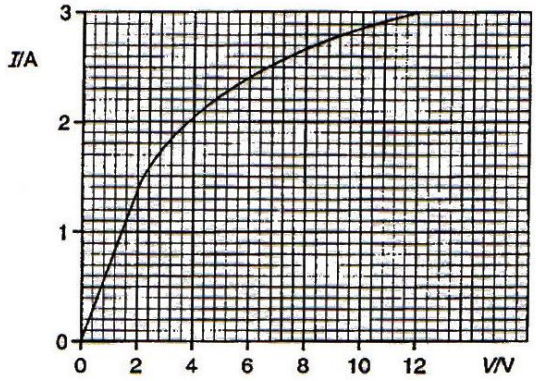


Fig. 4.1



(i) For the lamp operating under normal conditions with a potential difference of 12.0 V across it, calculate
1. the resistance of the lamp,
2. the power supplied to the lamp.

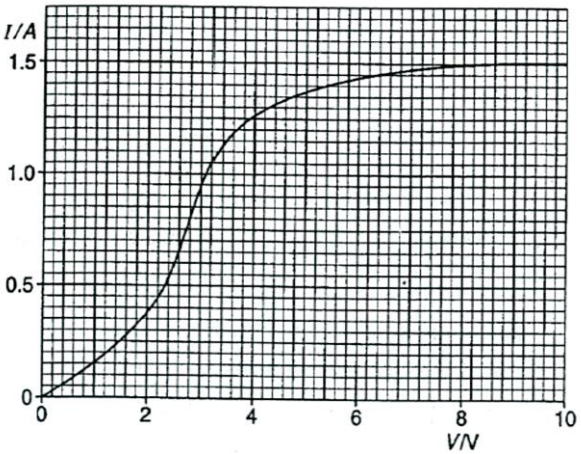
[4]

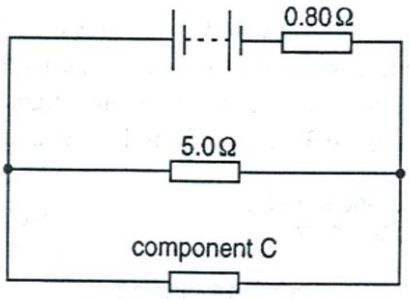
(ii) Calculate the resistance of the lamp when the potential difference across it is only 6.0 V. [1]

(iii) Sketch a graph showing the resistance of the lamp with the potential difference across it. [4]

D12. **N07/III/3**

(a) An electrical component C has an I-V characteristic as shown in the figure below.





(i) Calculate the resistance of component C when a p.d. of 6.0 V is applied across it. [1]

(ii) Deduce the minimum value of the resistance of component C over the range 0 – 10 V. [2]

(b) Component C is then connected into a circuit with a supply of internal resistance 0.80 Ω and a resistor of constant resistance 5.0 Ω, as shown in the diagram on the right.

The current through the 5.0 Ω resistor is found to be 0.85 A.

Calculate

(i) the p.d. across component C, [2]
(ii) the total current from the supply, [2]
(iii) the e.m.f. of the supply, [2]
(iv) the energy supplied to component C in 20 minutes. [3]

D13. **N11/III/8 modified**

(d) A battery has an e.m.f. of 12.0 V and negligible internal resistance. A resistor of resistance 3.0 Ω is permanently connected to one its terminals. The battery and resistor form a power supply with terminals X and Y, as shown in Fig. 8.2.

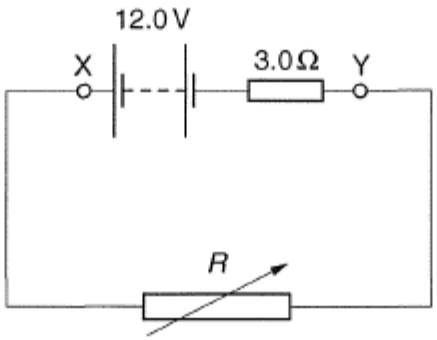


Fig. 8.2

A variable resistor of resistance *R* is connected across terminals X and Y and its resistance is varied. The variation with *R* of the power *P* dissipated in the variable resistor is shown in Fig. 8.3.

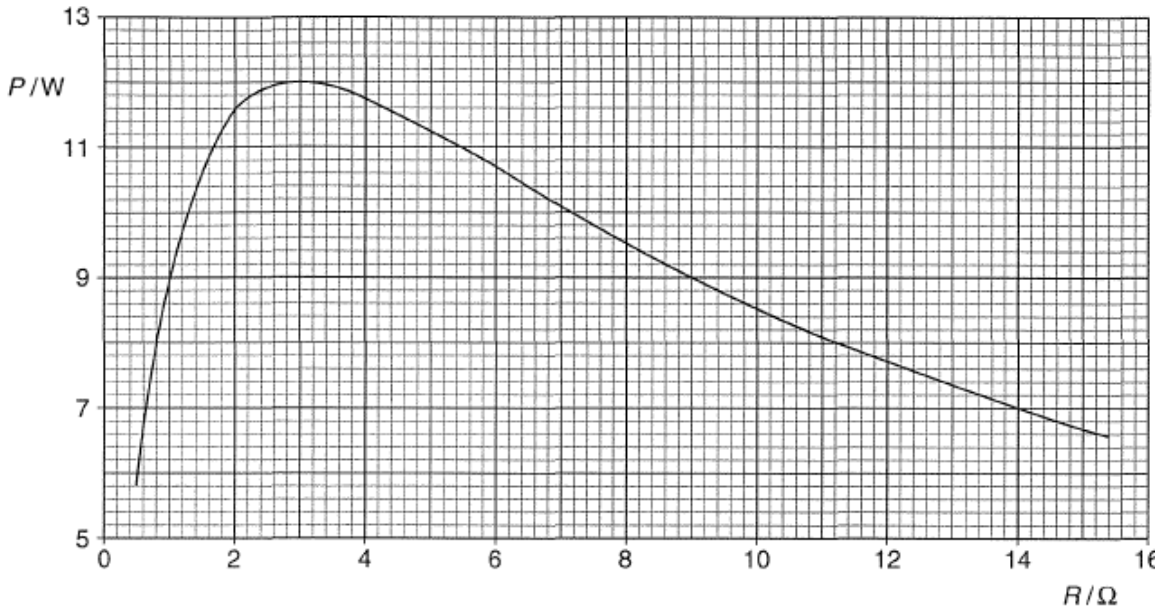


Fig. 8.3

(i) State the two values of R at which the power dissipated in the variable resistor is 9.0 W. [1]

(ii) For each of the two values in (i), calculate the efficiency of transfer of power from the supply to the variable resistor. [5]

$$\text{efficiency} = \left(\frac{\text{Power output}}{\text{Power supplied}} \right) \times 100\%$$

(iii) Calculate the efficiency of transfer of power from the supply to the variable resistor at maximum power *P*. [1]

(e) A battery is to be designed to provide the power for a component that has a fixed resistance. The internal resistance *r* of the battery is to be chosen to provide maximum efficiency of energy transfer to the component. Use your answers in (d) to state and explain the choice of *r* compared to the component's fixed resistance. [2]

D14 **Young & Freedman** A source with e.m.f. E and internal resistance r is connected to an external circuit.

(a) Show that the power output of the source is maximum when the current in the circuit is half the short circuit current of the source.

(b) If the external circuit consists of a resistance R, show that the power output is maximum when $R = r$, and that the maximum power is $\frac{E^2}{4r}$.

D15 **N22/III/2**

At room temperature, a metal wire has a diameter of 1.02mm and a resistance per unit length of 1.73 Ωm^{-1}. The wire is connected to a d.c. power supply. When the power supply is first switched on, the wire heats up from room temperature to a higher constant temperature in a time of 60 s.

(a) Calculate the resistivity of the wire at room temperature. [3]

(b) (i) The first law of thermodynamics, when applied to a system, can be expressed as:

$$\Delta U = q + w$$

where ΔU is the increase in internal energy of the system,
q is the heat supplied to the system and
w is the work done on the system.

Use the words positive, negative and zero to complete Table 2.1 for the three terms in the equation for the wire at the times shown. You may use each word once, more than once or not at all. [2]

Table 2.1

time after being switched on/s	ΔU	q	w
0–59			
60–100			

(ii) At its higher constant temperature, the current in the wire is 12A and the potential difference across it is 230 V.

Complete Table 2.2 to show the values of ΔU, q and w for a time of 40 s at this higher constant temperature. [3]

Table 2.2

ΔU/J	q/J	w/J

Tutorial 13B: DC Circuits

Self-Review Questions

S1. **J02/I/34**

When four identical lamps P, Q, R and S are connected as shown in diagram 1, they have normal brightness.

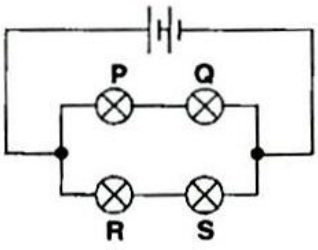


diagram 1

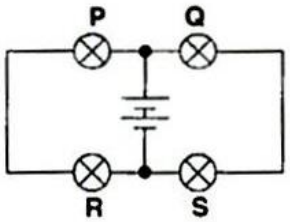


diagram 2

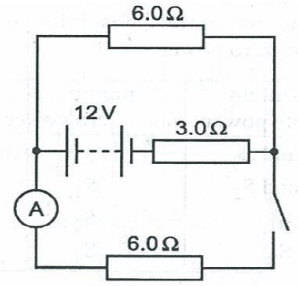
When the four lamps are connected as shown in diagram 2, which statement is correct?

A The lamps do not light.
B The lamps are less bright than normal.
C The lamps have normal brightness.
D The lamps are brighter than normal.

S2. **N07/I/28**

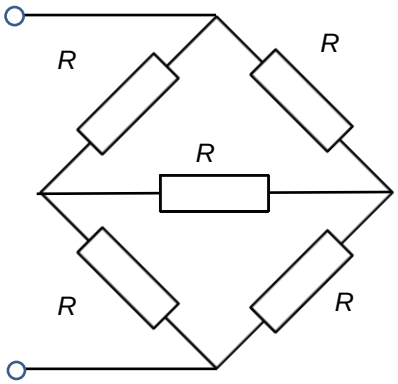
In the circuit, the battery has an e.m.f. of 12 V and an internal resistance of 3.0 Ω. The ammeter has negligible resistance.

The switch is closed. What is the reading on the ammeter?



A 0.50 A **B** 1.0 A **C** 1.3 A D 2.0 A

S3. Five identical 4 Ω resistors are connected as shown below.
Find the equivalent resistance.



S4. **J02/II/8**

A student has available some resistors, each of resistance 100 Ω.

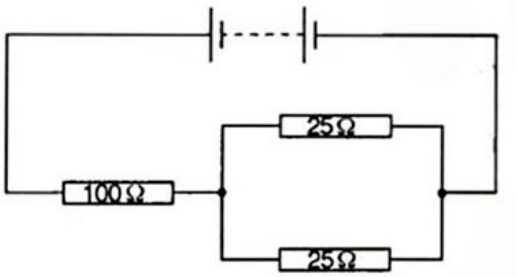
(a) Draw circuit diagrams, one in each case, to show how a number of these resistors may be connected to produce a combined resistance of

(i) 200 Ω ,

(ii) 50 Ω,

(iii) 40 Ω.

(b) The arrangement of resistors shown below is connected to a battery.



The power dissipation in the 100 Ω resistor is 0.81 W. Calculate

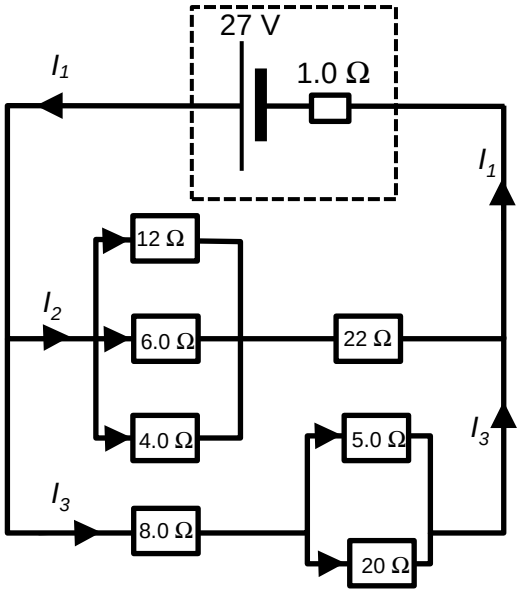
(i) the current in the circuit and the current in each branch,

(ii) the power dissipation in each of the 25 Ω resistors.

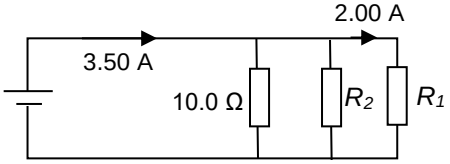
S5. The 27 V battery for the circuit below has an internal resistance of 1.0 Ω. Find

(a) the currents I_1, I_2 and I_3,

(b) the current passing through the 12 Ω resistor.



S6. In the circuit shown, the rate at which R_1 is dissipating electrical energy is 20.0 W.



(a) Find R_1 and R_2.

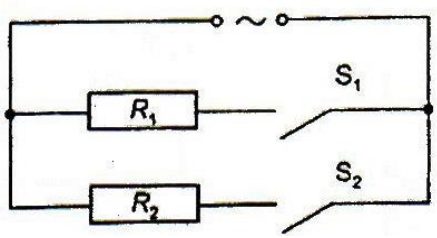
(b) What is the e.m.f. of the battery?

(c) Find the current through both R_2 and the 10.0 Ω resistor.

(d) Calculate the total electrical power consumption in all the resistors and the electrical power delivered by the battery. Show that your results are consistent with conservation of energy.

S7. **N05/I/24**

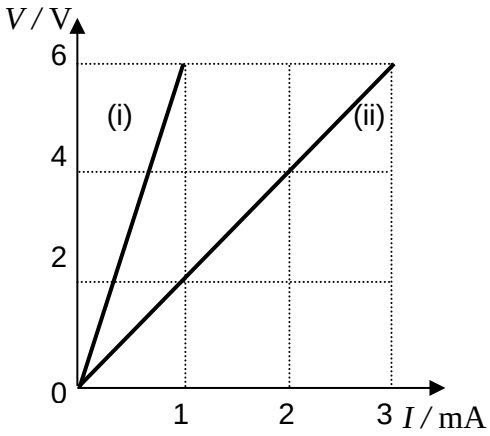
An electric heater can be represented as two resistors of resistances R_1 and R_2 and two switches S_1 and S_2. The resistance R_2 is greater than that of R_1.



Which switches must be closed so that the heater produces the maximum possible power and the minimum non-zero power?

	maximum possible power	minimum non-zero power
A	S_1 and S_2	S_2
B	S_1 and S_2	S_1
C	S_1	S_2
D	S_2	S_1

S8. In the figure on the right, graph (i) shows the V–I characteristics of a resistor P. P is then arranged in a particular combination with a resistor Q to give graph (ii).

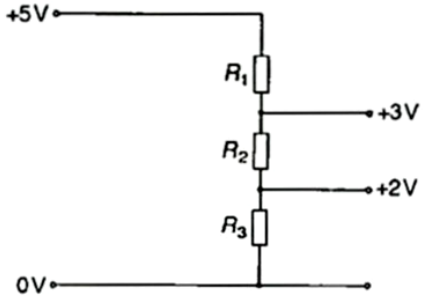


From the graph, it can be deduced that Q has a resistance of

A 2 kΩ and is in series with P.
B 2 kΩ and is in parallel with P.
C 3 kΩ and is in parallel with P.
D 4 kΩ and is in series with P.

S9 **J02/I/35**

A potential divider is used to give outputs of 2 V and 3 V from a 5 V source, as shown.

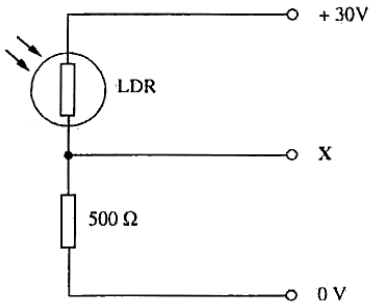


What are possible values for the resistances R_1, R_2 and R_3?

	R_1/kΩ	R_2/kΩ	R_3/kΩ
A	2	1	5
B	3	2	2
C	4	2	4
D	4	6	10

S10 **N91/I/15**

A light dependent resistor (LDR) and a 500 Ω resistor from a potential divider between voltage lines held at +30 V and 0 V as shown in the diagram.

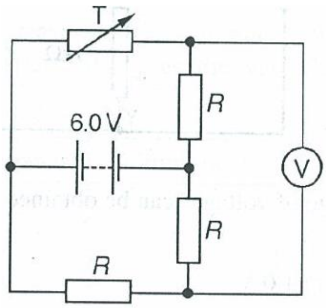


The resistance of the LDR is 1000 Ω in the dark but then drops to 100 Ω in bright light. What is the corresponding change in the potential at X?

A a rise of 25 V
B a rise of 15 V
C a fall of 10 V
D a fall of 15 V
E a fall of 25 V

Discussion Questions

D1. A battery of emf 6.0 V and negligible internal resistance is connected to three resistors, each of resistance *R*, and a variable resistor T, as shown.

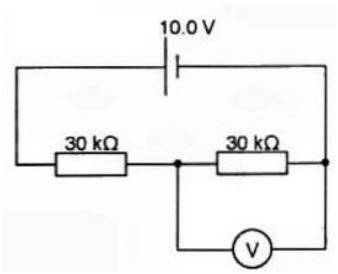


The resistance of T changes from *R* to 5*R*. What is the change in the magnitude of the reading of the high resistance voltmeter?

A zero **B** 2 V **C** 4 V **D** 5 V

D2. **N02/I/23**

In the circuit, a 10.0 V supply of negligible internal resistance is joined to two 30 kΩ resistors in series. A voltmeter of resistance 60 kΩ is connected in parallel with one of the resistors.

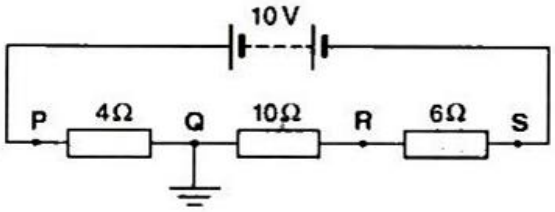


What is the reading on the voltmeter?

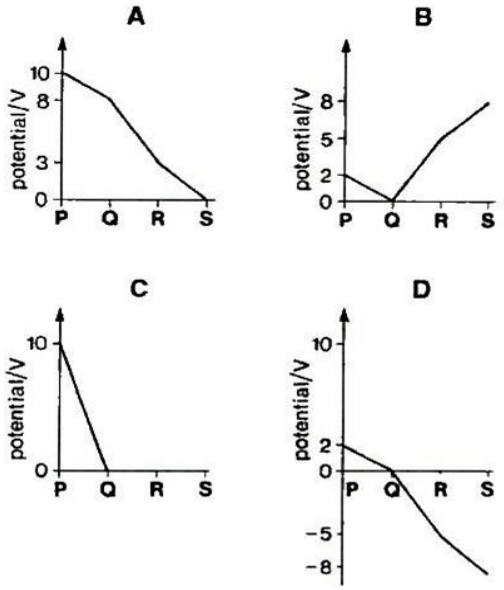
A 4.0 V
B 5.0 V
C 6.0 V
D 6.7 V

D3. **J95/I/15**

The diagram shows three resistors of resistances 4 Ω, 10 Ω and 6 Ω connected in series. A potential difference of 10 V is maintained across them, with point Q being earthed.

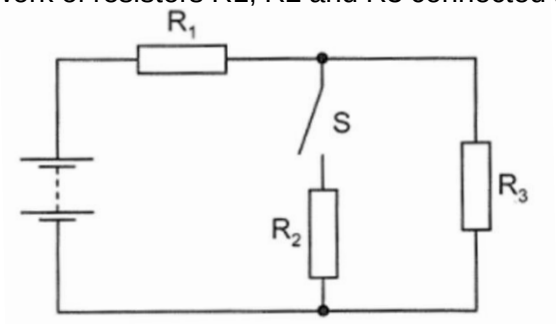


Which graph represents the change in potential along the resistor network



D4. **N11/I/26**

The diagram shows a network of resistors R1, R2 and R3 connected to a battery of negligible internal resistance.



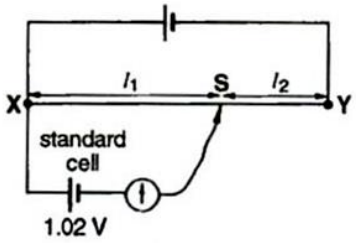
When the switch S is closed, the potential difference (p.d.) across R2 (originally zero) rapidly increases to a steady value.

What happens to the potential difference (p.d.) across each of the other two resistors, and to the power output of the battery?

	p.d. across R_1	p.d. across R_3	battery power output
A	decreases	decreases	decreases
B	decreases	stays the same	decreases
C	increases	decreases	increases
D	increases	stays the same	increases

D5. **N88/I/17**

A standard cell of emf 1.02 V is used to find the potential difference across the wire XY as shown in the diagram. It is found that there is no current in the galvanometer when the sliding contact is at S, l_1 from X and l_2 from Y.

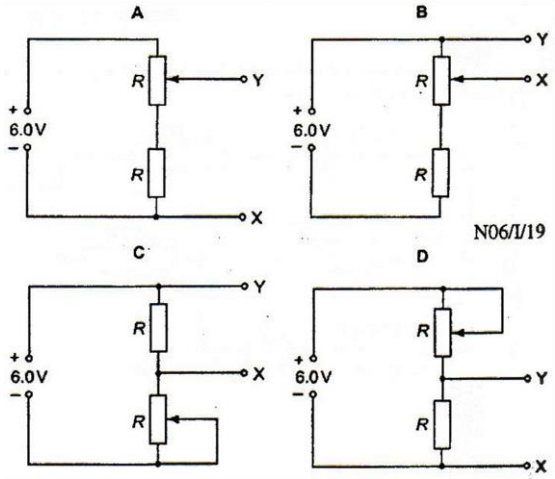


What is the potential difference across XY?

A 1.02 V **B** $1.02\left(\frac{l_2}{l_1}\right)$ V **C** $1.02\left(\frac{l_1}{l_2}\right)$ V **D** $1.02\left(\frac{l_1+l_2}{l_2}\right)$ V **E** $1.02\left(\frac{l_1+l_2}{l_1}\right)$ V

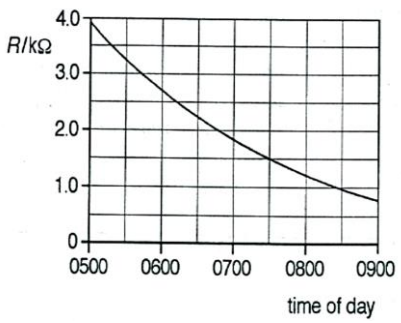
D6. **N06/I/19**

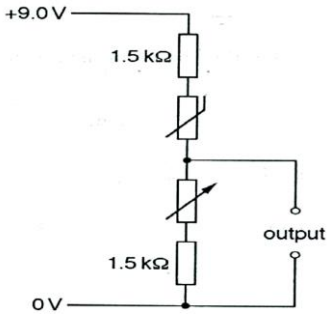
A potential divider has a constant supply of 6.0 V as shown in the diagrams.
Which circuit will provide a potential difference between X and Y that can be varied between zero and 3.0 V?



D7 **N08/I/28**

The graph shows how the resistance *R* of a thermistor varies during a part of a day. The thermistor is connected in the potential divider circuit shown.



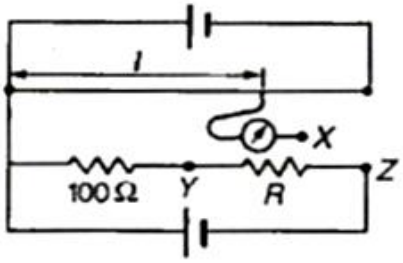


In order to obtain an output of 6.0 V at time 0730, to which value should the variable resistor be set?

A 1.5 kΩ **B** 3.0 kΩ **C** 4.5 kΩ **D** 6.0 kΩ

D8. **J81/II/23**

The diagram shows a circuit which may be used to compare the resistance *R* of an unknown resistor with a 100 Ω standard. The distances *l* from one end of the potentiometer slide-wire to the balance point are 400 mm and 588 mm when X is connected to Y and Z respectively. The length of the slide-wire is 1.00 m.

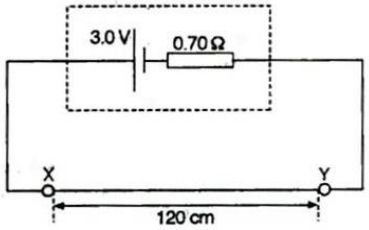


What is the value of resistance R?

A 32 Ω **B** 47 Ω **C** 68 Ω **D** 147 Ω **E** 150 Ω

D9. **N02/II/6**

(a) A uniform wire XY of length 120 cm and radius 0.55 mm is connected in series with a cell of e.m.f. 3.0 V and internal resistance 0.70 Ω, as shown in the figure below.

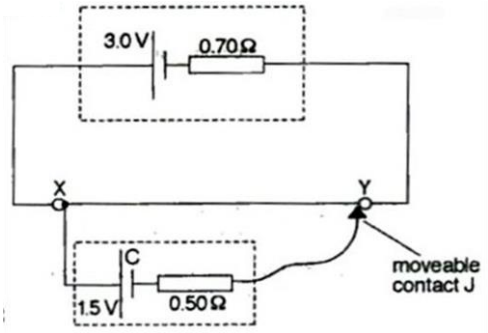


(i) The resistivity of the material of the wire is 1.1×10^{-6} Ω m. Show that the resistance of the wire is 1.4 Ω.

(ii) Calculate the potential difference (p.d.) per unit length of XY.

[5]

(b) A cell C of e.m.f. of 1.5 V and internal resistance 0.50 Ω is connected to the circuit as shown.



The moveable contact J can be connected to any point along the wire XY.

(i) Initially, the contact J is connected to end Y. The p.d. across the wire XY is then equal to the p.d. applied across cell C. On the above diagram, mark with an arrow the direction of the current through the cell C.

(ii) Determine the position of the contact J on the wire XY such that there is no current through the cell C.

(iii) Suggest one way in which the circuit in the above figure may be modified so that the position found in (ii) is nearer to end Y.

[5]

D10. **N12/II/3**

The variation with potential difference V of current I for a semiconductor diode is shown in Fig. 3.1.

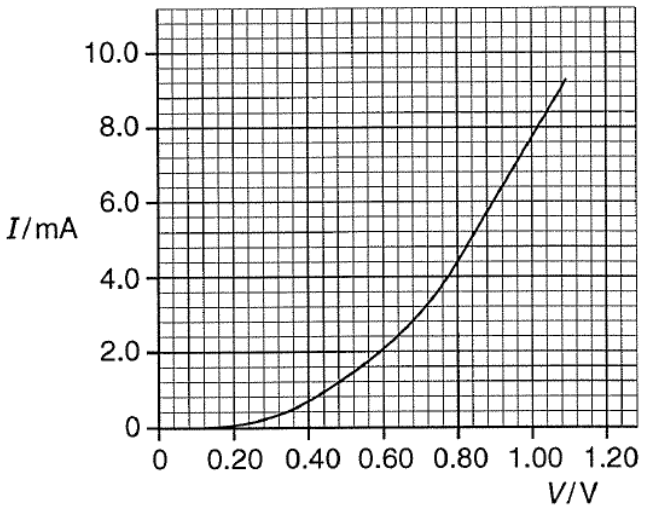


Fig. 3.1

(a) (i) Describe the main features of Fig. 3.1 that show the characteristics of the diode in terms of current, voltage and resistance. [4]

(ii) Use Fig. 3.1 to determine the resistance of the diode at 0.80 V. [2]

(b) In an attempt to obtain the graph of Fig. 3.1 for the diode, a student sets up a circuit as shown in

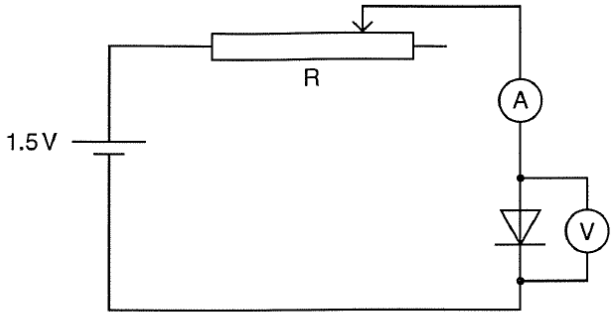


Fig. 3.2

The cell has an e.m.f. of 1.5 V and negligible internal resistance. The resistor R has a resistance that can be varied between zero and 100 Ω.
State and explain why the circuit shown in Fig. 3.2 is inappropriate for determining the graph of Fig. 3.1. [2]

(c) (i) Draw a circuit diagram, using the same components as shown in Fig. 3.2, from which the graph of Fig. 3.1 may be determined. [1]

(ii) Explain why this circuit is appropriate for determining the graph of Fig. 3.1. [1]

Suggested Solution (COE):

S2. 0.125×10^{-3} m s^{-1}
S6. 54 Ω
S8. (i) 13.0 A (ii) 17.6 Ω (iii) 1.50×10^{7} J
D2. (b) 7.37×10^{-4} m s^{-1}
D7. 2 mA
D9 1.68 kW
D11. (i) 1. 4.0 Ω 2. 36 W (ii) 2.5 Ω
D12. (a) (i) 4.2 Ω (ii) 2.95 Ω (b) (i) 4.25 V (ii) 2.18 A (iii) 5.99 V (iv) 6760 J
D13. (d) (i) 1 Ω, 9 Ω (ii) 25%, 75% (iii) 50%.
D15 (a) 1.41×10^{-6} Ω m

Suggested Solutions (DC):

S1. C
S2. B
S3. 4 Ω
S4. (b)(i) 0.090 A (ii) 0.051 W
S5. (a) I_1 = 3.0 A; I_2 = 1.0 A; I_3 = 2.0 A (b) 0.17 A
S6. (a) 5.00 Ω, 20.0 Ω (b)10.0V (c) 0.50 A, 1.0 A (c) 35.0 W
S7. A
S8. C
S9. C
S10. B

D9. (a)(i) 1.4 Ω (ii) 1.67 V m^{-1} (b)(i) to the right (ii) 0.90 m
D10. (a)(ii) 180 Ω

Chapter 14

Electromagnetism

This picture shows Shanghai Maglev, the world's first high-speed commercial magnetic levitation line, whose trains run on non-conventional tracks and reach a top speed of 430 km/h (267 mph). In comparison, conventional railway speed is below 100 km/h, and commercial aircraft cruise at around 800 km/h.

Maglev (derived from magnetic levitation) is a system of train transportation that uses two sets of magnets: one set to repel and push the train up off the track, and another set to move the elevated train ahead, taking advantage of the lack of friction. Along certain "medium-range" routes (usually 320 to 640 km), maglev can compete favourably with high-speed rail and airplanes.

With maglev technology, there is just one moving part: the train itself. The train travels along a guideway of magnets which control the train's stability and speed. Propulsion and levitation require no moving parts. This is in stark contrast to electric multiple units that may have several dozen parts per bogie. Maglev trains are therefore quieter and smoother than conventional trains and have the potential for much higher speeds.

(adapted from https://en.wikipedia.org/wiki/Maglev)

H2 Physics Syllabus 9749

Content

- Concept of a magnetic field
- Magnetic fields due to currents
- Force on a current-carrying conductor
- Force between current-carrying conductors
- Force on a moving charge

Learning Outcomes

Candidates should be able to:

(a) show an understanding that a magnetic field is an example of a field of force produced either by current-carrying conductors or by permanent magnets

(b) sketch flux patterns due to currents in a long straight wire, a flat circular coil and a long solenoid

(c) use $B = \mu_0 I / 2\pi d$, $B = \mu_0 N I / 2r$ and $B = \mu_0 n I$ for the flux densities of the fields due to currents in a long straight wire, a flat circular coil and a long solenoid respectively

(d) show an understanding that the magnetic field due to a solenoid may be influenced by the presence of a ferrous core

(e) show an understanding that a current-carrying conductor placed in a magnetic field might experience a force

(f) recall and solve problems using the equation $F = BIl \sin\theta$, with directions as interpreted by Fleming's left-hand rule

(g) define magnetic flux density

(h) show an understanding of how the force on a current-carrying conductor can be used to measure the flux density of a magnetic field using a current balance

(i) explain the forces between current-carrying conductors and predict the direction of the forces

(j) predict the direction of the force on a charge moving in a magnetic field

(k) recall and solve problems using the equation $F = BQv \sin\theta$

(l) describe and analyse deflections of beams of charged particles by uniform electric and uniform magnetic fields

(m) explain how electric and magnetic fields can be used in velocity selection for charged particles.

Table of Contents

14 Introduction 4

14.1 The Magnetic Field and Magnetic Flux Density 4
14.2 Currents can generate Magnetic Fields 5
14.3 Currents Experiencing a Force when in a Magnetic Field 12
14.4 Defining the Magnetic Flux Density and the Tesla 15
14.5 Forces Between 2 Long Parallel Current-Carrying Conductors 16
14.6 Revisiting the DC Motor 18

Tutorial 14A: Magnetic Field and Magnetic Force 19

14.7 Magnetic Force on a Moving Electric Charge 28
14.8 Motion of a Charged Particle in a Uniform Magnetic Field 29

Tutorial 14B: Moving Charges in Magnetic Field 32

14.9 Moving Charges in Uniform Electric and Magnetic fields 36
14.10 Velocity Selector 37

Tutorial 14C: Moving Charges in Electric and Magnetic Fields 38

APPENDIX A: Earth as a Magnet 43

APPENDIX B: The Cyclotron 44

APPENDIX C: Charges Moving in Non-Uniform Magnetic Fields 45

List of symbols

The following are the main symbols used in diagrams:

⊗ Into the page

⊙ Out of the page

Playlist of Worked Examples can be found at: https://for.edu.sg/82qoe9

14 Introduction

More than 2000 years ago, the Chinese discovered magnetic phenomena and they have since been using them for navigation. However, it was only in the 1700s that magnetic properties and phenomena were studied in greater depth. Major discoveries made in the last 300 years have made a significant impact on our lives today. Scientists who have contributed significantly to our understanding of magnetic phenomena include Ampère, Oersted, Gauss, Henry and Faraday.

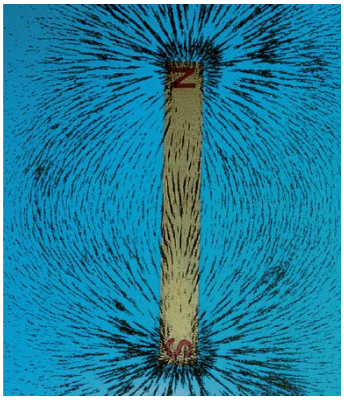
Magnetism has many useful applications in our everyday lives. Earth's magnetic field not only allows us to find our way around, but also enables birds to navigate when they migrate South during winter in the North. Nuclear Magnetic Resonance (NMR), which makes use of atoms' magnetic moments, allows us to capture images of our body. Mobile phones, speakers and microphones make use of the concept of electromagnetic induction. The Great Industrial Revolution would not have been possible without knowing that the cutting of magnetic flux can generate electricity.

In this topic, you will learn that magnetic fields can be generated by electric currents and how magnetic forces act on moving charges as well as current-carrying conductors. In the next topic on Electromagnetic Induction (EMI), you will learn how electric currents can be created by changing magnetic fields.

14.1 The Magnetic Field and Magnetic Flux Density

Just like the gravitational and electric fields, the magnetic field is an example of a '*field of force*'. The concept of a field illustrates '*action at a distance*', which means that a suitable object placed in the field can experience a force. A magnetic field is a region of space where a magnetic material or a moving charge/current-carrying conductor experiences a magnetic force.

As with electric fields, we can illustrate magnetic fields by drawing magnetic field lines. The pictorial representation of magnetic field lines is very useful in visualizing the strength and direction of the magnetic field. Field lines can be visualized quite easily in the real world. This is commonly done with iron filings scattered on a surface around a magnet. Each iron filling is magnetized and acts like a tiny compass needle with north and south poles. Thus, the filings rearrange themselves along the magnetic field lines.



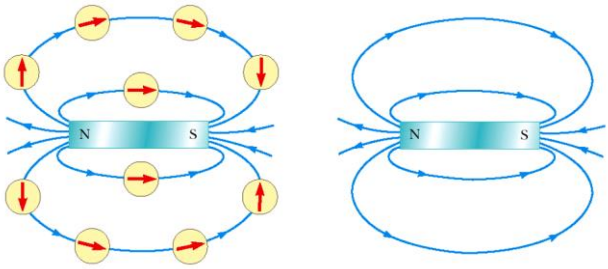


Figure 14.1.1 Plotting of magnetic field lines of a magnet using a compass.

Unlike gravitational field and electric field, we associate the direction of magnetic field lines with the direction of the north pole on a compass, rather than the direction of magnetic force.

- Magnetic field lines 'appear' to originate from the north pole and end on the south pole of a magnet. But to be exact, **magnetic field lines have no start or end point**. The field lines actually continue on within the magnet to form a complete loop.
- The *direction of a magnetic field at any point* is indicated by the **tangent to the magnetic field line** at that point. The direction of the magnetic field at any point in space is the direction indicated by the north pole of a small compass needle placed at that point.
- Magnetic field lines **do not cross each other**.
- The **strength of the magnetic field is indicated by the density of the field lines**; the magnetic field is stronger in regions where the field lines are relatively close together and weaker where they are relatively far apart.

We use the **direction of the field lines at a point to define the direction of a vector quantity called magnetic flux density *B*** (it will be defined later on pg 15.)

Furthermore, the **magnitude of this quantity will indicate the strength of the magnetic field**. Nearer to the source, the magnetic field is stronger and thus ***B*** will have a greater magnitude.

The SI unit for magnetic flux density is the **tesla (T).**

The magnetic field is sometimes called the *B*-field.

14.2 Currents can generate Magnetic Fields

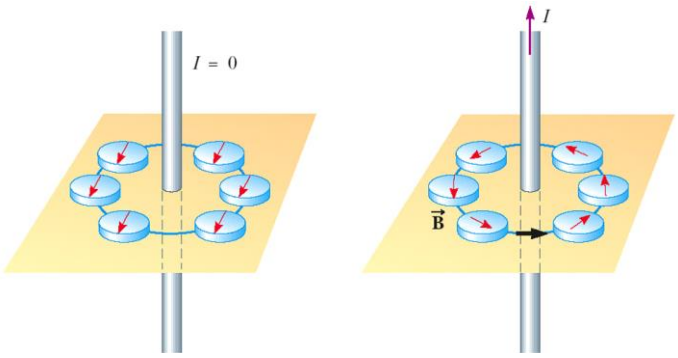
Before the 1800s, magnetism and electricity were considered to be two independent topics. Although there were a number of scientists who thought the two topics were related, they were unsuccessful in finding the link. It was only in 1820 that a Danish physicist, Hans Christian Oersted, discovered the link. He observed that a compass needle swung almost perpendicular to a current-carrying conductor as if attracted by a powerful magnet. This discovery prompted Ampère to work out the mathematical relationship between currents and magnetic fields known as Ampère's force law (not required in the 'A' level syllabus). It also inspired Faraday's work on electromagnetic induction. Oersted's discovery, which linked the motion of electric charges with the creation of a magnetic field, marked the beginning of an important discipline called electromagnetism.

Two important observations were made about the **magnetic field due to current-carrying conductors**:

1. The direction of the magnetic field is always *perpendicular* to the direction of the current.

2. The strength of the magnetic field produced is dependent on the magnitude of the current.

Previously, we detected the presence of a magnetic field of a magnet by observing how the iron filings align themselves when sprinkled around the magnet. A similar effect is observed when we sprinkle iron filings around a current-carrying wire.



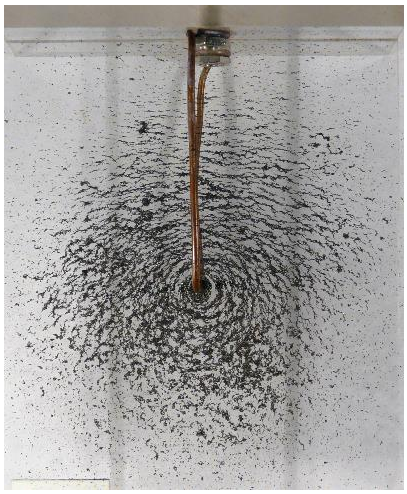
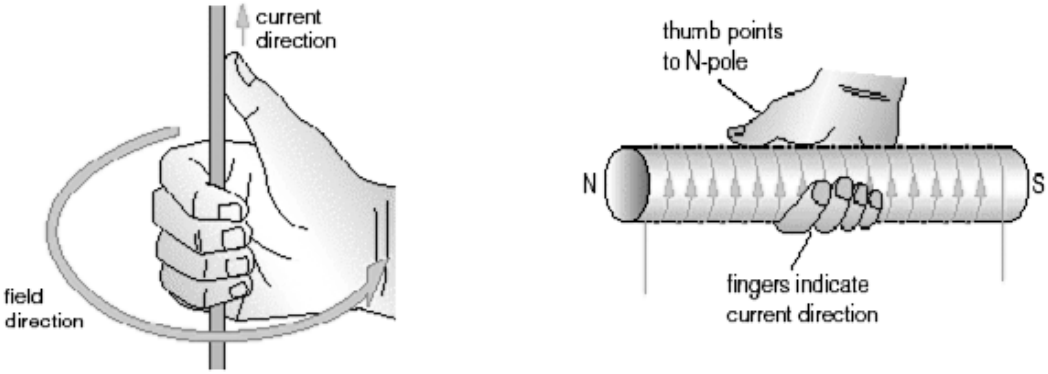


Figure 14.2.1 Compasses deflected when the current is on.

The direction of the magnetic field lines can be easily determined using the **Right-Hand Grip Rule**.



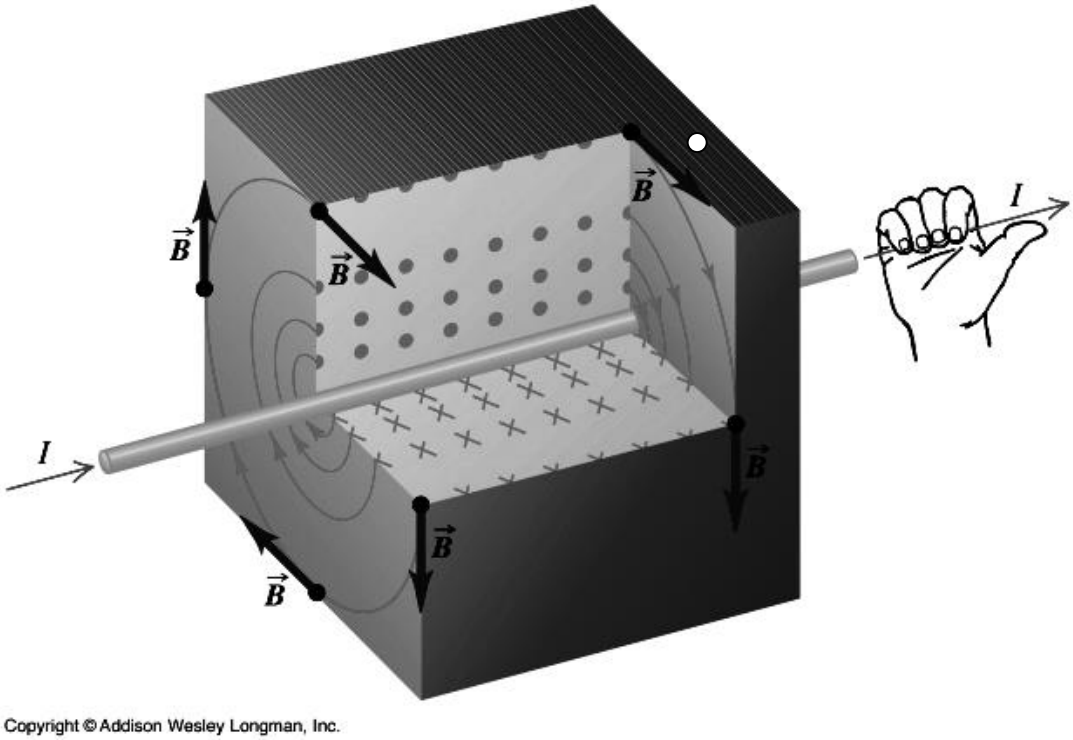
(a) Current-carrying conductor/wire (b) current carrying coil/solenoid.

Figure 14.2.2 Using the right-hand grip rule to determine the direction of the magnetic field due to a current-carrying conductor.

14.2.1 Magnetic-Field Patterns

14.2.1.1 Long Straight Current-Carrying Conductor

Figure 14.2.3 shows the magnetic-field pattern produced by a long straight current-carrying conductor.

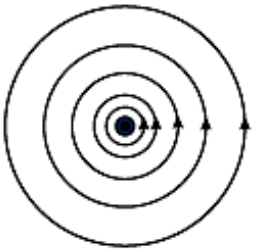


Copyright © Addison Wesley Longman, Inc.

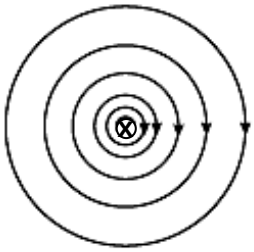
Figure 14.2.3 Magnetic-field pattern due to a long straight current-carrying conductor.

Helpful Tips

1. The magnetic field lines form concentric circles around the straight wire. Given a choice, *always draw the plan view as shown below.*



A circle with a DOT shows that the current is coming OUT from the plane



A circle with a CROSS shows that the current is moving INTO the plane

2. When drawing the lines, it is extremely important to show the *field lines spreading further and further apart as d increases*. This is because the B-field is getting weaker as we move away from the source (i.e. the wire).

The magnetic flux density, B, at a distance, d, from a long straight wire carrying a current, I, is given by

$$B = \frac{\mu_o I}{2\pi d}$$

where μ_o is the ***permeability of free space***, which has the value $4\pi \times 10^{-7}$ H m^{-1}.

Thus, $\boldsymbol{B}$ **is *directly proportional* to** $\boldsymbol{I}$ but is ***inversely proportional*** **to the perpendicular distance** $\boldsymbol{d}$ from the wire.

14.2.1.2 Flat Circular Coil

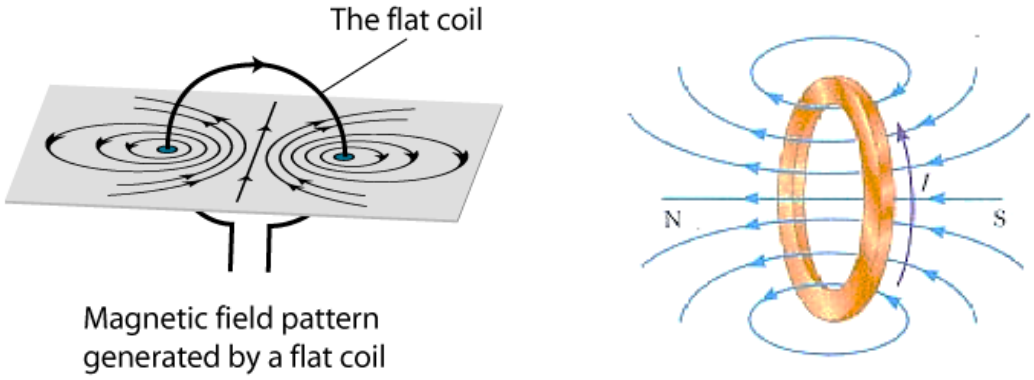
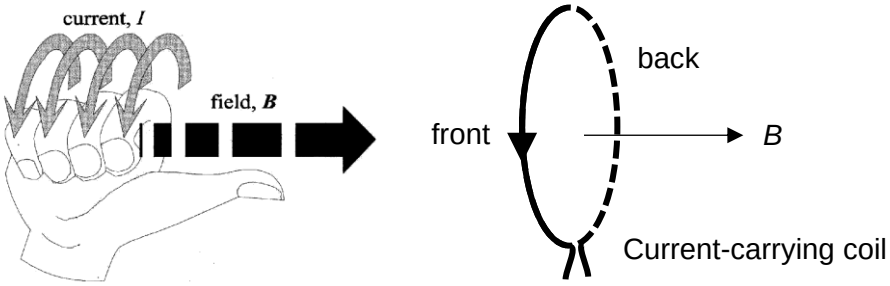


Figure 14.2.4 Magnetic-field pattern of current flowing in circular loop.

The right-hand grip rule can still be used to determine the direction of the magnetic field lines produced by a current in circular coils. In general, there are 2 different methods that can be used:

1. For the straight-current scenario, the thumb represents the current and the coiling fingers representing the B-field.

2. For "circular" currents, we can exchange the fingers and the thumb to let the coiling fingers represent the direction of the currents in the circular coils and the thumb the direction of the magnetic-field lines passing through the cross-sectional area of the circular coils.



1. Using the thumb to represent the B-field also allows us to determine the "north" pole of the circular coil, since B-field lines originate from the "north" pole of the magnet. This will be very useful later.

2. It is sometimes difficult to tell from the diagram which part represents the "front" of the coils. Standard conventions include using a thicker / bolder line for the "front" and a dotted line for the "back" of the coil.

The magnetic flux density, B, at the ***centre*** *of a flat circular coil* of N turns, radius, r, and carrying current, I, is given by

$$B = \frac{\mu_o NI}{2r}$$

where μ_o is the **permeability of free space** which has the value $4\pi \times 10^{-7}$ H m^{-1}.

Again, we observe the same relationship as the magnetic field due to a long straight current-carrying conductor, that is, $\boldsymbol{B}$ **is** ***directly proportional*** **to** $\boldsymbol{I}$ but is ***inversely proportional*** **to the radius of the circular coil,** $\boldsymbol{r}$**.**

14.2.1.3 Long Solenoid

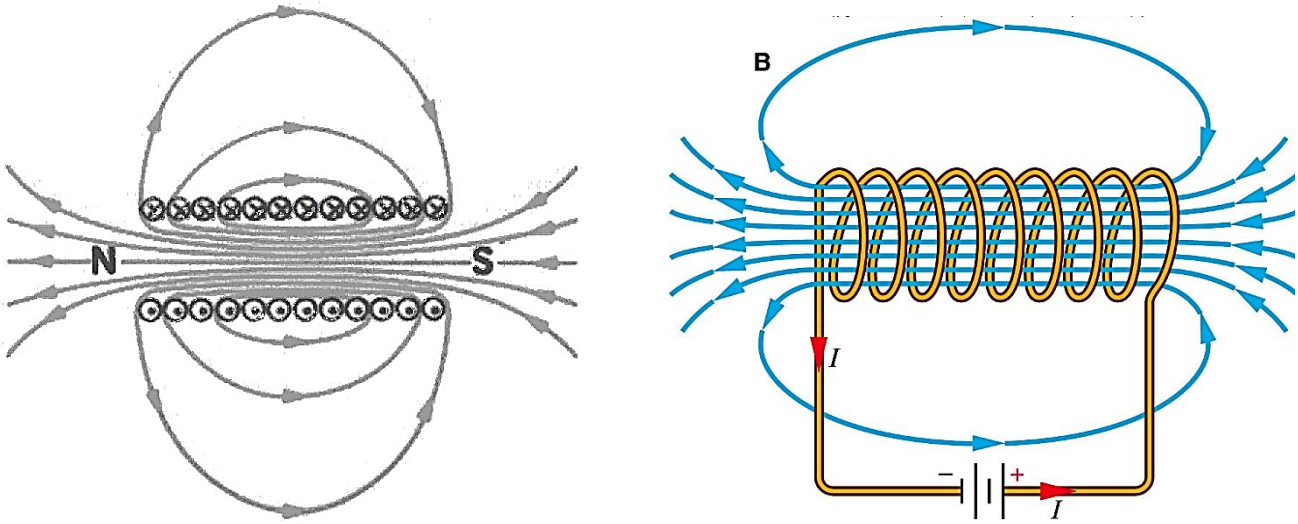


Figure 14.2.5 Magnetic-field pattern of a current-carrying solenoid.

1. The magnetic field produced by a long current-carrying solenoid is **uniform** for the length of the solenoid except for the ends of the solenoid. This is represented by the **parallel, equally-spaced lines along the axis of the solenoid.**

2. The solenoid also has a "north" and a "south" pole. In fact, its field pattern is very similar to that of a bar magnet.

3. The magnetic flux density **at the ends is half the magnetic flux density at the center** of the solenoid.

The magnetic flux density, B, at the ***centre*** *of a long solenoid* with n turns per unit length and carrying current, I, is given by

$$B = \mu_o n I$$

where μ_o is the **permeability of free space** which has the value $4\pi \times 10^{-7}$ H m^{-1}.

Here, $\boldsymbol{B}$ **is** ***directly proportional*** **to** $\boldsymbol{I}$ and ***n*** **(the number of turns per unit length)**.
Note that *n* (lower-case *n*) is the *number of turns per unit length*, whereas *N* (capital *N*) is the *total number of turns.*

For the A-level examination, the above formulae (magnetic flux density due to a long straight wire, a flat circular coil and a long solenoid) are provided in the formulae list and you are expected to know what the symbols mean.

Effect of a ferrous core

One way of strengthening the magnetic field of a solenoid is to insert a piece of soft iron (ferrous core) into the solenoid. The resultant magnetic flux density of the solenoid will be enhanced.

The magnetic domains in the soft iron cause the magnetic field to strengthen. Inserting the soft iron causes the permeability of free space, μ_0, to change to permeability of iron, μ. Materials that strengthen magnetic field in this manner are described as ***ferromagnetic***.

Examples of ferromagnetic materials include soft iron, cobalt and nickel. They can all be magnetized.

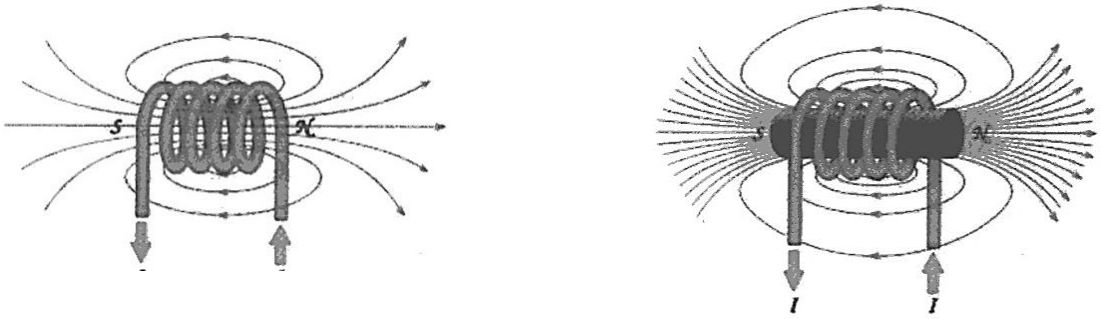
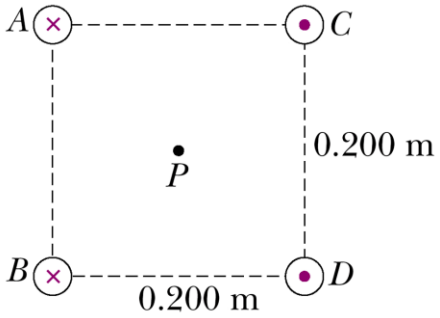


Figure 14.2.6 Increasing field lines due to a ferrous core

	Flux density at the centre	Flux density at either pole
Solenoid	$B = \mu_o nI$	$B = \frac{1}{2}\mu_o nI$
Solenoid with ferrous core	$B = \mu nI$	$B = \frac{1}{2}\mu nI$

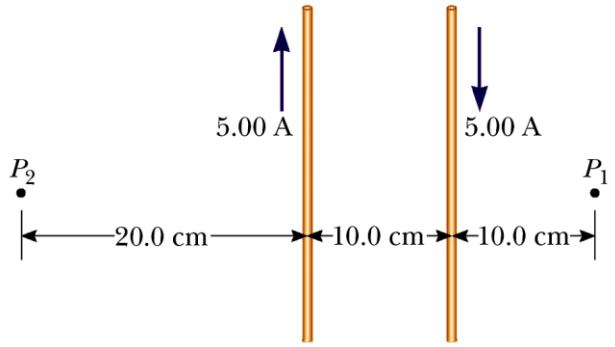
Worked Example 1

Determine the direction of the magnetic field at the center of the square assuming all currents are of the same value.



Worked Example 2

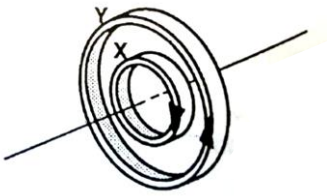
(a) Determine the direction of the magnetic field at points P_1 and P_2.



(b) Calculate the magnitude of the resultant magnetic flux density at P_1.

Worked Example 3 (J93/I/8)

Two coils X and Y, each with 100 turns, are arranged as shown in the diagram.



X has a radius 0.050 m and carries a current of 3.0 A. Y has radius 0.10 m and carries a current of 6.0 A in the opposite direction to X.

What is the magnitude of the total magnetic flux density at the centre of the coils?

14.3 Currents Experiencing a Force when in a Magnetic Field

Since currents can generate a magnetic field and magnetized conductors behave like magnets, maybe a current-carrying wire will experience a magnetic force when placed in a magnetic field too. Figure 14.3.1 shows the results of such an experiment.

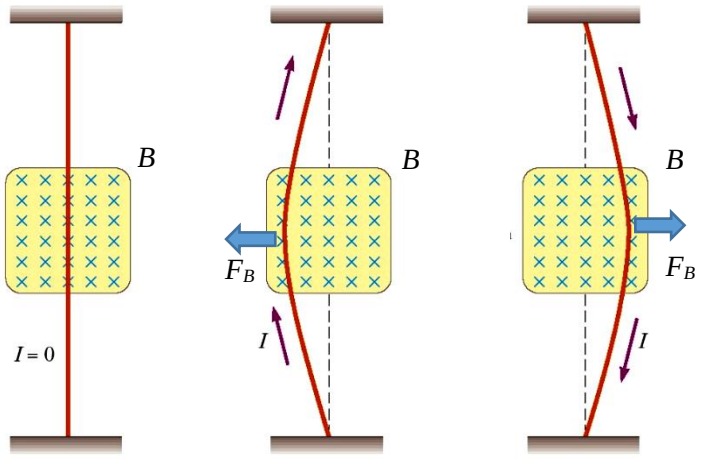


Figure 14.3.1 Each sketch shows the direction of the force exerted on a current-carrying wire placed in a magnetic field

Observations from this experiment show that there is a relationship between the magnetic flux density of the magnetic field present and the force experienced by the current-carrying wire.

Maximum force is experienced by the wire when the magnetic flux density $\boldsymbol{B}$ and the wire are **perpendicular to each other**. The direction of the force that results is perpendicular to both $\boldsymbol{B}$ and the wire. Quantitatively, in this orientation, we have

$$F_B = BIL$$

where F_B is the force exerted on the wire, I is the current in the wire and L is the length of the wire.

The relative directions are easily determined using **Fleming's Left-Hand Rule**,

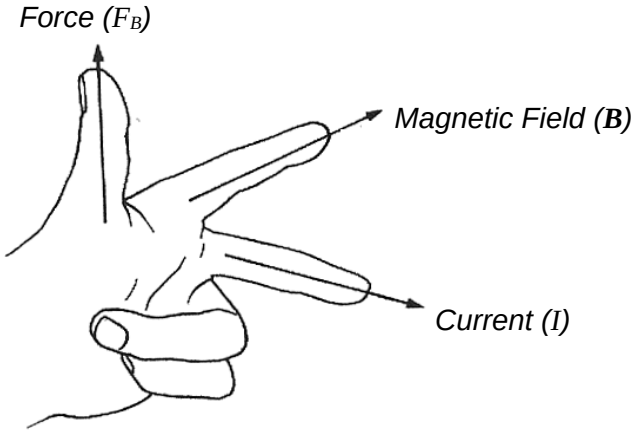


Figure 14.3.2 Fleming's Left-Hand Rule.

14.3.1 Calculating the Force on a Current in an External Magnetic Field

Fleming's Left-Hand Rule allows you to determine the direction of the magnetic force $\boldsymbol{F_B}$ on a current-carrying conductor placed perpendicular to a magnetic flux density $\boldsymbol{B}$. However, in general, the angle between $\boldsymbol{B}$ and $\boldsymbol{I}$ can be any angle θ. The magnetic force $\boldsymbol{F_B}$ experienced by the conductor is ***always*** **perpendicular** to **the plane containing $\boldsymbol{B}$ and $\boldsymbol{I}$**. Fleming's Left-Hand Rule can still be applied in such instances to determine the orientation of F_B.

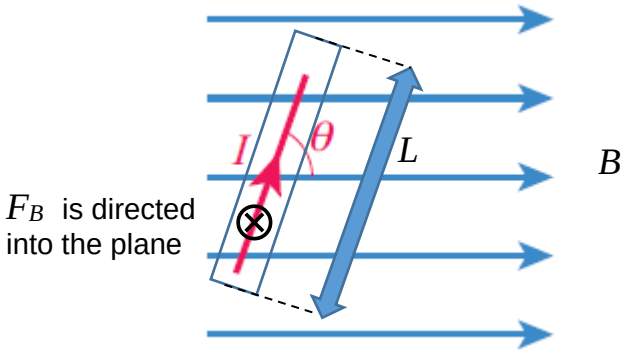
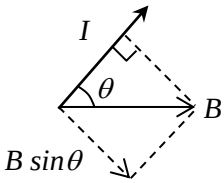


Figure 14.3.3 Magnetic force F_B on a current-carrying conductor

In Figure 14.3.3, the component of the magnetic flux density perpendicular to the conductor is given by

$$B_\perp = B\sin\theta$$



where θ is the angle between B and I.

Hence, the magnetic force on the straight conductor is given by

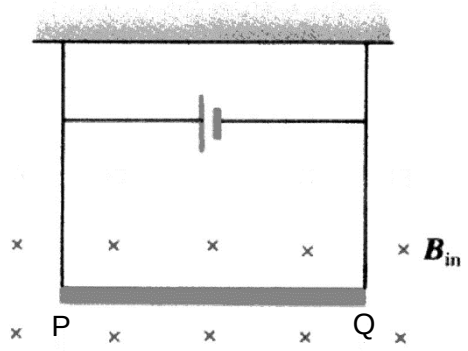
$$\boxed{F_B = B_\perp IL}$$

where L is the length of the conductor carrying current, I, in the magnetic field, $B_\perp$ is the magnetic flux density perpendicular to the current carrying conductor.

By Fleming's Left-Hand Rule, the magnetic force is directed into the page.

Worked Example 4

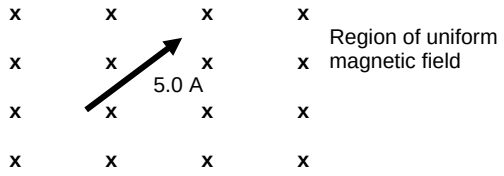
A conductor PQ suspended by two light flexible wires as shown in the figure below has a mass per unit length of 0.040 kg m^{-1}. What current must exist in the conductor in order for the tension in the supporting wires to be zero if the magnetic field in the region is 3.6 T into the page?



Worked Example 5 [J93/I/19]

A wire of length 3.0 cm is placed at right angles to a magnetic field of flux density 0.040 T. The wire carries a current of 5.0 A. What is the magnitude of the force which the field exerts on the wire?

A less than 0.006 N
B 0.006 N
C greater than 0.006 N but less than 0.6 N
D 0.60 N
E greater than 0.60 N



14.4 Defining the Magnetic Flux Density and the Tesla

We may now define quantitatively the magnetic flux density, $\boldsymbol{B}$.

Definition of Magnetic Flux Density, $\boldsymbol{B}$:

The **magnetic flux density** is defined as the **force per unit length per unit current** acting on a **straight** current-carrying conductor placed **perpendicular to a magnetic field**.

Definition of Tesla (T):

The magnetic flux density of a magnetic field is 1 tesla if the force per unit length per unit current which acts on a straight current-carrying conductor placed perpendicular to the magnetic field is 1 newton per metre per ampere.

Magnetic flux density, $\boldsymbol{B}$, is a **vector** quantity and the **S.I. unit is the tesla (T).**

For different shapes of current-carrying wires, the magnetic fields around the wires will in general be different. There are ways to measure experimentally the value of $\boldsymbol{B}$, which generally depends on the current that is generating the magnetic field as well as the distance from the wire.

14.4.1 Measuring Magnetic Flux Density using a Current Balance

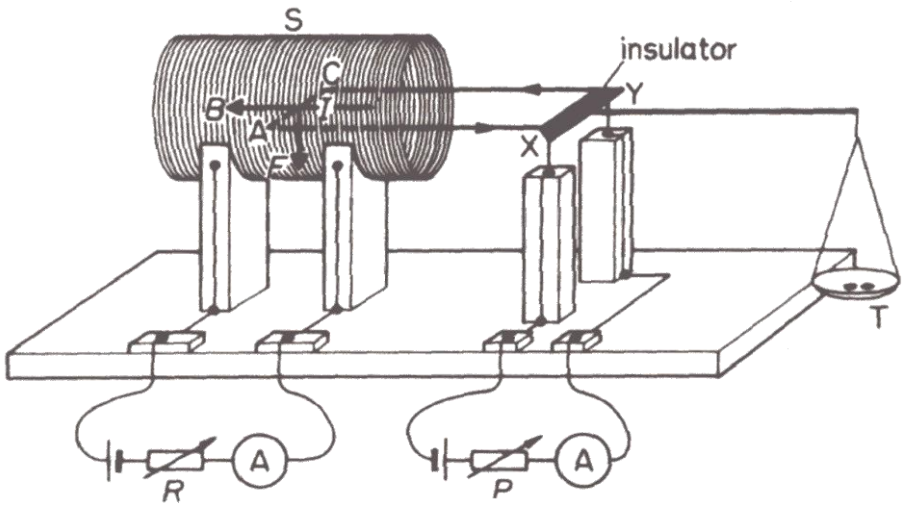
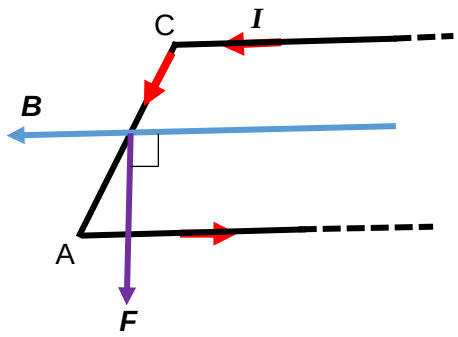


Figure 14.4.1 Current balance.

Frame inside the coil



Note: magnetic force F is perpendicular to the plane that contains both the magnetic field B and the current I.

How a current balance works:

1. A wire frame AXYC is balanced on 2 knife edges through which a current I can pass from a battery. The frame is arranged such that the side AC of the frame (of length L) lies within the magnetic field whose flux density B is to be determined.

2. When there is no current, the frame is horizontal, as the force acting on AC is zero.
   When current flows through, a magnetic force acts on AC and pushes that side of the frame downwards (by Fleming's Left-Hand Rule).

3. A known mass is loaded on the right side to restore the frame to its horizontal position.

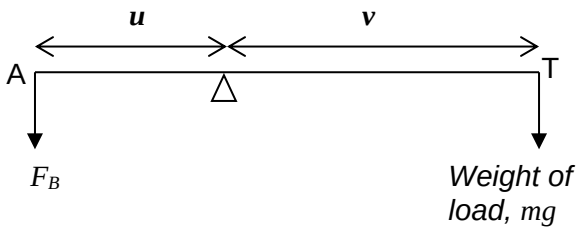
4. Consider the forces acting on the wire frame,

   taking moments about the knife edges,

   clockwise moments = anticlockwise moments

   $(mg)v = (BIL)u$

   $$B = \frac{mgv}{ILu}$$



Side view of wire

14.5 Forces Between 2 Long Parallel Current-Carrying Conductors

Now that we know that a currents can generate magnetic fields like magnets, and that they can also experience a magnetic force when they are near other magnets, let's consider what happens if we replace the magnet with another current-carrying wire.

When 2 parallel long wires are placed in a vacuum, **each wire experiences a force as it will be in a magnetic field generated by the other wire**.

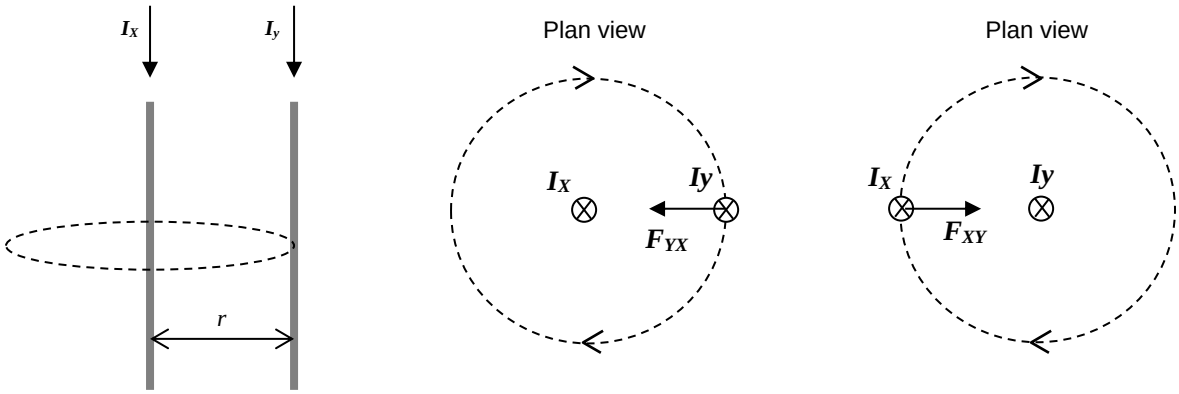


Figure 14.5.1 Two long parallel current-carrying conductors.

Current I_X sets up a magnetic field that interacts with current I_Y. Therefore the force acting on I_Y due to the magnetic field B_X (generated by I_X) is $F_{YX} = B_X I_Y l$ (l : length of the wire)

Recall that B_X is given by $B_X = \dfrac{\mu_0 I_X}{2\pi r}$

Hence the force $F_{YX} = \dfrac{\mu_0 I_X I_Y l}{2\pi r}$

The direction of the force on I_Y can be found using Fleming's Left-Hand Rule. In this case, the force on I_Y due to I_X is towards I_X. We can go through the same analysis to understand the force acting on I_X as a result of I_Y, or we can simply apply **Newton's 3rd law** and conclude that the force on I_X due to I_Y is of the same magnitude but in the opposite direction.

Hence the force on I_X due to I_Y is also $F_{XY} = \dfrac{\mu_0 I_X I_Y l}{2\pi r}$.

It is interesting to note that in this case where the **two currents are flowing in the same direction, the force between them is attractive**. We can apply a similar analysis to currents flowing in opposite directions and find that the force between them is repulsive.

A simple way of remembering the conclusion is

"LIKE CURRENTS ATTRACT, UNLIKE CURRENTS REPEL."

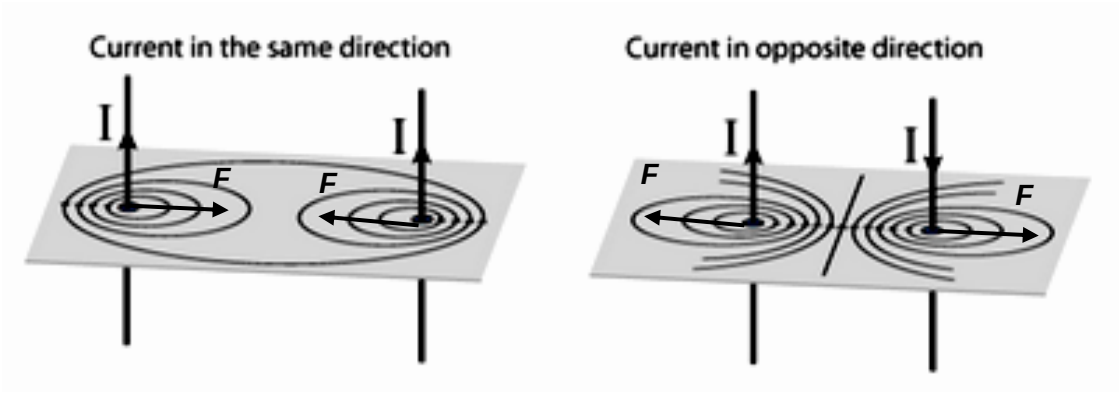
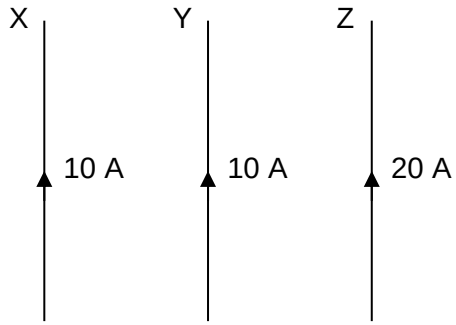


Figure 14.5.2 Sketch showing the magnetic fields and forces on parallel current-carrying wires.

Worked Example 6

X, Y and Z are three coplanar long parallel wires carrying currents as shown.

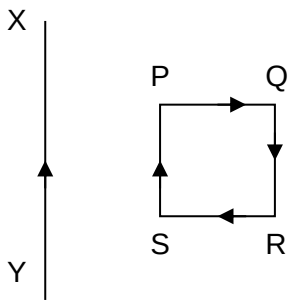


The resultant force on wire Y is

A zero
B towards X
C towards Z
D perpendicular to the paper

Worked Example 7

A long straight wire XY lies in the same plane as a square loop of wire PQRS which is free to move. The sides PS and QR are initially parallel to XY. The wire and loop carry steady currents as shown in the diagram.



What will be the effect on the loop?

A It will move towards the long wire.
B It will move away from the long wire.
C It will rotate about the axis parallel to XY.
D It will be unaffected.
E It will contract.

14.6 Revisiting the DC Motor

Consider a rectangular coil PQRS of *N* turns pivoted so that it can rotate freely about a vertical axis which is perpendicular to a uniform horizontal magnetic field of flux density B. Let the normal to the plane of the coil make an angle θ with the field (Figure 14.6.1 (b)).

The side PS of length *h* is perpendicular to the magnetic field (from N to S) and so the magnetic force *f* on it is given by

$$f = BIh$$

If the coil has *N* turns, the length of the conductor is increased *N* times and so the force acting on side PS is

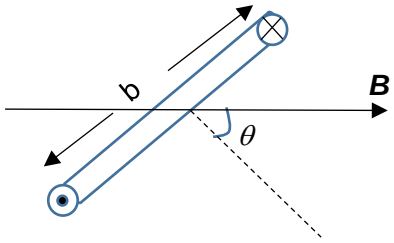
$$F = NBIh$$

The force acting on side QR is also given by $F = NBIh$, but its direction is opposite to that on PS, as shown in Figure 14.6.1 (b). The force on side PQ cancels out the force on side SR as these two forces are equal in magnitude but opposite in direction.

The two forces on sides PS and QR tend to turn the coil about a vertical axis passing through the middle of the coil. These two forces together form a couple and the torque due to this couple is given by

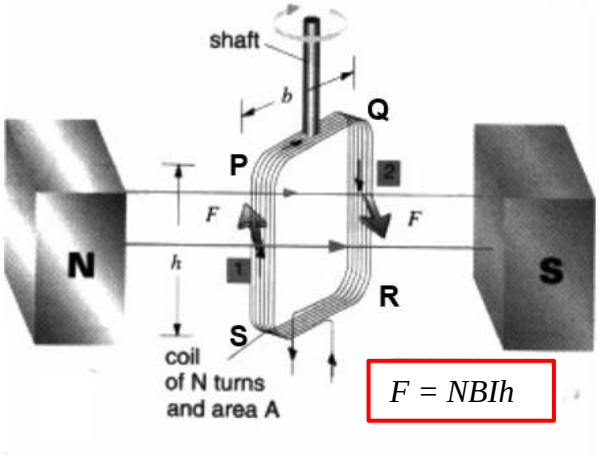
$$\tau = F(b\sin\theta)$$

$$\tau = NBIhb\sin\theta$$

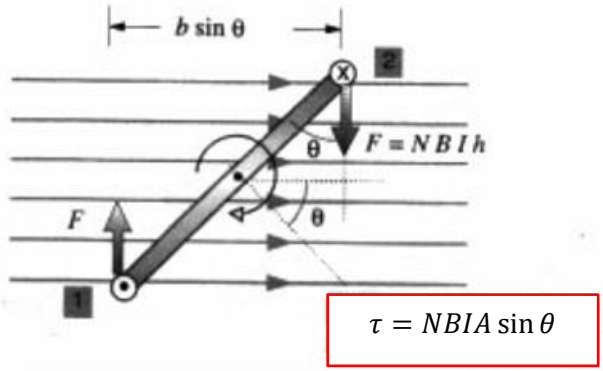


But $h \times b$ = Area, A of the coil. So

$$\tau = NBIA\sin\theta$$



(a) DC Motor



(b) Plan view of the coil and magnetic field

Figure 14.6.1

To provide continuous rotating motion in one direction, the current in the coil must periodically reverse direction.

Tutorial 14A: Magnetic Field and Magnetic Force

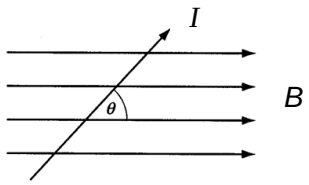
Self-Review Exercises

Use these questions to test your familiarity with the concepts. These questions should be sufficiently easy such that you can solve them on your own, with a little bit of thinking, without help from the tutors. The solutions are made available on Google Classroom for self-check.

S1 Define magnetic flux density and the tesla.

S2 UCLES H1 2008 P1 Q23

A wire carrying a current I is placed in a magnetic field of flux density B.



How is the magnitude F of the force acting on the wire related to the angle θ that it makes with the field?

A $F \propto \theta$ **B** $F \propto \sin\theta$

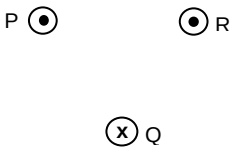
C $F \propto \cos\theta$ **D** $F \propto 1 / \sin\theta$

S3 UCLES N1993 P1 Q16 (modified)

Three long vertical wires pass through the corners of an equilateral triangle PQR. They carry equal currents into or out of the paper in the directions shown.

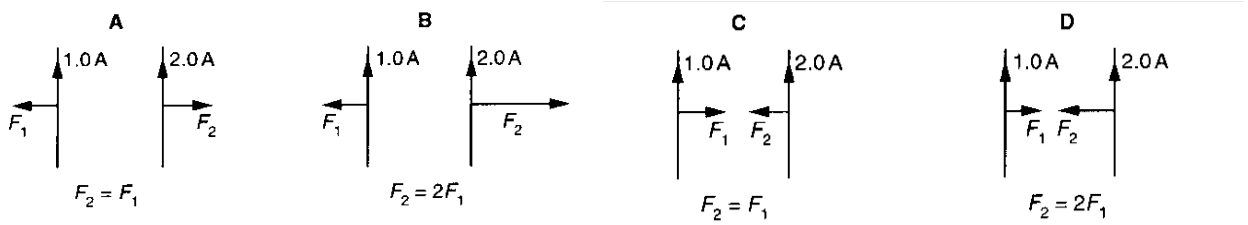
Draw an arrow to represent the resultant magnetic flux density on the wire Q, label the arrow as B_{res}.

Draw an arrow to represent the resultant force on the wire Q, label the arrow as F_{res}.



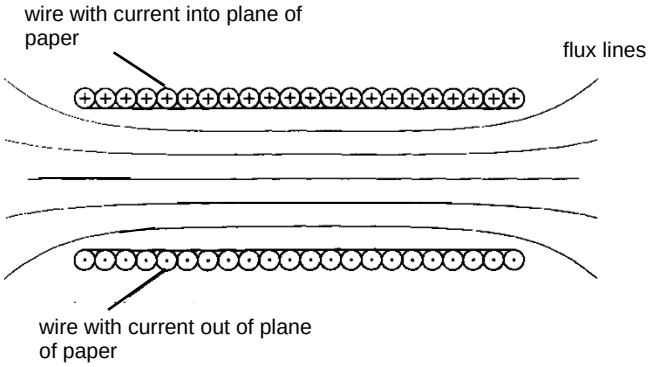
S4 UCLES N1996 P1 Q18

Two long, straight parallel wires carry currents of 1.0 A and 2.0 A. Which diagram shows the directions and relative magnitudes F_1 and F_2 of the forces per unit length on each of the wires?



S5 UCLES J1997 P2 Q3

The diagram illustrates the pattern of the magnetic flux due to a current in a solenoid.



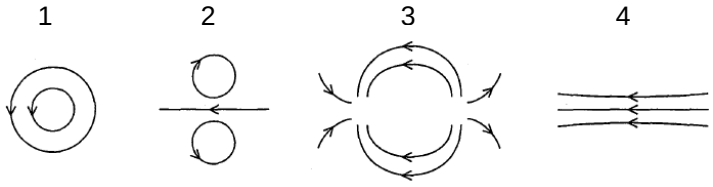
(i) Draw arrows to show the direction of the magnetic field in the solenoid.

(ii) Draw a line to represent a current-carrying conductor in the magnetic field which does not experience a force due to the magnetic field. Label the conductor **C**.

(iii) The coils of wire on an electromagnet are usually wound on a ferrous core. State two properties of the core which are important in its use as an electromagnet.

S6 UCLES H1 2011 P1 Q27, 2013 P1 Q26

Four magnetic fields are shown.

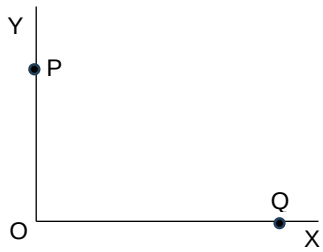


These magnetic fields are produced by a number of devices, which are listed below. What is the correct order for the fields for the devices given in the table?

	Flat circular coil	Bar magnet	Long solenoid	Long straight wire
A	1	3	2	4
B	1	2	4	3
C	2	3	4	1
D	2	4	1	3

S7 **UCLES J2000 P1 Q19**

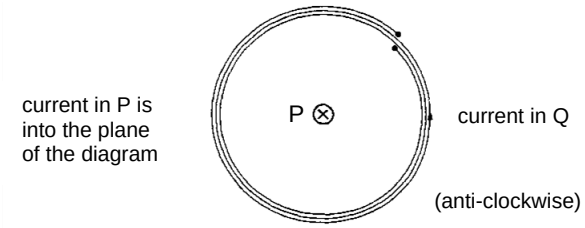
The diagram shows a flat surface with lines OX and OY at right angles to each other. Which current in a straight conductor will produce a magnetic field at O in the direction OX?



A at P into the plane of the diagram
B at P out of the plane of the diagram
C at Q into the plane of the diagram
D at Q out of the plane of the diagram

S8 **UCLES J1999 P1 Q19**

A long straight wire P is placed along the axis of a flat circular coil Q. The wire and the coil each carry a current as shown.

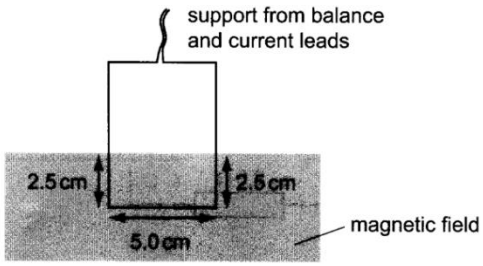


What can be deduced about the force acting on each part of Q due to the current P?

A The force is away from P.
B The force is towards P.
C The force is perpendicular to the plane of the diagram.
D There is no force in any direction.

S9 **UCLES 2007 P1 Q30**

A single-turn rectangular wire loop hangs from a balance reading in grams so that its lower part is in a region of uniform magnetic field. The direction of the field is at right-angles to the plane of the loop. The arrangement is as shown in the diagram.

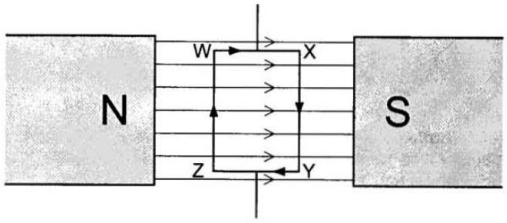


When there is no current in the loop, the reading of the balance is 10.060 g. When the current in the loop is 3.0 A, the balance reading is 10.040 g.
What is the magnitude of the flux density of the field?

A 6.5×10^{-4} T **B** 1.3×10^{-3} T
C 1.3×10^{-2} T **D** 6.6×10^{-1} T

S10 UCLES 2010 P1 Q31

In an electric motor, a rectangular coil WXYZ has 20 turns and is in a uniform magnetic field of flux density 0.83 T.



The lengths of sides XY and ZW are 0.17 m and of sides WX and YZ are 0.11 m. The current in the coil is 4.5 A.

What is the maximum torque provided by the motor?

A 0.070 Nm **B** 0.63 Nm

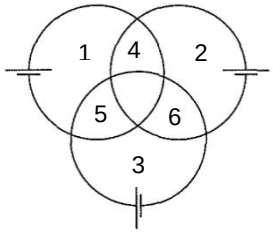
C 1.4 Nm **D** 2.8 Nm

Discussion Questions

These are questions for you to apply the concepts. Your tutor will select some questions from this set of questions to discuss in class. The rest of the questions will be for your own further practice after class.

D1 UCLES 2015 P1 Q30

Three separate coils of insulated wire are connected to cells as shown. They are placed on a table on top of each other, partially overlapping.



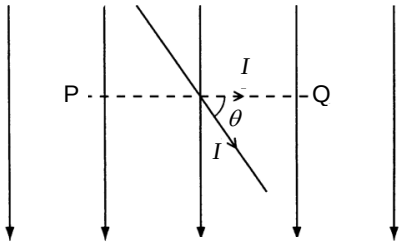
Six of the seven areas formed within the coils are numbered. In which areas do the magnetic fields of all the coils reinforce each other?

A 1 and 6 **B** 2 and 5

C 3 and 4 **D** 4

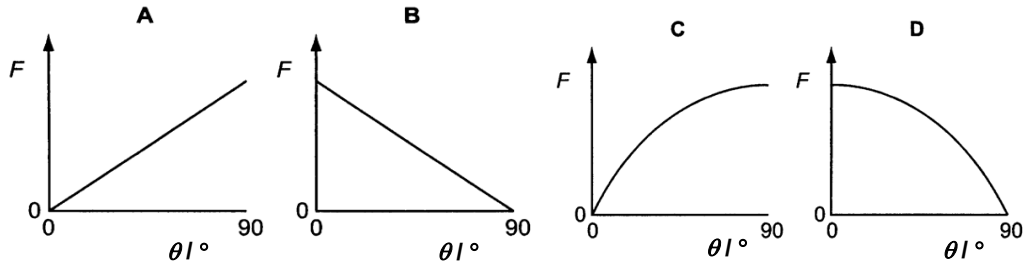
D2 UCLES 2006 P1 Q23

A straight wire PQ carrying a constant current I is placed at right angles to a uniform magnetic field, as shown by the dotted line in the diagram.



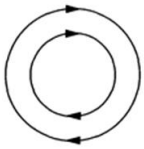
magnetic field

The wire is then rotated through an angle about an axis perpendicular to the plane of the diagram. Which graph shows how the magnitude of the magnetic force *F* on the wire varies with θ in the range 0° to 90°?



D3 UCLES 2000 P1 Q18

A small flat circular coil lies inside a similar larger coil. Each coil carries a current as shown. What is experienced by the small coil due to these currents?

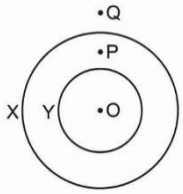


view from above

A a torque about the horizontal axis
B a torque about the vertical axis
C a vertical force along the axis
D no resultant force

D4 **UCLES 2009 P1 Q31 and 2019 Q26**

X and Y are two coaxial circular coils lying on a table. O, P and Q are three points on the table.

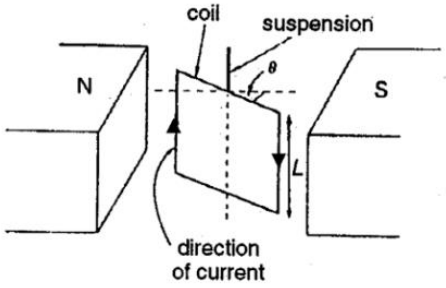


Initially, there is a constant current in coil X and no current in coil Y. A small current is now passed through coil Y, which decreases the magnitude of the magnetic flux density at O.
How does the magnitude of the flux density change at P and Q?

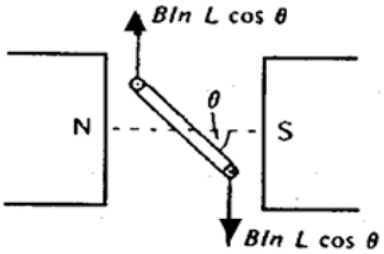
	P	Q
A	decreases	decreases
B	decreases	increases
C	increases	decreases
D	increases	increases

D5 **UCLES J1987 P1 Q18**

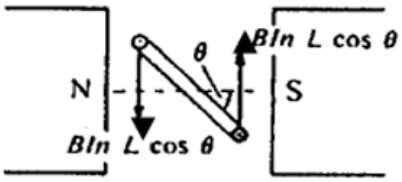
A current *I* is carried by a square coil of *n* turns and side *L*, suspended vertically as shown in a uniform horizontal magnetic field of flux density *B*. Which one of the plan diagrams correctly shows the magnitude and direction of the forces acting on the vertical sides of the coil?



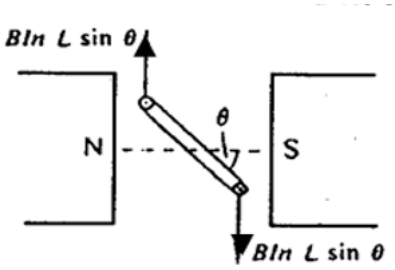
A



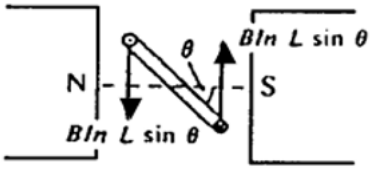
B



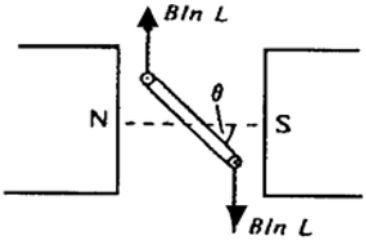
C



D

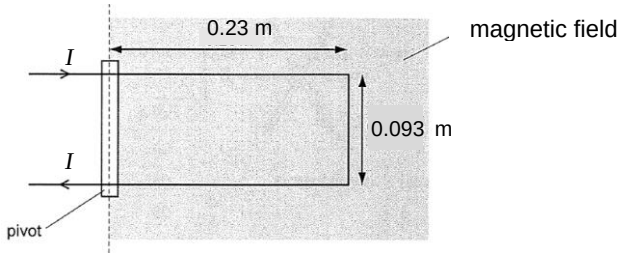


E



D6 UCLES 2012 P1 Q34

In order to determine the value of a current *I*, it is passed into a current balance. This consists of a U-shaped wire placed in a constant magnetic field of flux density 3.6×10^{-2} T. The U-shaped wire has length 0.23 m and the arms are 0.093 m apart, as shown in the diagram.



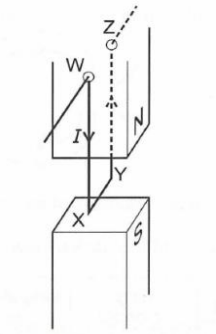
The U-shaped wire experiences a turning moment about the pivot of value 7.4×10^{-3} N m. What is the value of *I*?

A 0.044 A **B** 1.6 A
C 2.8 A **D** 9.6 A

D7 UCLES H1 2011 P1 Q26

A piece of wire WXYZ is pivoted freely about a horizontal axis at points W and Z. Section XY of the wire is situated between the North (N) and South (S) poles of a magnet.

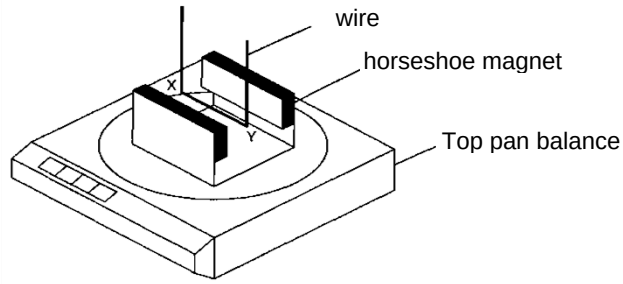
WXYZ is connected to an electrical circuit. The circuit is closed and there is a constant current *I* in the wire.
Which of the following describes the motion of the wire, if any, after the circuit is closed?



A Swings to the left and becomes stationary in a new position.
B Swings to the right and becomes stationary in a new position.
C Oscillates continually in the magnetic field.
D Remains at rest at its original position in the magnetic field.

D8 UCLES 2000 P1 Q19

A horseshoe magnet rests on a top-pan balance with a wire situated between the poles of the magnet. With no current in the wire, the reading on the balance is 142.0 g. With a current of 2.0 A in the wire in the direction XY, the reading changes to 144.6 g.



(a) What is the reading on the balance, when there is a current of 3.0 A in the wire in the direction YX?

(b) Identify the direction of magnetic field in the horseshoe magnet. Explain how you deduce your answer.

D9 UCLES 2019 P2 Q4 (modified)

Two long, straight wires X and Y are separated by a distance of 0.12 m. The current in wire X is 5.5 A. The current in wire Y is 8.4 A.

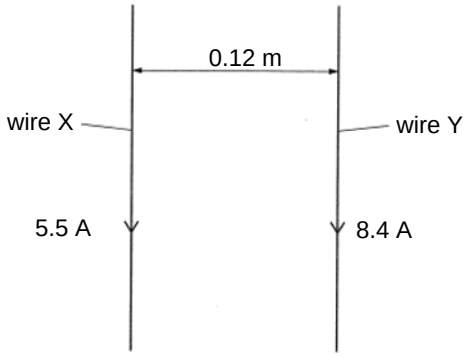


Fig. 4.2

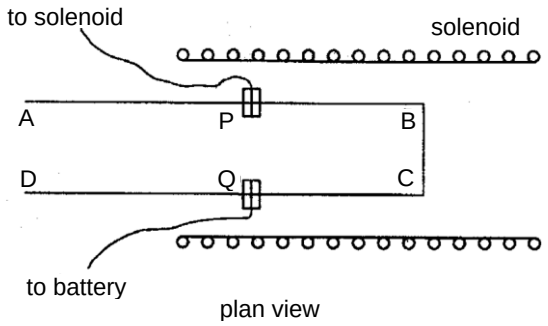
Both currents are in the same direction, as shown in Fig. 4.2.

(a) Explain, in terms of physical principles, the direction of the force experienced by both wires.

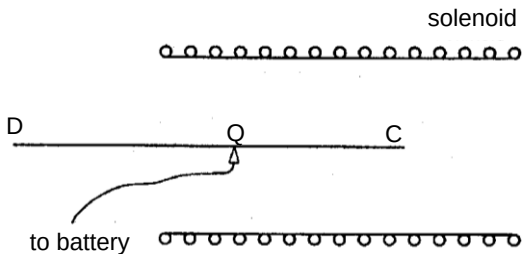
(b) Calculate the force per unit length on wire X.

D10 UCLES 1994 P2 Q4

A wire frame ABCD is supported on two knife edges P and Q so that PBCQ of the frame lies within a solenoid. Electrical connections are made to the frame through the knife edges so that the part PBCQ of the frame and the solenoid can be connected in series with a battery. When there is no current in the circuit, the frame is horizontal.



plan view



side view

(a) When the frame is horizontal and a current passes through the frame and the solenoid, what can you say about the direction of the force, if any, due to the magnetic field of the solenoid acting on

(i) side BC,

(ii) side PB?

(b)* State two ways in which you could reverse the direction of the force on the side BC.

(c) The solenoid has 700 turns per metre and carries a current of 3.5 A. Given that the magnetic flux density B on the axis of a long solenoid is $B = \mu_0 nI$,

(i) Calculate the magnetic flux density in the region of side BC of the frame.

(ii) Side BC has a length of 5.0 cm. Calculate the force acting on BC due to the magnetic field in the solenoid.

(iii) A small piece of paper of mass 0.10 g is placed on the side DQ and positioned to keep the frame horizontal. Given that QC is of length 15.0 cm, how far from the knife-edge must the paper be positioned?

D11 UCLES 2017 P2 Q4

Part of a long current-carrying conductor (wire) is shown in Fig. 4.1. The conductor is normal to and passes through a horizontal plane.

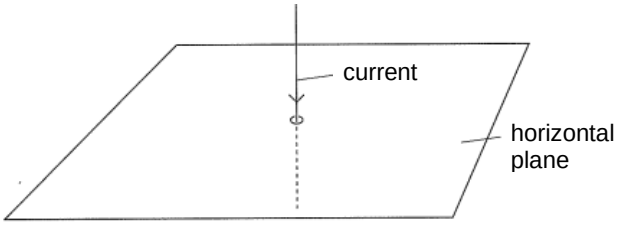


Fig. 4.1

Fig. 4.2

Fig. 4.2 shows the current in the wire directed perpendicularly into the plane of the paper.

(a) On Fig. 4.2, draw lines to present the magnetic field in the horizontal plane around the long wire. [2]

(b) The current in the wire in (a) is 8.5 A.
Calculate the magnetic flux density B due to the current in the wire at a distance of 19 cm from the centre of the wire in the horizontal plane. [2]

(c) The current in the wire is switched off and a small compass is placed 19 cm due North of the wire, as shown in Fig. 4.3a.

The current of 8.5 A is then switched on and the needle deflects by 12° of the compass as shown in Fig. 4.3b.

Fig. 4.3a

Fig. 4.3b

(i) Use your answer in (b) to calculate the horizontal component B_H of the magnetic flux density due to the Earth's magnetic field at this location. [2]

(ii) On Fig. 4.3b, mark with a letter X the position where the resultant horizontal magnetic flux density is zero. Explain your reasoning. [2]

Numerical Answers to Selected Questions in Tutorial 14A

S2 B **S4** C **S6** C **S7** B **S8** D **S9** B **S10** C

D8(a) 138.1 g **D9(b)** 7.7×10^{-5} N m^{-1}

D10(c)(i) 3.08×10^{-3} T **(c)(ii)** 5.39×10^{-4} N **(c)(iii)** 8.24 cm

D11(b) 8.9×10^{-6} T **(c)(i)** 4.2×10^{-5} T

14.7 Magnetic Force on a Moving Electric Charge

If we look into the microscopic view of the charges moving in a magnetic field, an expression for the magnitude of the force acting on a charge Q moving with velocity $\boldsymbol{v}$ in a magnetic field of flux density $\boldsymbol{B}$ is given by

$$F_B = B_\perp Q v$$

where $\boldsymbol{B_\perp}$ is the component of the magnetic flux density perpendicular to the velocity of the moving charges.

Note:

1. The magnetic force is maximum when the charge is moving perpendicularly to the magnetic field.
2. A charge in a magnetic field does not experience any force when
   - it is moving parallel to the direction of the magnetic field,
   - it is stationary in the field.
3. If the direction of the magnetic flux density and the velocity of the charged particles are perpendicular, the direction of the magnetic force on a charge moving in a magnetic field can be determined using Fleming's Left Hand Rule. By convention, the direction of conventional current is given by the direction of flow of positive charges.

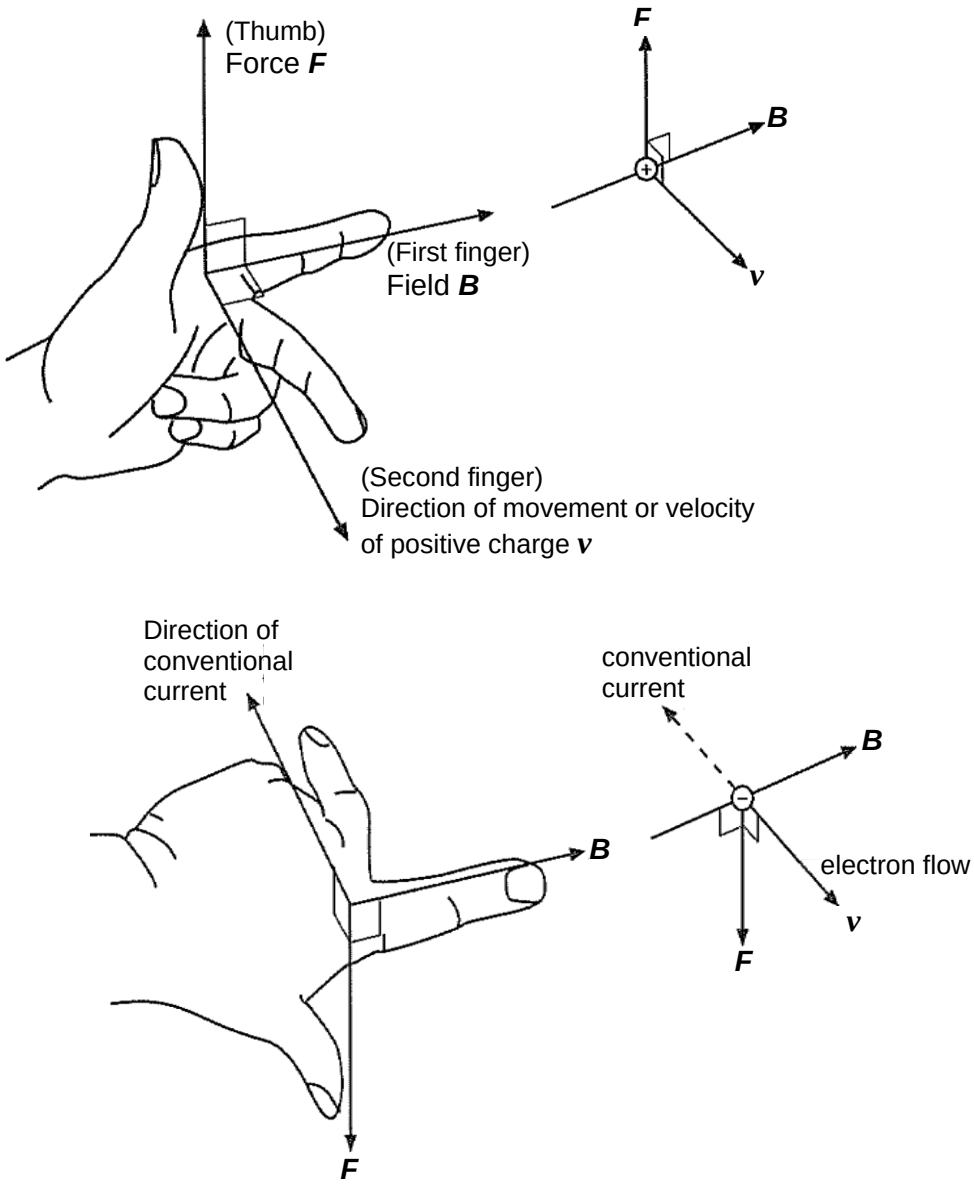
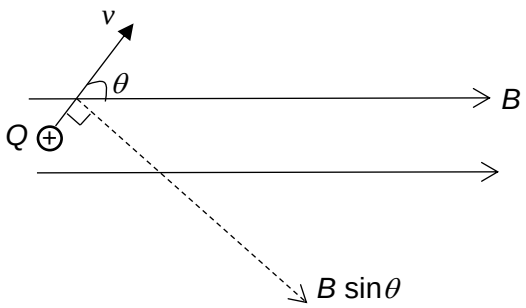


Figure 14.7.2 Applying Fleming's Left-Hand Rule on the motion of positive and negative charge

Example: For a particle with a positive charge Q moving with a velocity v at an angle θ to the magnetic field as shown below:



Magnetic force F acting on the moving charge is directed into the plane of the paper (***Fleming's Left-Hand Rule***)

$$F = (B \sin\theta)Q\, v$$
$$= BQv \sin\theta$$

14.8 Motion of a Charged Particle in a Uniform Magnetic Field

Let's now look at how moving charges that are not confined within a wire are affected by a magnetic field. We will again use Fleming's Left-Hand Rule to find the direction of this force.

Consider a positively charged particle moving at velocity v in a uniform magnetic field B.

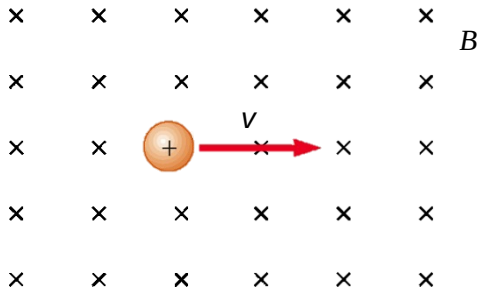


Figure 14.8.1 Charge moving at a velocity v in a uniform magnetic field

Using Fleming's Left Hand Rule, and the concept that the direction of the convention current is the direction of motion of the positively charged particle, we can deduce that the direction of the magnetic force acting on the charged particle at this instant is upwards within the plane of the plane. This causes the particle to be deflected from its initial path.

However, the direction of the force remains perpendicular to the directions of both velocity and magnetic field (Fleming's Left Hand Rule) even as the charged particle deflects, hence the charged particle will move in a circular path! The **magnetic force acting on the moving charge provides the centripetal force** for it to move in a circular path.

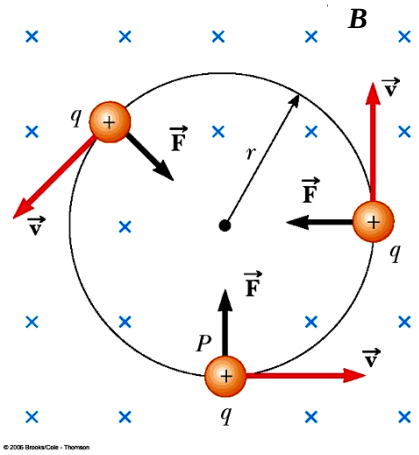


Figure 14.8.2 A positively charged particle performing circular motion in a magnetic field

The magnetic force on the positive charge provides the centripetal force:

$$Bqv = \frac{mv^2}{r}$$, where r is the radius of the circular path

$$r = \frac{mv}{Bq}$$

From this equation for the radius of the path, we can deduce that for particles entering the magnetic field with the **same velocity**, the **radius would be proportional to the mass-to-charge ratio** of the particle (i.e. $\frac{m}{q}$). The ratio of the charge on a particle to the mass of the particle (i.e. $\frac{q}{m}$) is known as the ***specific charge***.

If *T* is the time taken to complete one revolution of the circular path and *v* is the constant speed of the charge,

$$v = r\omega$$

$$v = r\frac{2\pi}{T}$$

$$v = \left(\frac{mv}{Bq}\right)\frac{2\pi}{T}$$

$$T = \frac{2\pi m}{Bq}$$

The period *T* is dependent on the mass, the magnetic flux density, and the charge of the particle. It is independent of the speed of the particle or the radius of the circular path.

1. Similar to the force acting on a straight current-carrying conductor placed in the magnetic field, we have to remember that the *angle* ***θ is the angle between the direction of motion and the magnetic field lines***. Examiners may try to trick you by definiting a different *θ*.

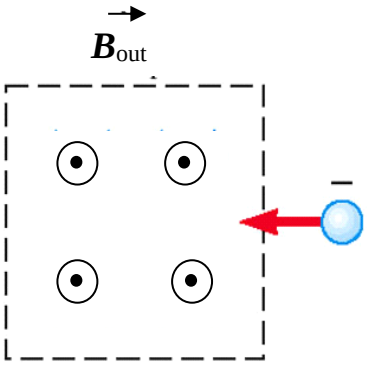
2. For negatively charged particles moving in a magnetic field, the magnitude of the force is still the same. However, when determining the direction of the force, for *positively charged particles, the direction of I is the direction of v, but for negatively charged particles, the direction of I is opposite to the direction of v.*

3. It is very common in exam questions to be asked to draw the path taken by moving charges in a magnetic field. Apply Fleming's Left Hand Rule (remember the sign of the charge is important!) to find the direction of the magnetic force, which is also the direction of the centripetal force. The centripetal force acts towards the centre of the circular motion, and hence deflects the moving charge in the direction of the force, or imagine where the centre of the circular motion is, and you should be able to draw the path correctly. Please ensure that the path is as circular as possible, that is, you should draw with a set of compasses. If you really don't have one in the exam, then label the path drawn as 'circular path'.

4. The **magnetic force does no work on the charged particle during the circular motion** as it is perpendicular to *v*. Hence, the speed and kinetic energy of the charge remain the same as it moves along the circular path.

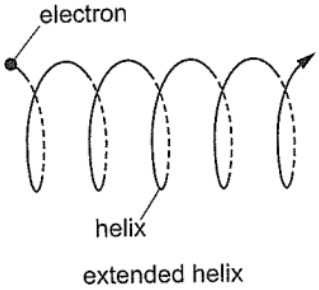
Worked Example 8

An electron with speed 2.0 x 10^6 m s^{-1} enters a magnetic field of flux density 2.0 mT perpendicular to its direction of motion. The magnetic field only exists within a square of size 8.0 mm. Draw the path taken by the electron as its moves through the magnetic field.



Worked Example 9 (**UCLES 2016 P1 Q31**)

The force on an electron is 7.3 x 10^{-16} N when moving at an angle of 23° to a uniform magnetic field of magnetic field strength 0.084 T.



extended helix

flat spiral

What is the speed of the electron and what is the shape of the path of the electron?

A 1.4 x 10^5 m s^{-1} in a helix

B 1.4 x 10^5 m s^{-1} in a spiral

C 2.1 x 10^4 m s^{-1} in a helix

D 1.1 x 10^4 m s^{-1} in a spiral

Tutorial 14B: Moving Charges in Magnetic Fields

Self-Review Exercises

Use these questions to test your familiarity with the concepts. These questions should be sufficiently easy such that you can solve them on your own, with a little bit of thinking, without help from the tutors. The solutions are made available on Google Classroom for self-check.

S1 There are two situations in which a charged particle in a magnetic field does not experience a magnetic force. State these two situations.

S2 **UCLES J1989 P1 Q26**

A charged particle is situated in a region of space and it experiences a force only when it is in motion. It can be deduced that the region encloses

A both an electric and gravitational field

B both a magnetic and an electric field

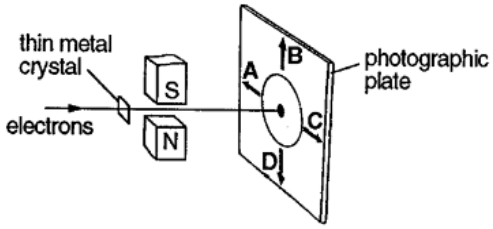
C both a magnetic and a gravitational field

D a magnetic field only

E an electric field only

S3 **UCLES 1990 P1 Q19**

GP Thomson's early experiments on the diffraction of electrons by crystals were criticized on the grounds that the beams affecting the photographic plates might be X-rays. He proved that this was not so by placing bar magnets, on each side of the beam as shown in the diagram. How would the magnetic field due to the magnets affect the diffraction ring?



A The ring would deflect in the direction of A.

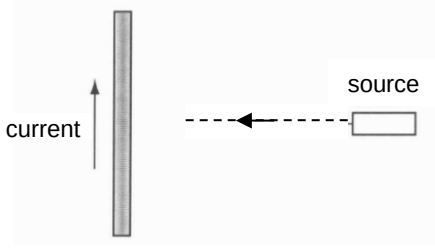
B The ring would deflect in the direction of B.

C The ring would deflect in the direction of C.

D The ring would deflect in the direction of D.

S4 **UCLES H1 2010 P1 Q27**

The diagram shows a long vertical wire carrying a steady current. A source fires a horizontal beam of positively-charged particles towards the wire.



As they approach the wire, the particles experience an electromagnetic force from the field of the current.

What is the direction of this force?

A In a direction parallel to the wire

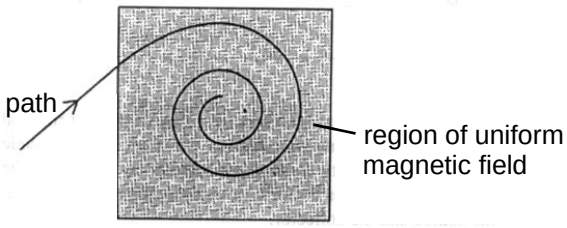
B In a horizontal direction at 90° to the original path

C In a horizontal direction towards the wire

D In a horizontal direction at 180° to the original path

S5 **UCLES 1993 P1 Q26**

A common way of investigating charged particles is to observe how they move in a plane at right angles to a uniform magnetic field. The diagram shows the path of a certain particle. Which of the following gives a satisfactory explanation for the path?



A The momentum of the particle is increasing steadily.

B The charge on the particle is decreasing steadily.

C The magnetic flux density is decreasing steadily.

D The mass of the particle is increasing steadily.

E The speed of the particle is decreasing steadily.

S6 **UCLES 1987 P1 Q14**

An electron of mass m_e, and charge e moves in a vacuum with uniform speed v. A uniform magnetic field of flux density B acts perpendicular to the direction of v.

(a) Explain why the electron moves in a circular orbit and show by means of a careful sketch the direction of its rotation relative to the direction of B.

(b) Find an expression for the radius of its orbit in terms of v, B e and m_e.

(c) Hence show that the time taken to complete one revolution in the orbit is given by

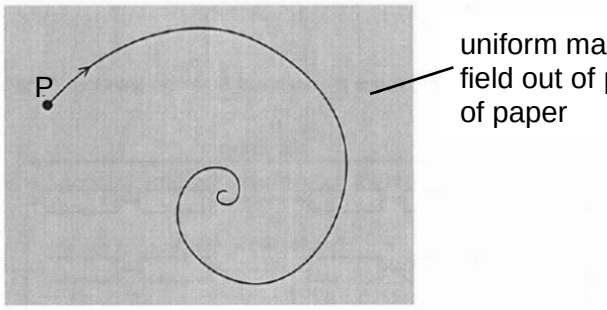
$$t = \frac{2\pi m_e}{Be}$$

Discussion Questions

These are questions for you to apply the concepts. Your tutor will select some questions from this set of questions to discuss in class. The rest of the questions will be for your own further practice after class.

D1 UCLES 2009 P1 Q30

A charged particle is moving in a region where there is a uniform magnetic field acting out of the plane of the paper. The path of the particle begins at P and is as shown.



What can be deduced about the charge on the particle and its speed?

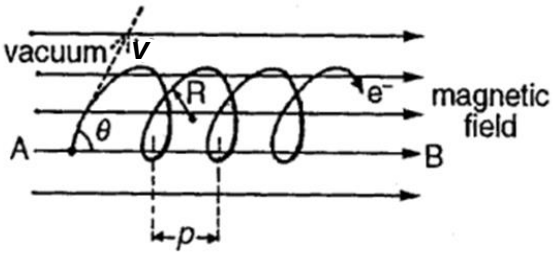
	charge	speed
A	negative	decreasing
B	negative	increasing
C	positive	decreasing
D	positive	increasing

D2 **(a)** Describe the shape of the paths followed by electrons injected

(i) at right angles into a uniform magnetic field

(ii) at right angles into a uniform electric field

(b) An electron is injected at a speed v of 7.0×10^6 m s^{-1} at an angle θ into a uniform magnetic field of flux density 3.14×10^{-5} T. It describes a helical path as shown in the diagram. Assuming that $\sin\theta = 0.6$ and $\cos\theta = 0.8$, show that the velocity component of the electron perpendicular to the field is 4.2×10^6 m s^{-1}.



(c) Hence calculate

(i) the radius of the helical path,

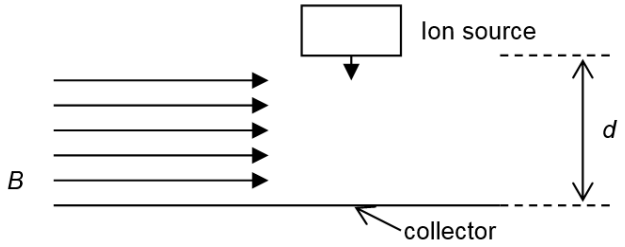
(ii) the time taken for the electron to complete one revolution in the helix,

(iii) the velocity component of the electron parallel to the field,

(iv) the pitch, p, of the helix.

D3 **UCLES J1988 P1 Q27**

An ion source is at distance *d* from a flat, horizontal collector at the same potential as the source. A magnetic field of flux density *B* acts horizontally as shown in the diagram. The field is uniform throughout the region between the source and the collector.



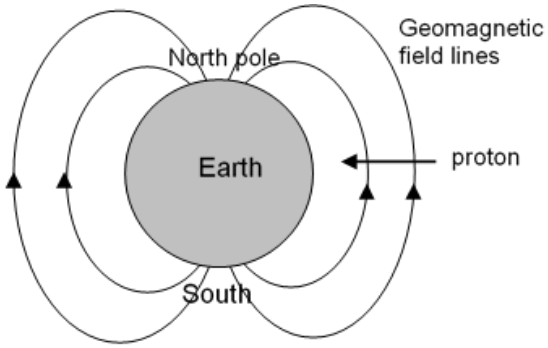
An ion of charge *q* and mass *m* is emitted vertically downwards at speed *v*. Under what conditions will the ion reach the collector?

A $v > \sqrt{\frac{2Bq}{m}}$ **B** $v < \sqrt{\frac{2Bq}{m}}$

C $v = \sqrt{\frac{dBm}{q}}$ **D** $v > \frac{dBq}{m}$

E $v < \frac{dBq}{m}$

D4 Cosmic rays are charged particles emitted by stars and they travel at very high speeds. These particles bombard the earth all the time. The majority of these particles are protons traveling close to the speed of light. A proton moves in the direction towards the earth as shown in the diagram.



(a) Find the direction of the magnetic force acting on the proton.

(b) Explain how the Earth is shielded from the protons by the geomagnetic field.

(c) Hence, explain which parts of the Earth are least shielded from the protons.

Numerical Answers for Tutorial 14B:

S2 D **S3** A **S4** A **S5** E **S6(b)** $r = \frac{m_e v}{Be}$

D2 (c)(i) 0.762 m **(ii)** 1.14 μs **(iii)** 5.6 x 10^6 m s^{-1} **(iv)** 6.38 m

14.9 Moving Charges in Uniform Electric and Magnetic fields

The table below summarizes the differences in the deflection of charged particles in a magnetic and electric field.

Deflection in a uniform magnetic field	Deflection in a uniform electric field
1. The magnetic field can exert a magnetic force only on a moving charged particle. Stationary charged particles experience no force.	1. The electric field always exerts an electric force on a charged particle, whether it is stationary or moving.
2. The magnetic force is perpendicular to the magnetic field and to the direction of motion of the charged particle. Deduced using Fleming's Left Hand Rule.	2. The electric force acts in the direction parallel to that of the electric field. Deduced from the law of electrostatics (i.e. like charges repel; unlike charges attract).
3. Magnetic force is dependent on the speed and direction of motion of the charged particle ($F_B = Bqv\sin\theta$).	3. Electric force is not dependent on the speed and direction of the charged particle ($F_E = qE$).
4. Circular motion is obtained when a charged particle enters a magnetic field perpendicularly.	4. Parabolic motion is obtained when a charged particle enters a uniform electric field perpendicularly.

For an electric field of 10^4 N C^{-1}, the ratio of the magnitude of the electric force qE to the magnitude of the gravitational force mg is of the order of 10^{14} for an electron and 10^{11} for a proton.

Thus, we can consider an electron only experiences an electric force in an E field. Once it leaves the E field, it will obey Newton's first law of motion and continue to move in a straight line at constant speed.

It is possible for a charged particle to move into a region where a uniform electric field and a uniform magnetic field are applied. The charged particle will experience an electric force and a magnetic force simultaneously.

We will use Newton's laws of motion to determine the resultant force acting on the charged particle, to predict its motion.

In the special case where the electric force and magnetic force are of **equal magnitude and opposite in direction**, the charged particle will travel in a straight line as the resultant force is zero.

Let us take a closer look at a practical example as follows.

14.10 Velocity Selector

In many experiments involving moving charged particles, it is important that all particles move with essentially the same speed. This can be achieved by applying a combination of a uniform electric field and a uniform magnetic field oriented as shown in Figure 14.10.

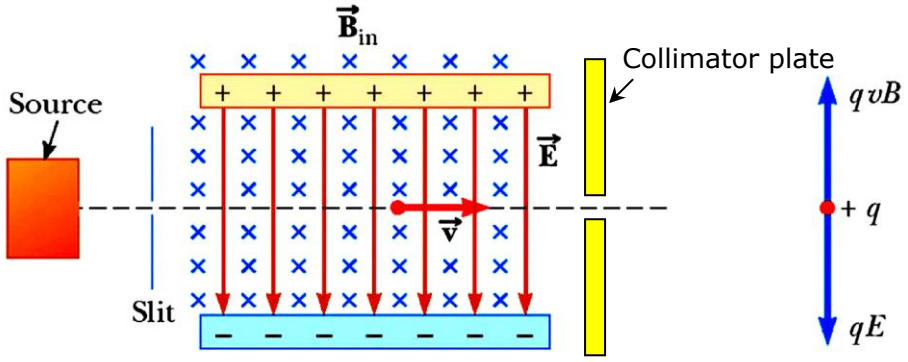


Figure 14.10 Velocity Selector

A uniform *E* field is applied vertically downwards and a uniform magnetic field is applied in the direction perpendicular to the electric field. Consider a positive charge entering the region from the left with velocity *v*.

It experiences an electric force *qE* that is downwards (in the same direction as the electric field) and a magnetic force *Bqv* that is upwards (Fleming's Left Hand Rule). By adjusting the magnitudes of the electric and magnetic fields, we can set up a scenario so that *qE* = *Bqv* and **the particle experiences no resultant force**, and moves through the fields in a straight horizontal line!

$$\Sigma F = 0$$
$$F_B = F_E$$
$$Bqv = qE$$
$$v = \frac{E}{B} = \frac{V}{Bd}$$

When we have **<u>many particles of *different speeds*</u> being projected into this region**, **only particles of velocity** $v = \frac{E}{B}$ **will pass through undeflected**.

For particles with speeds **greater than this**, the **magnetic force will be larger than the electric force** and deflect upwards.

For particles with **speeds less than this**, the **magnetic force will be less than the electric force**, and they deflect downwards.

Not all uniform E and B fields that are perpendicular to each other will work as a velocity selector. It is important to note that the concept behind this application is that the **particle experiences no resultant force when travelling in the crossed fields**. Hence it is important to ensure that the **direction of the electric force is opposite to the direction of the magnetic force** set up by the two fields.

Tutorial 14C: Moving Charge in Electric and Magnetic Fields

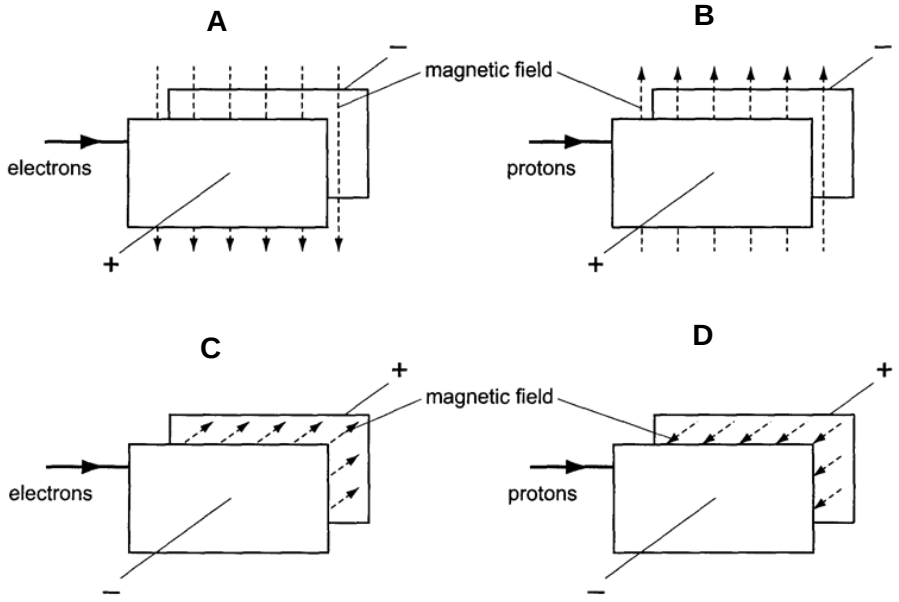
Self-Review Exercise

Use these questions to test your familiarity with the concepts. These questions should be sufficiently easy such that you can solve them on your own, with a little bit of thinking, without help from the tutors. The solutions are made available on google classroom for self-check.

S1 **UCLES 2008 P1 Q31**

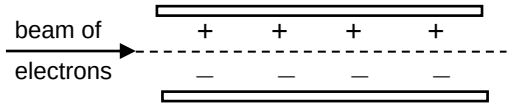
The diagrams show different particle beams entering a region between two metal plates in which there are uniform and magnetic fields.

In which arrangement would it be possible for the beam to pass through undeflected?



S2 **UCLES 2012 P1 Q33**

A beam of electrons, travelling horizontally towards the right, passes between two horizontal charged parallel metal plates.



There is a vertical electric field between the plates. There is no deflection of the beam because of the presence of a uniform magnetic field in the region between the plates.

In which direction must this magnetic field be?

A Out of the paper

B Into the paper

C Vertically upwards

D Vertically downwards

S3 A small mass experiences a force when placed in a field of force. The field may be electric, magnetic or gravitational. State the type of field when the mass is

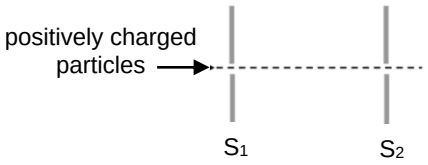
(a) uncharged and the force is in the direction of the field

(b) charged and the force is in the opposite direction to the field

(c) charged and the force is experienced only when it is moving

S4 UCLES J1993 P2 Q7

A narrow beam of identical positively charged particles passes through two slits S_1 and S_2 as shown.



A uniform magnetic field of flux density *B* is applied in the region between S_1 and S_2 in the direction out of the plane of the paper.

(a) A uniform electric field is applied in the space between the slits such that charged particles of only one speed *v* can pass through S_2. On the diagram, mark clearly with an arrow labelled *E* the direction of this field.

(b) Explain how this combination of magnetic and electric fields allows particles of only one speed to pass through S_2. Deduce an expression for *v* in terms of *B* and *E*.

(c) Sketch on the diagram, a possible path in the region of the electric and magnetic fields, of particles having a speed greater than *v*.

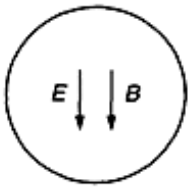
Discussion Questions

These are practice questions for you to apply the concepts. Your tutor will select some questions from this set of questions to discuss in class. The rest of the questions will be for your own further practice after class.

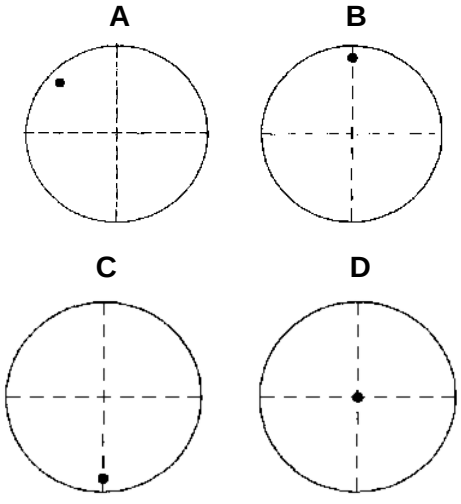
D1 UCLES J1997 P1 Q27

In a cathode ray oscilloscope tube, the electron beam passes through a region where there are electric and magnetic fields directed vertically downwards as shown. The deflections of the spot from the centre of the screen produced by both the electric and magnetic field acting separately are of equal magnitude.

front view of screen

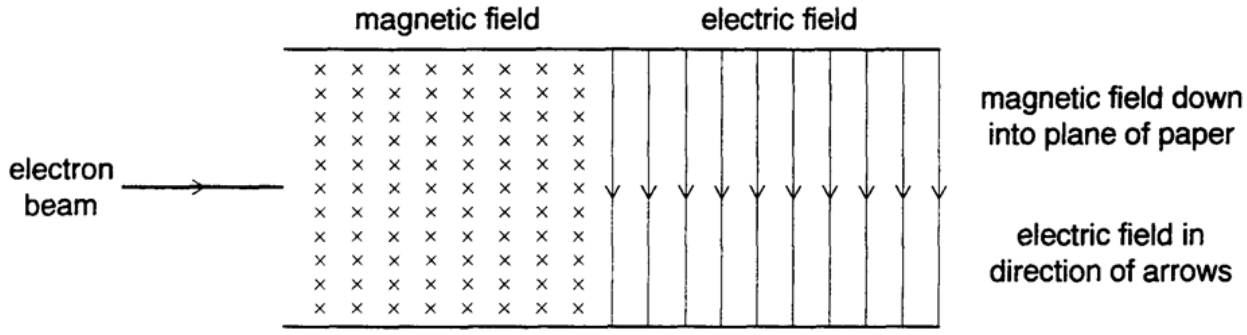


Which diagram shows a possible position of the spot on the screen when both fields are operating together? (HINT: the electrons travel out of the paper since the screen is shown on the paper.)



D2 UCLES J1999 P2 Q6 (part)

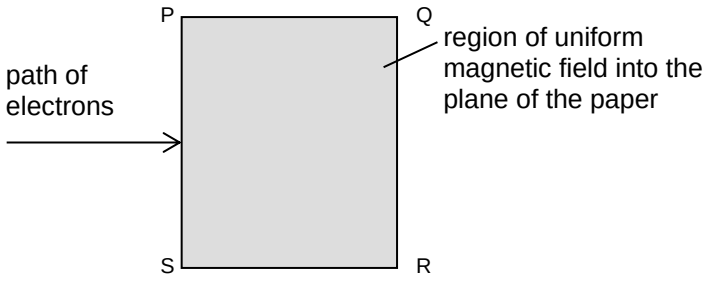
An electron beam passes through adjacent magnetic and electric fields as shown.



On the diagram, sketch a possible path of an electron through both fields.

D3 UCLES 2001 P2 Q8

The diagram shows a region PQRS of a uniform magnetic field directed downwards into the plane of the paper. Electrons, all having the same speed, enter the region of the magnetic field.



(a) On the diagram, show the path of the electrons as they pass through the magnetic field, emerging from the side QR.

(b) A uniform electric field is also applied in the region PQRS so that the electrons now pass undeflected through this region. On the diagram, mark with an arrow labelled *E*, the direction of the electric field.

(c) The undeflected electrons in (b) each have a charge *-e*, mass *m* and speed *v*. State and explain the effect, if any, on particles entering the region PQRS of the same magnetic and electric fields as in (b) if the particles each have

(i) charge *-e*, mass *m* and speed *2v*

(ii) charge *+e*, mass *m* and speed *v*

D4 **UCLES 2010 P3 Q3**

Force-fields may be represented using lines that have direction.

(a) State

(i) what is meant by a *field of force*, [1]

(ii) how, using lines of force, changes in the strength of a force-field are represented. [2]

(b) Conventionally, arrows on field lines define the direction of a force acting on an object.

State the property of the object that experiences a force in this direction for

(i) a gravitational field, [1]

(ii) an electric field, [1]

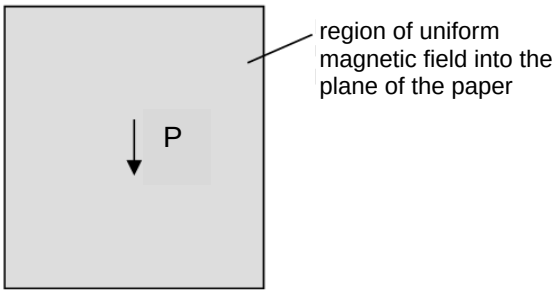
(iii) a magnetic field. [1]

(c) Explain how an electric field and a magnetic field may be used for the velocity selection of charged particles. You may draw a diagram if you wish. [4]

D5 **UCLES 2003 P2 Q6 (part)**

(b) An electron is travelling at right angles to a uniform magnetic field of flux density 1.5 mT, as shown. The field is directed into the plane of the paper. When the electron is at P, its velocity is 2.9×10^7 m s^{-1} in the direction shown.

This is normal to the magnetic field.



(i) Sketch the path of the electron on the diagram, assuming that it does not leave the region of the magnetic field. [2]

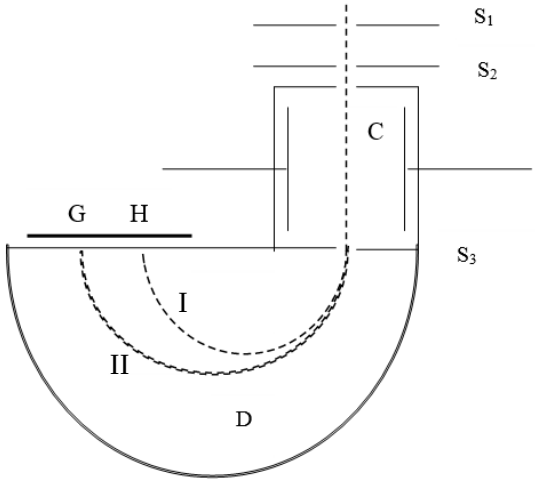
(ii) Calculate for this electron

1. the force on it due to the magnetic field,
2. the radius of its path. [4]

(iii) A uniform electric field is now produced in the same direction and in the same region as the magnetic field. Suggest the shape of the resultant path of the electron. You may draw a sketch to illustrate the path if you wish. [2]

D6 **HCI 2004 Prelim P2 Q6**

The diagram shows a mass spectrometer in which singly-ionised atoms of Ne-20 and Ne-22 pass through slits S_1 and S_2 before entering compartment C. The charge of the ions is positive. C is a velocity selector and consists of mutually perpendicular magnetic field *B* of strength 0.0870 T and electric field of field strength 3.06×10^3 V m^{-1}. Compartment D has only a magnetic field strength of the same magnitude and direction as compartment C. The selected ions of both Ne-20 and Ne-22 after passing through S_3 are deviated by the magnetic field in D and travel different paths reaching the photographic plate at G and H.



(a) Indicate on the diagram

(i) the direction of the magnetic field *B*, [1]

(ii) the polarities of the plates in compartment C using "+" and "-" signs. [1]

(b) Calculate the velocities of the ions of both Ne-20 and Ne-22. [3]

(c) Explain why the paths of both the isotopes of neon ions are circular. [1]

(d) Identify the ions travelling in path I. [1]

(e) Determine the separation GH when the isotopes arrive at the photographic plate. [3]

Numerical Answers on Selected Questions in Tutorial 14C:

S1 B **S2** B

D5 (b)(ii)1. 6.96×10^{-15} N **(b)(ii)2.** 0.11 m

D6 (b) 3.52×10^4 m s^{-1} **(e)** 0.0168 m

APPENDIX A: Earth as a Magnet

A suspended magnet or compass points northward because Earth itself is a huge magnet. The compass aligns with the magnetic field of the Earth. The magnetic poles of the Earth, however, do not coincide with the geographic poles – in fact, the magnetic and geographical poles are widely separated. The magnetic pole in the northern hemisphere, for example, is now located nearly 1800 km from the geographic pole, somewhere in the Hudson Bay region of northern Canada. The other pole is located south of Australia. This means that compasses do not generally point to the true north. The discrepancy between the orientation of a compass and true north is known as the ***magnetic declination***.

Also, Earth's magnetic field lines are not parallel to the surface at all points. Near the magnetic poles, the field lines are almost perpendicular to the surface of the Earth. At the magnetic Equator, the magnetic field lines are almost parallel to the surface of the earth. The angle that the magnetic field makes with respect to the surface at any point is known as the ***angle of dip or inclination***.

We do not know exactly why Earth itself is a magnet. The configuration of Earth's magnetic field is like that of a strong bar magnet placed near the centre of Earth. But Earth is not a magnetized chunk of iron like a bar magnet. It is simply too hot for individual atoms to hold to a proper orientation. So the explanation must lie with electric currents deep in the interior. About 2000 km below the outer rocky mantle (which itself is almost 3000 km thick) lies the molten part that surrounds the solid centre. Most Earth scientists think that moving charges looping around within the molten part of Earth create the magnetic field. Some Earth scientists speculate that the electric currents are the result of convection currents from heat rising from the central core and that such convection currents combined with the rotational effects of the Earth produce Earth's magnetic field. Because of the Earth's great size, the speed of moving charges need only be about a millimetre per second to account for the field. A firmer explanation awaits more study.

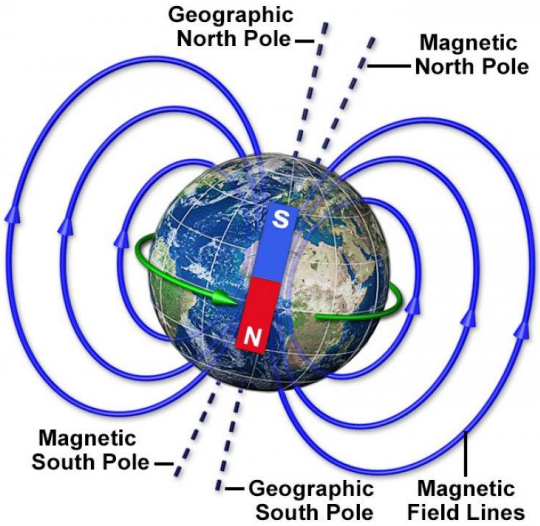


Figure A.1 Earth's magnetic poles versus geographic poles

Whatever the cause, the magnetic field of the Earth is not stable; it has wandered throughout geologic time. Evidence of this comes from the analysis of magnetic properties of rock strata. Iron atoms in a molten state are disoriented because of thermal motion, but a slight predominance of the iron atoms align with the magnetic field of the Earth. When cooling and solidification occurs, this predominance records the direction of the Earth's magnetic field in the resulting igneous rock. It is similar for sedimentary rocks, where magnetic domains in grains of iron that settle in sediments tend to align themselves with Earth's magnetic field and become locked into the rock that forms. The slight magnetism that results can be measured with sensitive instruments. As samples of rock are tested from different strata formed throughout geologic time, the magnetic field of the Earth for different periods can be charted. This evidence shows that there have been times when the magnetic field of the Earth has diminished to zero, followed by reversal of the poles. More than twenty reversals have taken place in the past 5 million years. The most recent occurred 700,000 years ago. Prior reversals happened 870,000 and 950,000 years ago. Studies of deep-sea sediments indicate that the field was virtually switched off for 10,000 to 20,000 years just over 1 million years ago. We cannot predict when the next reversal will occur because the reversal sequence is not regular. But there is a clue in recent measurements that show a decrease of more than 5% of the Earth's magnetic field strength in the last 100 years. If this change is maintained, we may well have another reversal within 2000 years.

The reversal of magnetic poles is not unique to the Earth. The Sun's magnetic field reverses regularly, with a period of 22 years. This 22-year magnetic cycle has been linked, through evidence in tree rings, to periods of drought on Earth. Interestingly enough, the long-known, 11-year sunspot cycle is just half the time during which the Sun gradually reverses its magnetic polarity.

Varying ion winds in Earth's atmosphere cause more rapid but much smaller fluctuations in Earth's magnetic field. Ions in this region are produced by the energetic interactions of solar ultraviolet rays and X-rays with atmospheric atoms. The motion of these ions produces a small but important part of Earth's magnetic field. Like the lower layers of air, the ionosphere is churned by winds. The variations in these winds are responsible for nearly all of the fast fluctuations in the Earth's magnetic field.

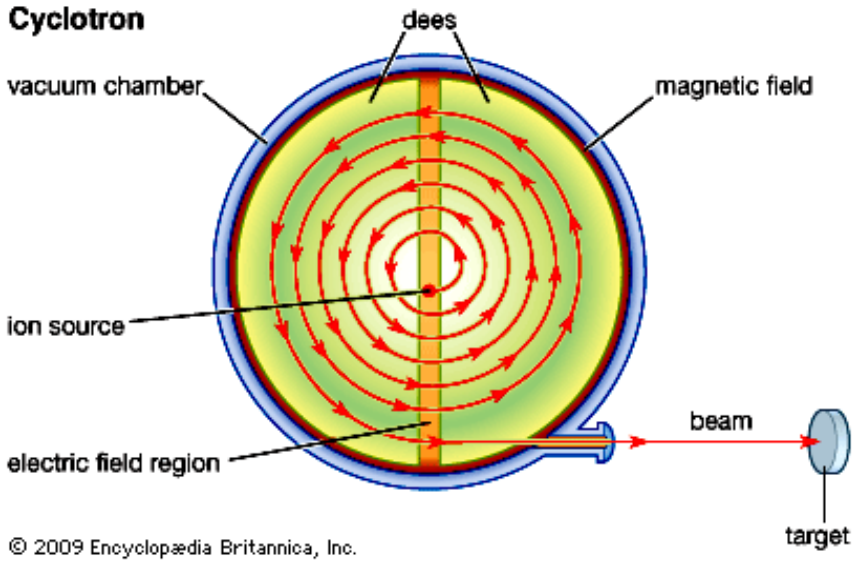
(Adapted from Paul G. Hewitt's Conceptual Physics)

https://www.conceptualacademy.com/course/conceptual-physics/248-earth%E2%80%99s-magnetic-field

APPENDIX B: The Cyclotron

A cyclotron is a device that can accelerate charged particles to very high speeds. The energetic particles produced are used to bombard atomic nuclei and thereby produce nuclear reactions of interest to researchers.

Both electric and magnetic forces play a key role in the cyclotron. The charges move inside two semi-circular containers D_1 and D_2, referred to as dees. A high frequency alternating potential difference is applied across the dees and a uniform magnetic field is directed perpendicularly to them. A positive ion released near the centre of the magnet in one dee moves in a semi-circular path and arrives back at the gap in a time interval of $T/2$ where T is the time needed to make one complete trip around the dees.



The frequency of the applied potential difference is adjusted so that the polarity of the dees is reversed at the same time interval during which the ion travels around one dee. The ion then accelerates across the gap from D_1 to D_2 and its kinetic energy increases by qV where V is the potential difference set up across the dees.

It now moves with a larger radius in D_2 due to the increase in speed. After $T/2$, it arrives at the gap again, coinciding with the reversing of the polarity of the potential difference set up across the dees. The ion thus accelerates again across the gap and the cycle repeats itself. Every time the ion passes the gap, it gains an energy of qV. By the time it exits the cyclotron, it would have garnered about 20 MeV.

Source	**Magnetic Flux Density (T)**
Atomic nucleus	10^{12}
Medical MRI unit	1.5
Bar magnet	10^{-2}
Surface of Sun	10^{-2}
Surface of Earth	5×10^{-5}
Inside human brain (due to nerve impulses)	10^{-13}

APPENDIX C: Charges moving in non-uniform magnetic fields

When charged particles move in a non-uniform magnetic field, the motion becomes complex. An example of such field is one where it is strong at the ends and weak in the middle as shown in Figure 1 below. This configuration is known as the magnetic bottle. A charged particle starting at one end spirals along the field lines until it reaches the other end, where it reverses its path and spirals back. The charged particle is therefore trapped within it.

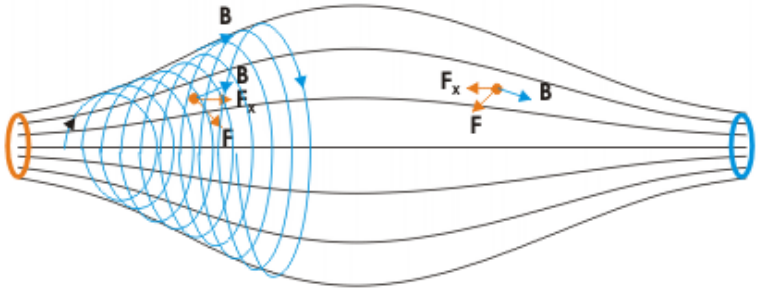


Figure 1 Magnetic bottle

Another example are the Van Allen radiation belts consisting of charged particles (mostly electrons and protons) surrounding the Earth in doughnut-shaped regions (Figure 2).

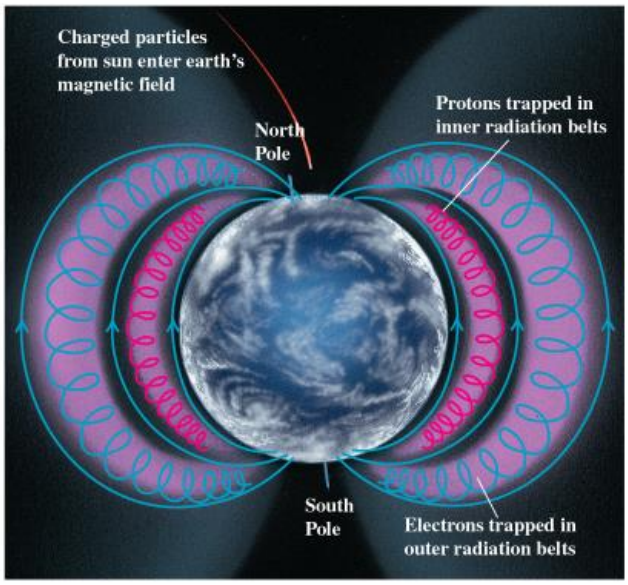


Figure 2 The Van Allen belts

Figure 3 The Aurora Borealis

Charged particles originate mainly from the sun (and other stars and heavenly bodies) and are commonly known as cosmic rays; they consist of protons, alpha particles, and other atomic nuclei, as well as some high-energy electrons. In any event, they travel through space at fantastic speeds and make up the cosmic radiation that is hazardous to astronauts. This radiation is intensified when the Sun is active and contributes added energetic particles. Cosmic rays are also a hazard to electronic instrumentation in space; impacts of heavily ionizing cosmic-ray nuclei can cause computer memory bits to "flip" or small microcircuits to fail. Fortunately for those of us on Earth's surface, most of these charged particles are deflected away by the magnetic field of the Earth.

Due to the Earth's non-uniform magnetic field, some of these particles are trapped and spiral around the field lines from pole to pole, making the Van Allen belts. When the particles are located near the poles, they sometimes collide with the atoms in the atmosphere, causing atoms to emit visible light. Such collisions are the origins of the beautiful Aurora Borealis (Northern Lights) and Aurora Australis (Southern Lights).

Chapter 15

ELECTROMAGNETIC INDUCTION

Electromagnetic induction effects are crucial in the operation of an electric guitar.

Electromagnetic Induction
2023
Syllabus 9749

Content

- Magnetic flux
- Laws of electromagnetic induction

Learning Outcomes

Candidates should be able to:

(a) define magnetic flux as the product of an area and the component of the magnetic flux density perpendicular to that area.

(b) recall and solve problems using $\phi = BA$.

(c) define magnetic flux linkage.

(d) infer from appropriate experiments on electromagnetic induction:
- i) that a changing magnetic flux can induce an e.m.f.,
- ii) that the direction of the induced e.m.f. opposes the change producing it,
- iii) the factors affecting the magnitude of the induced e.m.f.

(e) recall and solve problems using Faraday's law of electromagnetic induction and Lenz's law.

(f) explain simple applications of electromagnetic induction.

Table of Contents

15.1 Introduction 4
15.2 Magnetic Flux and Magnetic Flux Linkage 4
15.3 Induced e.m.f. in a Moving Conductor (Motional e.m.f.) 6
15.4 Laws of Electromagnetic Induction 9
15.4.1 Faraday's Experiment and Findings 9
15.4.2 Faraday's Law 10
15.4.3 Lenz's Law 10
15.5 Practical Applications of Electromagnetic Induction 13
15.5.1 Electric Generator (rotating coil in uniform magnetic field) 13
15.5.2 Speedometer of a Car 15
15.5.3 Ground Fault Interrupters 16
15.5.4 Electric Guitar 16
15.6 Eddy Currents 17
Appendix 1 A Faraday's Disc with area *A* rotating with a frequency *f* in a magnetic field *B* 18
Appendix 2 Volcanic Eruption as a Consequence of Eddy Currents in Action 19
Tutorial 15: Electromagnetic Induction 20
Electromagnetic Induction Summary 32

15.1 Introduction

In the previous chapter, we discussed the production of magnetic fields due to electric currents. As nature is often symmetric, this led scientists to suspect that the reverse (i.e. the production of electric currents due to magnetic fields) could also be possible. Experiments conducted by Michael Faraday and independently by Joseph Henry in 1831 showed that a changing magnetic field could indeed produce electric current in a circuit. These results led to a fundamental law known as Faraday's Law. In this chapter, we will discuss Faraday's Law and see how his experimental findings have helped set the foundation for many of modern world's technologies.

Figure 15.1.1
Michael Faraday (1791 – 1867),
British Physicist and Chemist

15.2 Magnetic Flux and Magnetic Flux Linkage

We have learnt that magnetic flux density, ***B***, is a measure of magnetic fields. Before we learn about electromagnetic induction, we need to understand what is meant by the term magnetic flux.

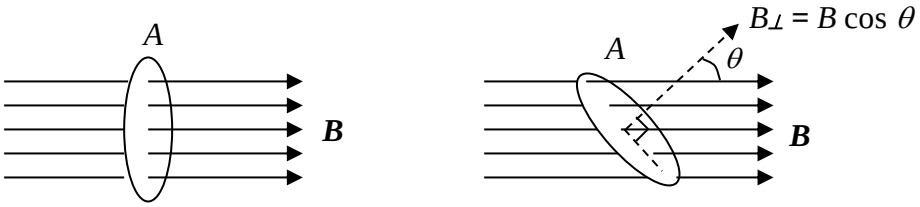
Definition of Magnetic Flux, ϕ:

The **magnetic flux** through a plane surface is the product of the area and the component of the magnetic flux density perpendicular to that area,

$$\phi = B_{\perp} A$$

where ϕ: magnetic flux through the plane surface
$B_{\perp}$: magnetic flux density perpendicular to the plane surface
A : area of the surface

The magnetic flux for two different cases is given below.



$\phi = B_{\perp} A = BA$ $\phi = B_{\perp} A = (B\cos\theta)A$

Figure 15.2.1 Magnetic flux.

The S.I. unit for magnetic flux, ϕ, is the weber (Wb).

Definition of weber (Wb):

One weber is the amount of magnetic flux through a surface if a magnetic field of flux density of one tesla exists perpendicularly to an area of 1 m^2.

1 Wb = 1 T x 1 m^2

In the previous chapter, we learnt that the S.I. unit for the magnetic flux density, B, is the tesla (T). An alternative unit for B is Wb m^{-2}.
Note that magnetic flux density, B, is a vector, but magnetic flux, ϕ, is a scalar.

It is common to encounter coils with multiple turns (each turn of the coil has the same area).

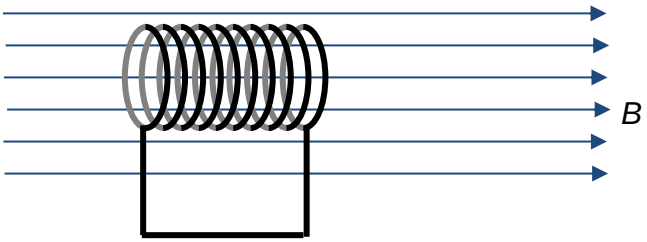


Figure 15.2.2 Magnetic flux linkage.

The **magnetic flux linkage** is defined as the product of the number of turns of the coil and the magnetic flux through each turn.

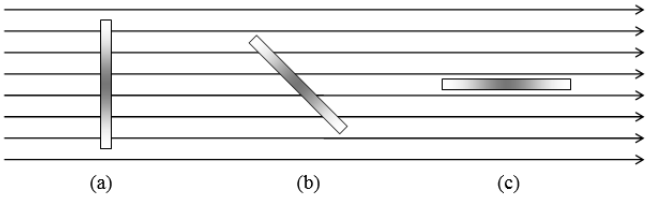
$$\phi_{linkage} = N\phi = NB_{\perp}A$$

The S.I. unit for magnetic flux linkage is weber turns (Wb turns).

Note: In many textbooks, there is no distinction between the symbols used for magnetic flux and magnetic flux linkage, both are usually represented by Φ, hence it is important for us to know clearly in each scenario whether Φ refers to magnetic flux or magnetic flux linkage.

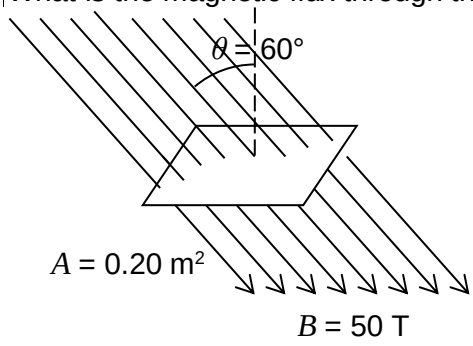
Example 1 Effect of orientation of magnetic field on magnetic flux

Illustrated is a loop placed at different orientation in a uniform magnetic field (viewed from the side).
Which orientation gives
1) zero magnetic flux,
2) non-zero but not maximum magnetic flux and
3) maximum magnetic flux?



Example 2 Determination of magnetic flux

A flat rectangular loop of wire is placed in a uniform magnetic field which is directed at an angle to the normal of the loop (for clarity, only one row of arrows is drawn to represent the uniform magnetic field). What is the magnetic flux through the loop?



15.3 Induced e.m.f. in a Moving Conductor (Motional e.m.f.)

In the previous topic on Electromagnetism, we learnt that the magnetic force experienced by a current-carrying conductor in a magnetic field is given by $F_B = BIL\sin\theta$, where θ is the angle between *B* and *I*.

Similarly, the magnetic force acting on a moving charge *Q* with velocity *v* in a magnetic field of flux density *B* is given by $F_B = BQv\sin\theta$, where θ is the angle between *B* and *v*.

Next, we shall look at the effects of a moving conductor in a magnetic field.

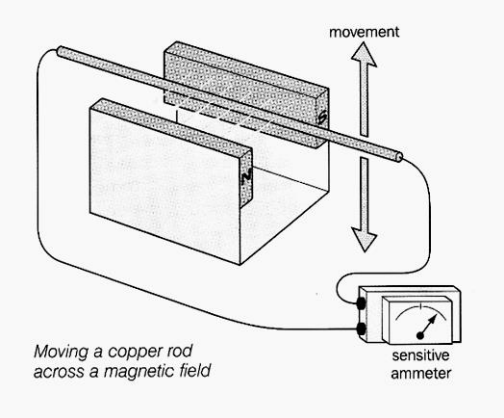


Figure 15.3.1 Conductor cutting through a magnetic field.

Figure 15.3.1 shows a copper rod that moves between the poles of a magnet, while being connected in series to a sensitive ammeter.

- When the copper rod is moved downwards, the pointer on the meter flicks to a particular direction suggesting that an e.m.f. has been induced.
- When the rod is moved upwards, the meter flicks to the opposite direction, suggesting that an e.m.f. has now been induced in the opposite direction.
- If the rod is held stationary in the field or made to move *along* the magnetic field lines, there is no induced e.m.f.
- The **faster** the copper rod is moved, the **larger** the current.

It appears that ***an e.m.f. is induced in a moving conductor if it is moving through a magnetic field.*** **The *magnitude of this induced e.m.f. is proportional to the speed.***

We shall now derive an expression for the induced e.m.f. for a moving conductor. Consider a straight conductor or moving rod of length *L* moving perpendicularly at a constant velocity *v* under the influence of some external agent through a uniform magnetic field *B* directed into the page as shown in Figure 15.3.2.

Scan the QR code to see 3-D animation of motional e.m.f.
(https://youtu.be/WXGoNw0TOqs)

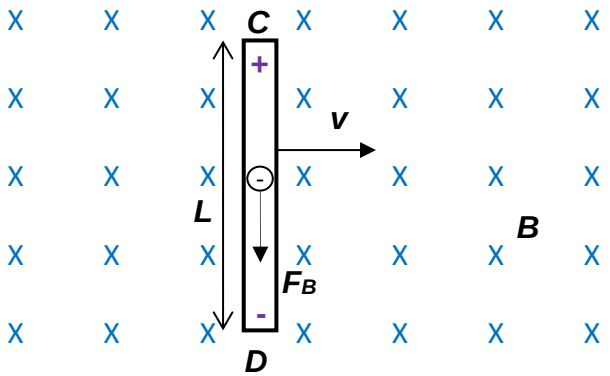


Figure 15.3.2 Isolated moving rod. The crosses indicate that the magnetic field is pointing into the page.

The electrons in the conductor experience a magnetic force $F_B = BQv$ that is directed along the length *L*, perpendicular to both *v* and *B,* where *Q* is the charge of the electron. Under the influence of this force, the electrons move to the lower end of the conductor and accumulate there, leaving a net positive charge at the upper end. As a result of this charge separation, an electric field *E* is produced inside the conductor.

The electrons accumulate at *D*, leaving a net positive charge at *C*, until the downward magnetic force $F_B = BQv$ is balanced by the upward electric force $F_E = QE$. At this point, the electrons stop moving.

The condition for equilibrium requires that

$$BQv = QE$$
$$E = Bv$$

In the topic on electric fields, you have learned that the electric field between two charged parallel plates is given by $E = \Delta V / L$, where ΔV is the potential difference between the plates and *L* is the distance between the plates. Thus, the electric field produced in the conductor in Figure 15.3.2 (once the electrons stop moving and *E* is constant) is related to the **potential difference across the ends of the conductor** by

$$V_{CD} = EL = BLv$$

where the upper end ***C* is at a higher electric potential than the lower end *D***. Thus, from the above equation, we can see that a potential difference is maintained between the ends of the conductor as long as the conductor continues to move through the uniform magnetic field. If the direction of the motion is reversed, the polarity of the potential difference is also reversed.

In the scenario above, **only an e.m.f. is induced across the rod.** As the circuit is not complete or closed, there is *no induced current* flowing in the rod.

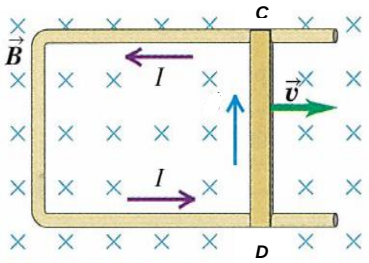


Figure 15.3.3 Rod connected to stationary conductor.

Now suppose the moving rod slides along a stationary U-shaped conductor, forming a complete circuit. Using the knowledge from Figure 15.3.2, the charges move in the moving part CD, causing a current to flow in an anti-clockwise direction as shown in Figure 15.3.3. We call this e.m.f. a motional electromotive force, denoted by ε. The magnitude of this e.m.f. is given by the rate of change of magnetic flux, Φ (see discussion on Faraday's Law in section 15.4.2)

$$\varepsilon = \frac{d\phi}{dt} = B\frac{dA}{dt} = BLv$$

If the total resistance of the U-shaped conductor and the sliding rod is *R*, the induced current I in the circuit is given by $BLv = IR$.

(i) Factors affecting magnitude of induced e.m.f.

From the formula $\varepsilon = BLv$, we can easily infer that increasing one or more of the following factors will increase the magnitude of the induced e.m.f. .

- The magnetic field strength *B*;
- The length of the moving conductor within the magnetic field *L*;
- The speed at which the conductor is moving *v*.

Conversely, decreasing one or more of these factors will lead to a decreased magnitude of the induced e.m.f..

(ii) Direction of motional e.m.f.

The direction of motional e.m.f. (and induced current, if any) can be predicted by **Fleming's Right-Hand Rule**.

Often we are asked to state the polarity of the induced e.m.f., or which end of the conductor is at the higher/lower potential. To answer that question correctly, we must keep in mind that when the Fleming's Right-hand rule is used, we are ***modelling the conductor as an e.m.f. source****, with its two ends acting like the two terminals of a battery.*

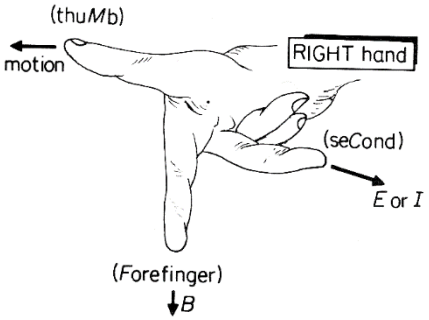
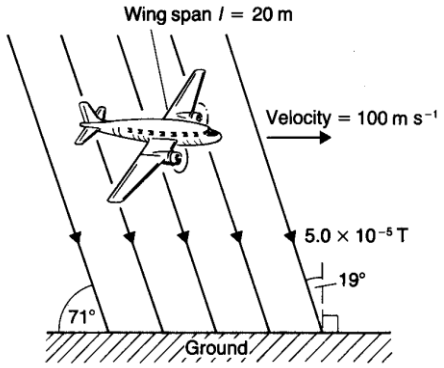


Figure 15.3.4 Fleming's Right-Hand Rule.

Example 3 Magnitude and polarity of induced e.m.f.

An airplane is travelling at 100 m s^{-1} in a direction which is horizontal. Calculate the e.m.f. induced between the tips of its wings, which have a span of 20 m. Take the Earth's magnetic flux density to be 5.0 x 10^{-5} T and the angle of dip at the place to be 71°. Indicate which tip is at a higher potential with a "+".



15.4 Laws of Electromagnetic Induction

15.4.1 Faraday's Experiment and Findings

In the last section we saw that relative motion between a magnetic field and a conductor can induce an e.m.f. in the conductor. An experiment carried out by Michael Faraday in 1825 showed how the same effect can be achieved with no apparent relative motion. Figure 15.4.1 shows the setup that led to his discovery.

A primary coil is connected to a battery and a switch. The coil is wrapped around a wooden core, and a current in the coil produces a magnetic field when the switch is closed.

A secondary coil also is wrapped around the wooden core and is connected to a sensitive galvanometer. No battery is connected in the secondary circuit and the secondary coil is not connected to the primary coil.

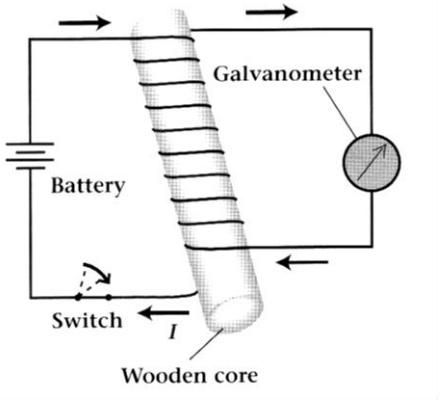


Figure 15.4.1 Faraday's experiment.

Faraday observed that:

- At the instant the switch is closed, the galvanometer needle deflects in one direction and returns to zero.
- When there is a steady current in the primary coil, the galvanometer always shows null deflection, no matter how large the current is.
- At the instant the switch is opened, the galvanometer needle deflects in the opposite direction and returns to zero.

In order to explain these observations, Faraday invoked the concept of changing magnetic flux linkage between the coils. His given explanation was:

(1) When the switch is closed, the **magnetic field produced by the current in the primary circuit** increases from zero to a constant value over some finite time. The **secondary coil experiences this changing magnetic flux linkage produced by the primary circuit** and the induced current only exists for a short time while the magnetic flux linkage through the secondary coil is increasing.

(2) Once the magnetic field produced by the primary coil reaches a steady value, the **magnetic flux linkage through the secondary coil is constant** and there is no current in the secondary coil.

(3) When the switch is opened, the opposite occurs – the **magnetic field produced by the current in the primary circuit decreases** from the constant value to zero over some finite time. Again, the **secondary coil experiences this changing magnetic flux linkage produced by the primary circuit** and the induced current flows [in the opposite direction to that in (1)] while the magnetic flux linkage through the secondary coil is decreasing.

In effect, when induced current flows in the secondary coil, the secondary circuit behaves as though a source of e.m.f. were connected to it. Hence, Faraday concluded that **there will be an induced e.m.f. in the secondary circuit if it experiences a changing magnetic flux linkage.**

With further experiments, he concluded that the **induced e.m.f. is directly proportional to the rate of change of magnetic flux linkage,** which led to the Law of electromagnetic induction.

15.4.2 Faraday's Law

The ***magnitude*** of the induced e.m.f. ε is ***directly proportional*** to the ***rate of change*** **of magnetic flux linkage** or **rate of cutting of magnetic flux.**

$$\varepsilon \propto \frac{d\phi}{dt}$$

The S.I unit for induced e.m.f., ε, is the volt (V). An alternative unit for induced e.m.f. is Wb turns s^{-1}.

15.4.3 Lenz's Law

The Russian physicist Heinrich Lenz (1804 – 1865) made a discovery that allows the polarity of the induced e.m.f. to be determined. This discovery led to Lenz's Law.

Lenz's Law states that the **polarity of the induced e.m.f.** is such that it tends to produce a current that **creates a magnetic field so as to <u>oppose the change in magnetic flux</u>.**

Consider the following scenario:

If the bar magnet is moved rightwards towards the stationary metal loop (Figure 15.4.2), the flux through the loop increases with time. The increasing flux induces an e.m.f. in the metal loop (Faraday's Law), and since the loop forms a closed circuit, a current is produced. This induced current produces its own magnetic field. By Lenz's law, the direction of the induced current must be such as to produce a magnetic field directed in the **<u>opposite direction</u>** to the **<u>strengthening external field.</u>**

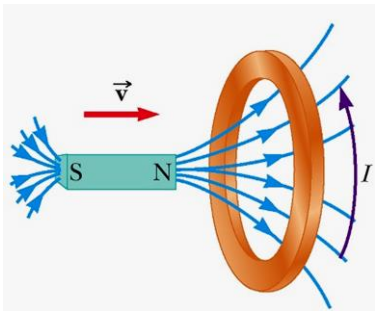
<u>Action</u>	<u>Effect</u>
	*The magnetic flux through the loop is **increasing** with time. By Faradays' law, e.m.f is induced in the closed circuit. By Lenz's law, induced current flows in the metal loop in anti-clockwise direction (when viewed from left).*

Figure 15.4.2 Magnet approaching loop.

If the magnet is moved to the left (Figure 15.4.3), the flux through the loop decreases with time. The decreasing flux again induces an e.m.f. in the metal loop (Faraday's Law). However, this time the direction of the induced current is such as to produce a magnetic field that is directed in **<u>the same direction</u>** as the **<u>weakening external field.</u>**

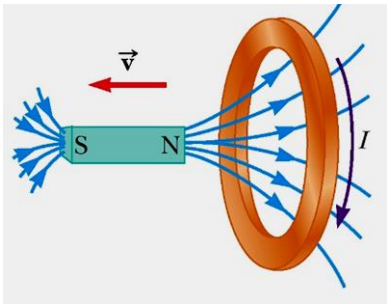
<u>Action</u>	<u>Effect</u>
	*The magnetic flux through the loop is **decreasing** with time. By Faradays' law, e.m.f is induced in the closed circuit. By Lenz's law, induced current flows in the metal loop in clockwise direction (when viewed from left).*

Figure 15.4.3 Magnet moving away from loop.

Notice also that the interaction of the two magnetic fields (produced by the magnet and the loop, respectively) produces magnetic forces. Note that **the direction of the induced current is such that the magnet experiences a repulsion force when approaching the loop and an attractive force when moving away from the loop.**

This leads to the important insight that **Lenz's Law is merely a re-statement of the law of conservation of energy**. The direction of induced current is always such as to (produce a magnetic force to) oppose the motion of the magnet (which was the cause of the electromagnetic induction). Since the motion of the magnet is opposed, work must be done in order to maintain the motion of the magnet. This mechanical work enables the conversion of energy into electrical energy. This is further illustrated in Figure 15.4.4.

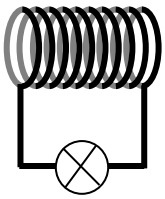
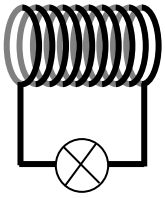
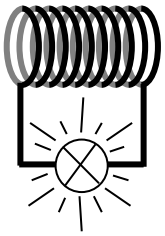
I external force → S N 	An initial force is applied to the magnet to start its motion for a very short duration.
II S N → u 	The magnet starts to move with an initial speed u towards the coil. Hence, the magnet has a certain amount of kinetic energy.
III S N → $v < u$ 	As the magnet approaches the coil, an induced e.m.f. is set up in the coil. As the coil is a part of a closed circuit, an induced current will flow in a way that tends to oppose the increase in magnetic flux. Hence, the magnet slows down while bulb lights up. Thus, the kinetic energy of the magnet is first converted into electrical energy in the circuit, followed by the electrical energy being converted into light and thermal energy in the bulb.

Figure 15.4.4 Illustration of Lenz's Law in view of the law of conservation of energy.

In summary,

(1) **Faraday's Law** enables us to determine the **magnitude** of induced e.m.f. (applicable to both open and closed circuits).
(2) **Lenz's Law** allows us to determine the **polarity** of induced e.m.f. or **direction of induced current** (applicable only to closed circuits). If circuit is open, we usually think in terms of what would happen if it was closed and hence find the direction of the induced current)

Combining both Faraday's and Lenz's Laws, we get:

$$\varepsilon = -\frac{d\phi}{dt}$$

The **negative sign** in the above equation is a **consequence of Lenz's law**.

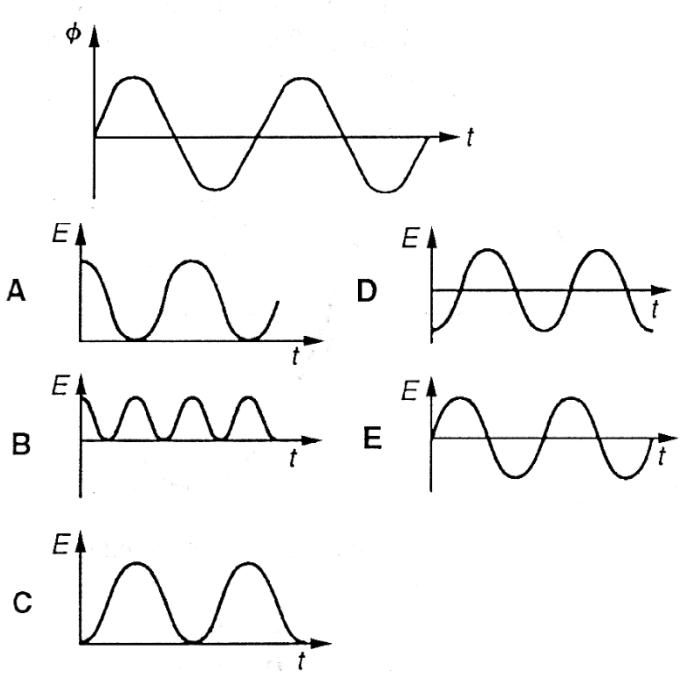
Example 4 Determination of average induced e.m.f.

A flat circular coil of diameter 30 mm has 500 turns and is situated so that the plane of the coil is perpendicular to a uniform magnetic field of flux density 20 mT. The flux density is reduced to zero and then increased to 20 mT in the opposite direction at a constant rate. The time taken for the whole operation is 60 ms.

What is the average value of the e.m.f. induced in the coil?

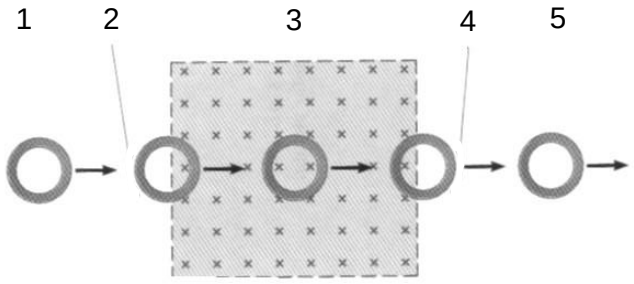
Example 5 Graphs of varying magnetic flux and e.m.f. induced

The magnetic flux ϕ through a coil varies with time as shown in the diagram. Which graph best represents the variation with time of the induced e.m.f. E in the coil?



Example 6 Application of Lenz's Law

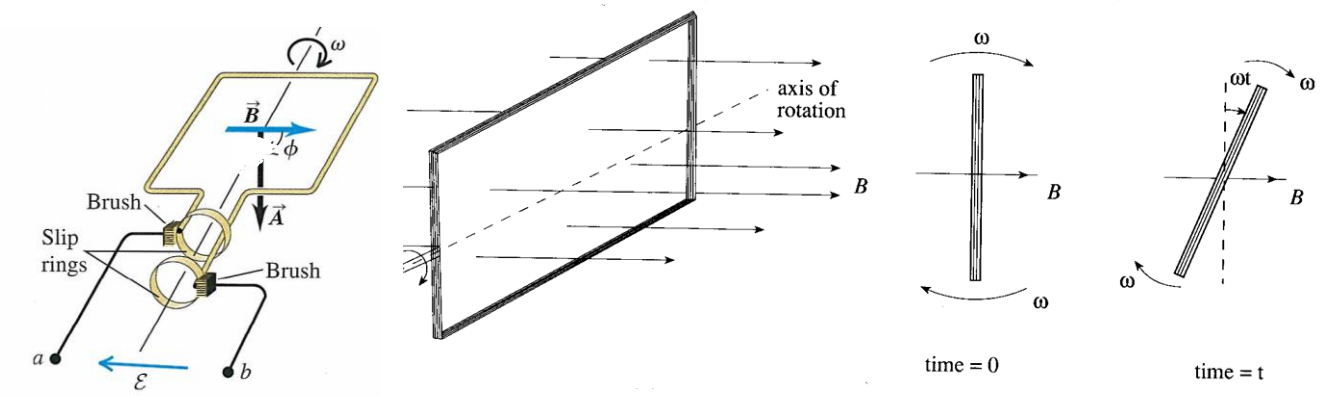
In the figure below, there is a constant magnetic field in a rectangular region of space. This field is directed perpendicularly into the plane of the paper. Outside this region, there is no magnetic field. A copper ring slides through the region, from position 1 to position 5. For each of the five positions, determine if an induced current exists in the ring and, if so, find the direction of the current.



15.5 Practical Applications of Electromagnetic Induction

15.5.1 Electric Generator (rotating coil in uniform magnetic field)

A coil of N turns, with area A, rotates at a constant angular velocity ω in a uniform magnetic field B.



(a) (b) (c) (d)

Figure 15.5.1 Rotating coil in a magnetic field, generating a current.

As the coil rotates in the magnetic field, it experiences a ***changing magnetic flux linkage*** as the component of the magnetic field perpendicular to the plane of the coil keeps changing. By Faraday's law, an e.m.f. is induced in the coil.

<u>Derivation of an expression for the induced e.m.f.</u>

1. Magnetic flux linkage through the coil

$$\phi = NB_{\perp}A = NBA\cos\theta = NBA\cos\omega t$$

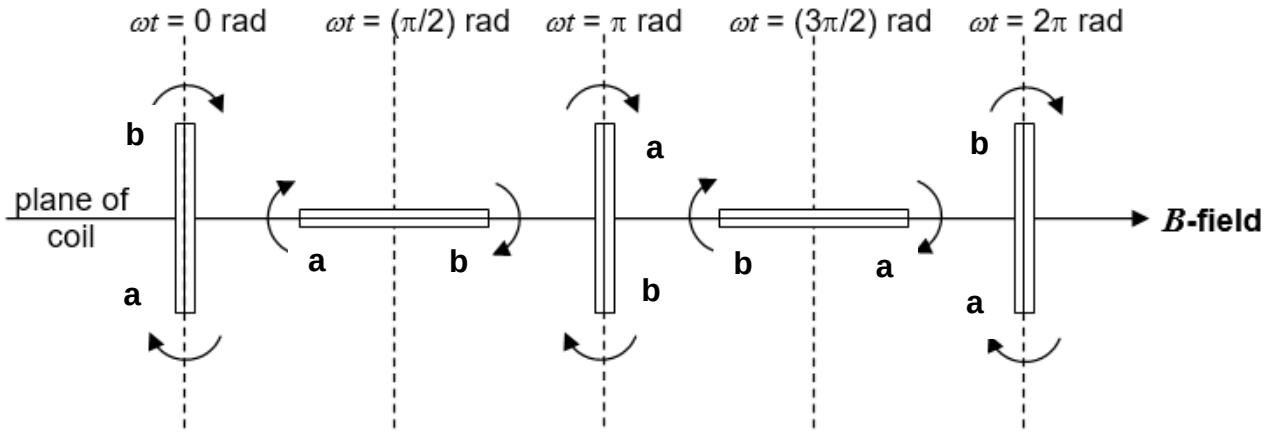
2. Induced e.m.f.

$$\varepsilon = -\frac{d\phi}{dt} = -\frac{d}{dt}(NBA\cos\omega t) = -NBA\frac{d}{dt}(\cos\omega t)$$

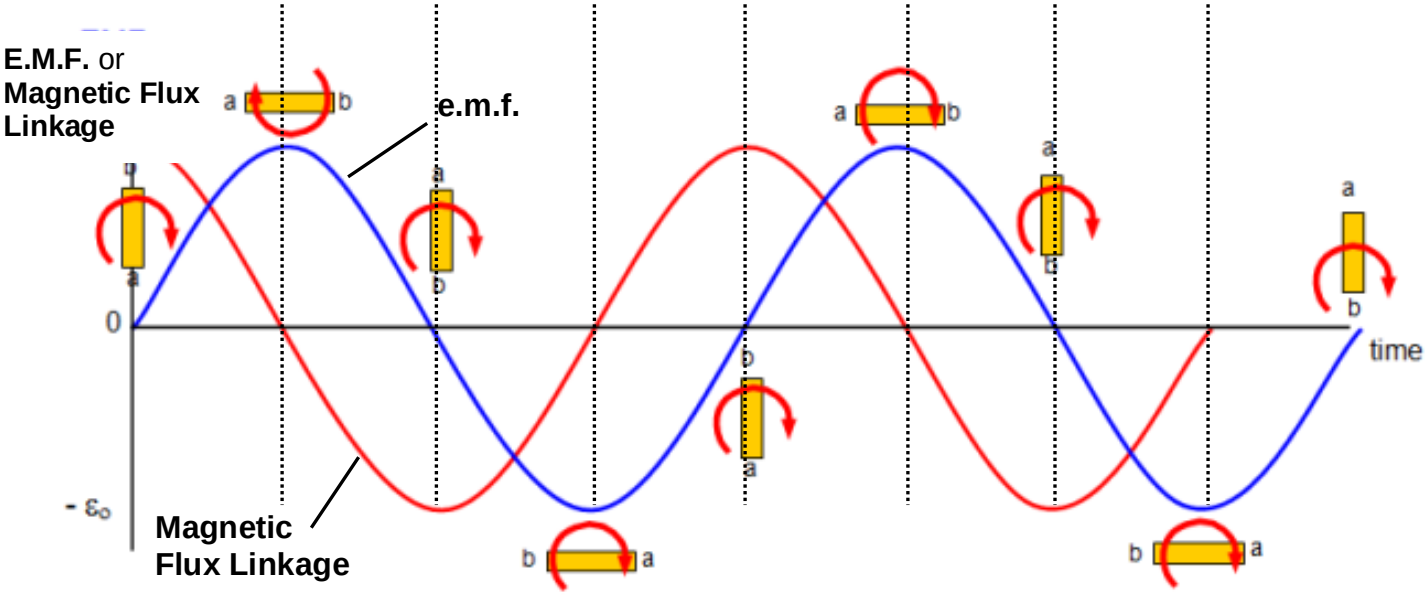
$$\varepsilon = NBA\omega\sin\omega t$$

Scan the QR code to see 3-D animation of AC generator (https://youtu.be/YkHhFho6L2Y)

Hence the induced e.m.f. is dependent on *N, B, A* and *ω*. **By Lenz's law, the direction of the induced e.m.f. is such that it sets up an induced current to oppose the rotation of the coil.**



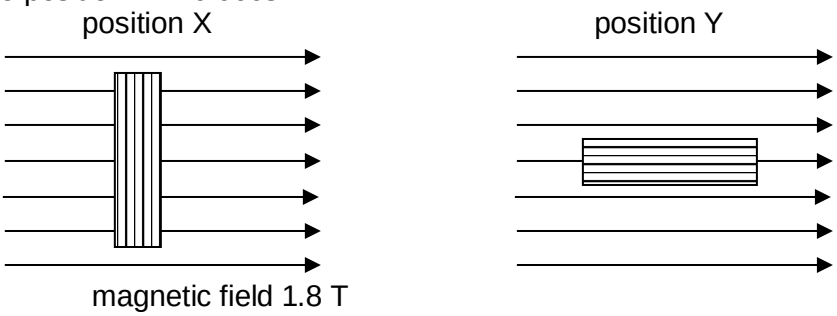
Graph of the magnetic flux linkage versus the time and Graph of induced e.m.f. versus time.



Note that the **magnitude of the induced e.m.f. is largest** when the **magnetic flux linkage is zero**. Conversely, when the magnitude of the magnetic flux linkage is largest, there is no induced e.m.f. in the coil.

Example 7 Determination of the e.m.f. induced in a rotating coil (N07/I/31)

A circular coil of diameter 0.020 m has 3000 turns. It is rotated in a uniform magnetic field of 1.8 T from position X to position Y in 0.060s.



(a) What is the average e.m.f. induced in the coil during the rotation X to Y?
(b) What is the average e.m.f. induced in the coil for one complete revolution, assuming the coil is rotating at a constant angular velocity?

15.5.2 Speedometer of a Car

A permanent magnet is attached to the wheel of the car. A sensor consisting of a coil of wire is connected to the frame of the car. When the car moves, the magnet rotates with the wheel. As the magnet passes over the sensor, the magnetic flux in the coil rises and then falls. This induces an e.m.f. and a current in the coil. The induced current, in the form of pulses, is sent to an electronic device which counts the number of turns performed by the wheel in each second. With this information, the speed of the car can be calculated if the radius of the wheel is known and the speed can thus be displayed on the speedometer.

Figure 15.5.2 Speedometer.

15.5.3 Ground Fault Interrupters

The ground fault interrupter is a safety device that protects people against an electric shock when they touch appliances and power tools. In the diagram below, wire 1 leads from the wall outlet to the appliance to be protected and wire 2 leads from the appliance back to the wall outlet. An iron ring surrounds the two wires to confine the magnetic field set up by each wire. A sensing coil, which can activate a circuit breaker when changes in magnetic flux occur, is wrapped around part of the iron ring.

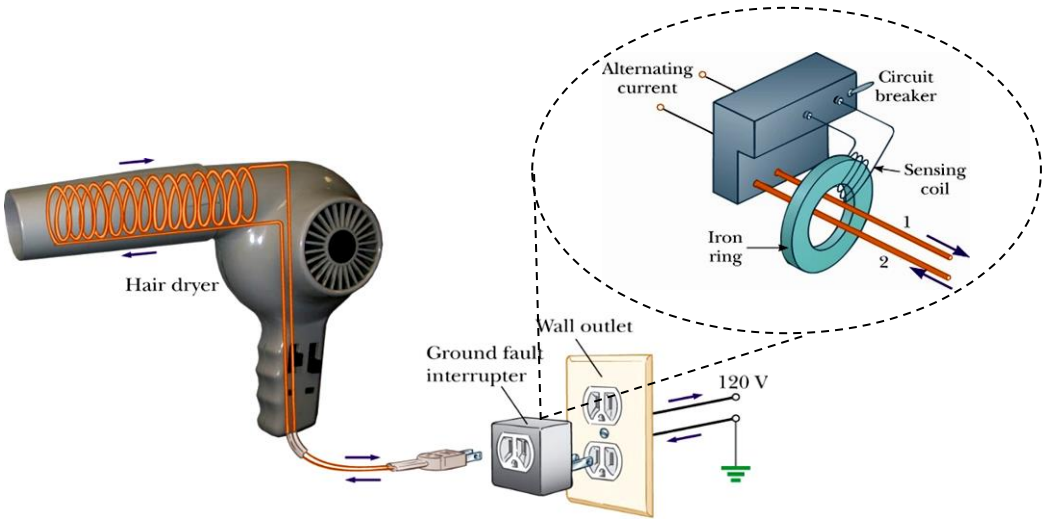


Figure 15.5.3 How the ground fault interrupter works.

Since the currents in the wires are in opposite directions, the net magnetic field through the sensing coil due to the currents is zero. However, if a short circuit occurs in the appliance so that there is no returning current, the net magnetic field through the sensing coil is no longer zero. This can happen for example if one of the wires loses its insulation, providing a path through the person to ground if the person happens to be touching the appliance. As the current is alternating, the magnetic flux through the sensing coil changes with time, thus inducing an e.m.f. in the coil. This induced e.m.f. will trigger the circuit breaker, stopping the current quickly in about a millisecond before it reaches a level that might be harmful to the person using the appliance.

15.5.4 Electric Guitar

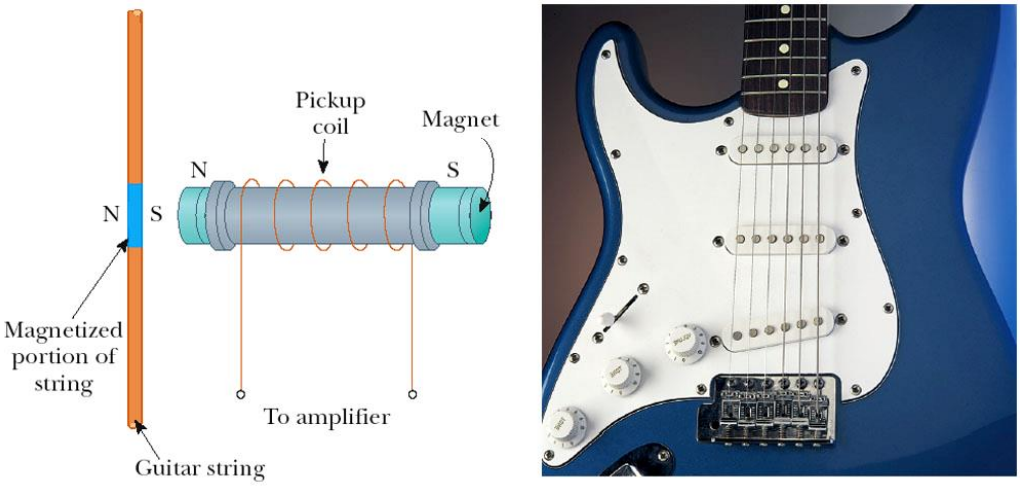


Figure 15.5.4 How the electric guitar works.

The pickup coil is placed near the vibrating guitar string, which is made of a metal that can be magnetized. The permanent magnet inside the coil magnetizes the portion of the string nearest the coil. When the guitar string vibrates at some frequency, its magnetized segment produces a changing magnetic flux through the pickup coil. The changing magnetic flux induces an e.m.f. in the coil. This induced e.m.f. or voltage is fed to an amplifier, the output of which is sent to the loudspeakers.

15.6 Eddy Currents

As we have seen, e.m.f. and current can be induced in a circuit by a changing magnetic flux. This also results in eddy currents being induced in bulk pieces of metal moving through a magnetic field.

Consider a flat copper or aluminium plate attached to a rigid rod which swings back and forth through a magnetic field.

As the plate enters the field, the increasing magnetic flux induces an e.m.f. in the plate, which in turn causes the free electrons in the plate to move, producing swirling currents (eddy currents). Similarly, as the plate leaves the field, the decreasing magnetic flux causes eddy currents in the opposite direction. As the currents flow, the plate will heat up; its internal energy will increase. By the principle of conservation of energy, this means that the mechanical energy in the oscillating plate must reduce. This loss in mechanical energy will eventually bring the plate to rest.

Alternatively, according to Lenz's law, the eddy currents give rise to repulsive forces that opposes the motion of the plate. This retarding magnetic force will eventually bring the plate to a rest.

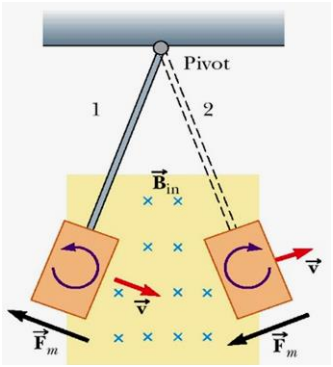


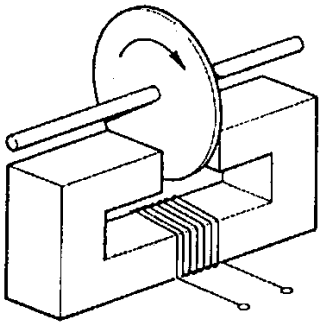
Figure 15.6.1 Eddy currents.

In order to reduce the amount of energy transformed from mechanical energy into internal energy, slots are cut into the plate, reducing the eddy currents and the corresponding retarding force. Also, conducting parts in devices are often laminated, which means they are built up in thin layers separated by a non-conducting material. This layered structure greatly increases the resistance of eddy current paths and effectively confines the currents to individual layers. Such a laminated structure is used in transformer cores.

However, eddy currents are often used in the braking system of many subway and rapid-transit cars. An electromagnet attached to the train is positioned near the steel rails. The braking action occurs when a large current is passed through the electromagnet. The motion of the magnet relative to the rails induces eddy currents in the rails, and the direction of these currents produces a drag force on the moving train. As the train slows down, the eddy currents decrease, hence the braking effect is quite smooth.

Example 8 Magnetic Braking

A copper disc spins freely between the poles of an unconnected electromagnet as shown. Describe and explain what will happen to the speed of rotation of the disc when a direct current flows in the electromagnet.



Answer:

As the copper disc rotates, different parts of the copper disc pass through the magnetic field due to the electromagnet.

Regions of the copper disc which are entering the electromagnet's magnetic field experience strengthening flux, leading to induced e.m.f. proportional to the rate of flux change (by Faraday's law). The induced e.m.f. causes the free electrons in the disc to move, producing eddy currents. The disc will heat up, and by the principle of conservation of energy, it will slow down and eventually come to rest. The polarity of the induced e.m.f. and current is such that it creates a magnetic flux in a direction opposite to the electromagnet's magnetic field, creating a repelling pole (by Lenz's law).

Regions of the copper disc leaving the electromagnet's magnetic field experience weakening flux, leading to an induced e.m.f. proportional to the rate of flux change (by Faraday's Law). The polarity of the induced e.m.f. and current is such that it creates a magnetic flux in the same direction as the electromagnet's magnetic field, creating an attracting pole (by Lenz's law).

For both regions, the direction of the magnetic force is always opposite to the direction of motion. The magnetic force retards the motion and slows down the spinning.

Appendix 1 A Faraday's Disc with area *A* rotating with a frequency *f* in a magnetic field *B*

The Faraday's disc, an example shown in Figure 1A, is the first homopolar generator. A homopolar generator is a DC electrical generator comprising an electrically conductive disc or cylinder rotating in a plane perpendicular to a uniform static magnetic field.

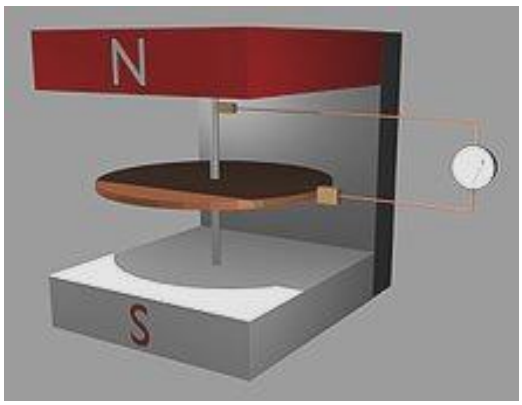


Figure 1A

Derivation of expression for induced emf
(method one)
Consider an elemental strip
Magnetic flux it cuts in time *t*, where *A* is area of sector mapped out in time *t*,

$$\phi = B_{\perp}A = B\left(\frac{1}{2}R^2\theta\right)$$

The induced e.m.f. is given by the rate at which magnetic flux is cut by the conductor. The magnitude is

$$\mathcal{E} = \frac{d\phi}{dt} = \frac{d}{dt}\left(\frac{1}{2}BR^2\theta\right) = \frac{1}{2}BR^2\frac{d\theta}{dt} = \frac{1}{2}BR^2\omega$$

For a complete revolution, since $\omega = 2\pi f$ and $A = \pi R^2$,

$$\mathcal{E} = \frac{1}{2}BR^2\omega = BAf$$

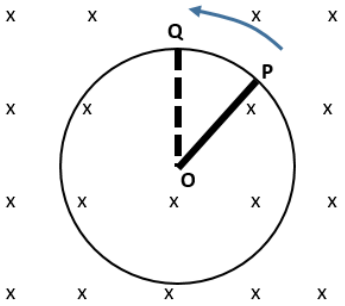


Figure 1B
(Plan view of disc in Figure 1A)

(method two)

Consider a free electron on the line OP. As the disc rotates,
the electron moves in an anticlockwise circle resulting in a clockwise current in the magnetic field. As such, the electron experiences a magnetic force F_B directed towards the outer rim of the disc, pushing the electron outwards.

The accumulation of electrons on the outer rim of the disc led to a potential difference being set up between the center O and the outer rim of the disc. This led to an electric force F_E on those electrons still in the disc.

At steady state, for any electron still in the disc, the electric force on it is balanced by the magnetic force,

$$F_E = F_B$$

$$-q\frac{dV}{dr} = Bqv = Bq\omega r$$

$$-\int_{\xi}^{0} dV = \int_{0}^{R} B\omega r\, dr$$

$$\xi = \frac{1}{2}B\omega R^2$$

Appendix 2 Volcanic Eruption as a Consequence of Eddy Currents in Action

(Extracted from Young and Freedman, University Physics, 12th edition)

A particularly dramatic example of eddy currents in action is Jupiter's moon Io, which is slightly larger than the earth's moon. Io moves rapidly through Jupiter's intense magnetic field (Figure 2A), and this sets up strong eddy currents within Io's interior. These currents dissipate energy at a rate of 10^{12} W, equivalent to setting off a one kilo-ton nuclear weapon inside Io every 4 seconds! This dissipated energy helps to keep Io's interior hot and so helps to cause volcanic eruptions on its surface, like those in Figure B. (However, note that gravitational effects from Jupiter cause even more heating.)

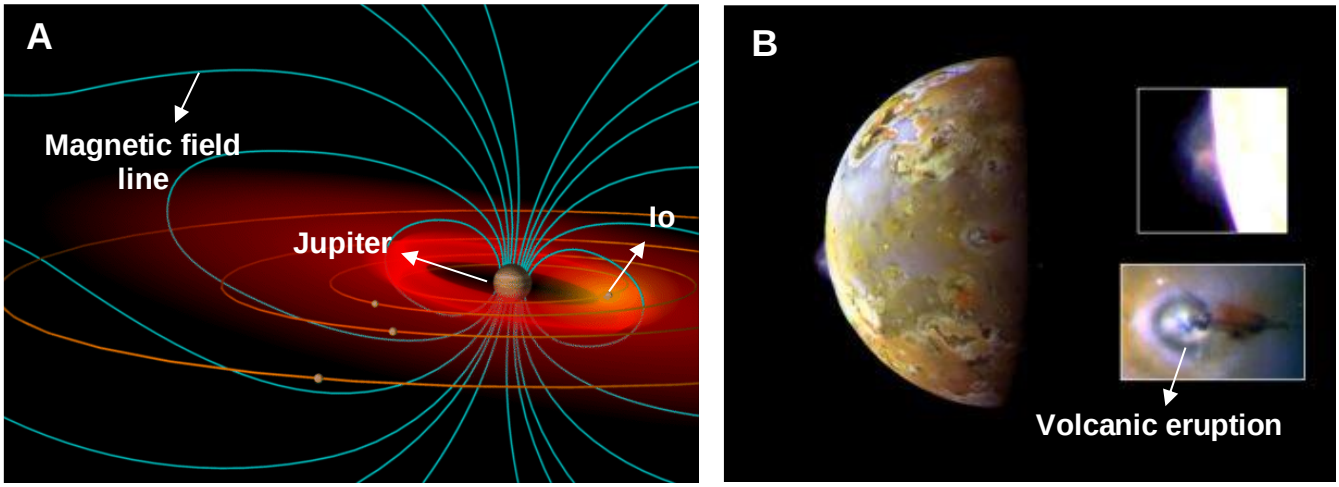


Figure 2 **A** Powerful magnetic field lines of Jupiter induce eddy currents in its moon Io, as Io moves in its orbit around Jupiter.
B Close-up view of a volcanic eruption, triggered in part by eddy-current heating.

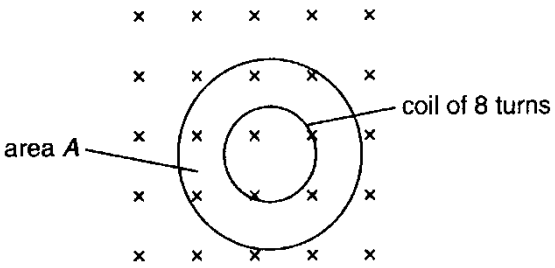
Tutorial 15: Electromagnetic Induction

Self-Practice Questions:

S1 [2018, H2PH, P3 Q4a]

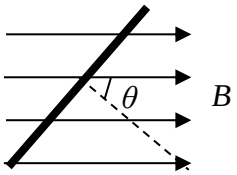
(i) Define *magnetic flux density*. [3]
(ii) State what is meant by *magnetic flux linkage*.[2]

S2 [J99/I/20] A uniform magnetic field of flux density *B* passes normally through a plane of area *A*. In this plane lies a coil of eight turns of wire, each of area ¼ *A*. What is the magnetic flux linkage for the coil?



A ¼ *BA*
B *BA*
C 2 *BA*
D 8 *BA*

S3 A square loop 2.00 m on a side is placed in a magnetic field of magnitude 0.300 T. Given that the field makes an angle of $\theta = 50.0^{o}$ with the normal (dotted line) to the plane of the loop, calculate the magnetic flux through the loop.



S4 A square-single turn coil 0.20 m on a side is placed with its plane perpendicular to a uniform magnetic field. An e.m.f. of 18 mV is induced in the coil winding when the area of the coil decreases at a rate of 0.10 $m^2\ s^{-1}$. What is the magnitude of the magnetic flux density of the magnetic field?

S5 [J99/I/20]

Diagram 1 shows an aluminium rod, moving at right angles to a uniform magnetic field. Diagram 2 shows the variation with time *t* of the distance *s* of rod from O.

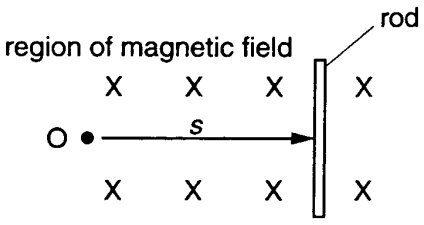


diagram 1

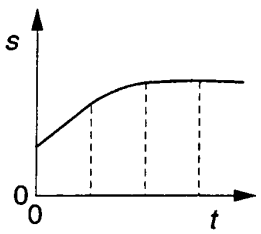
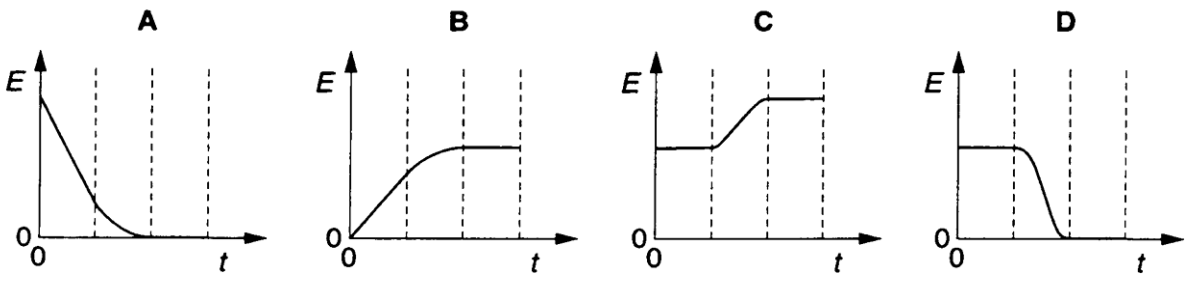
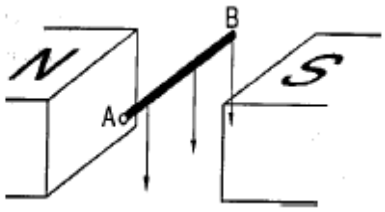


diagram 2

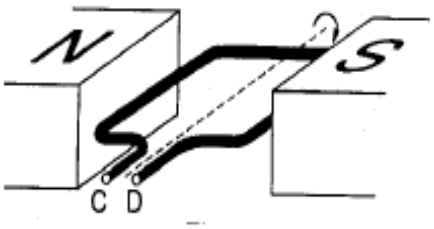
Which graph best shows the variation with *t* of the e.m.f. *E* induced in the rod?



S6 N94/III/4: A straight wire AB is moved across a magnetic field, as shown.

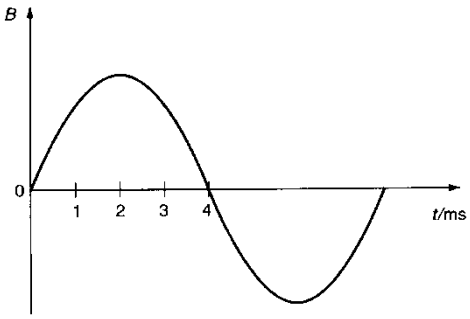
a) 

State three factors that determine the value of the induced e.m.f. measured between A and B.

b) 

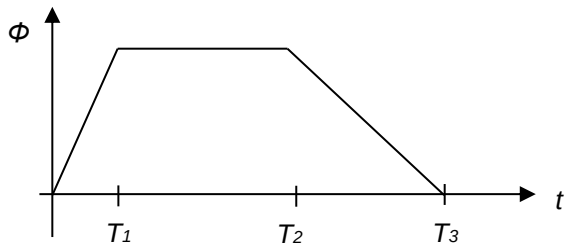
The wire AB is then replaced by a single loop of wire which is rotated with constant angular velocity in the magnetic field. Sketch a graph of the e.m.f. induced across CD, the two ends of the loop, against time t for one revolution. Assume that $t = 0$ for the coil in the position shown, that $t = T$ after one complete revolution and that the magnetic field within which the coil rotates is uniform. Mark the values $t = 0$ and $t = T$ on the time axis of your graph.

S7 [N97/I/19] A loop is subjected to a changing magnetic field and an e.m.f. is induced in it. The flux density B of the field varies with time t as shown. At which value of t is the magnitude of the e.m.f. induced a maximum?

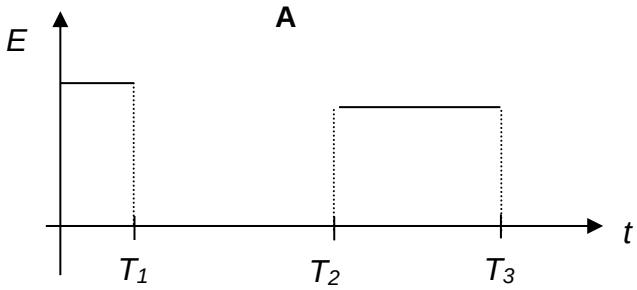


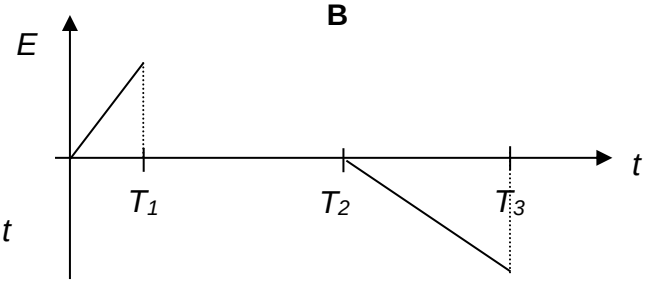
A 1 ms
B 2 ms
C 3 ms
D 4 ms

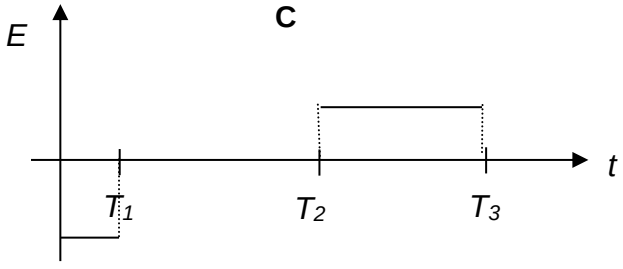
S8 [HCI 2007 BT1]
The magnetic flux Φ through a coil varies with time t as shown in the diagram.

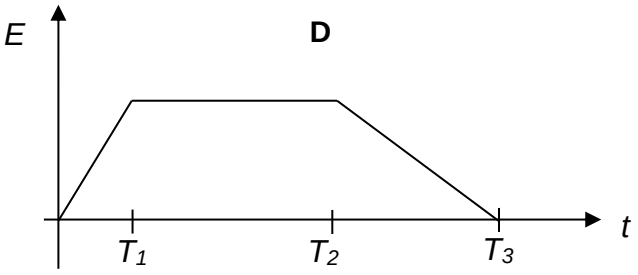


Which graph best represents the variation with time t of the e.m.f. E induced in the coil?

A


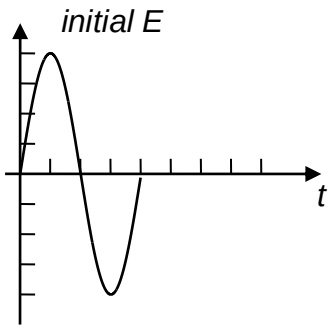
B


C


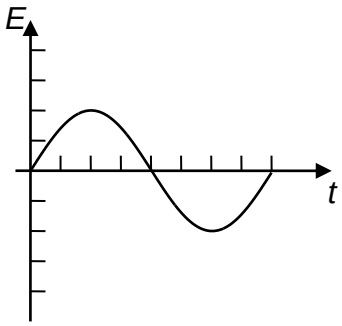
D


S9 [HCI 2008 BT2] When a coil is rotated in a uniform magnetic field at a certain frequency, the variation with time *t* of the induced e.m.f. *E* is shown on the left side.

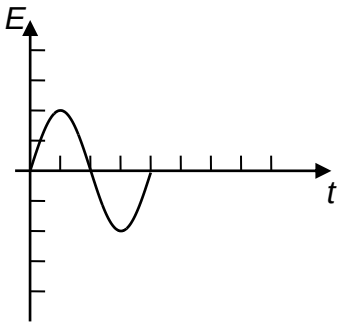
The frequency of rotation of the coil is reduced to **one half** of its initial value. Which **one** of the following graphs correctly shows the new variation with time *t* of the induced e.m.f. *E*?



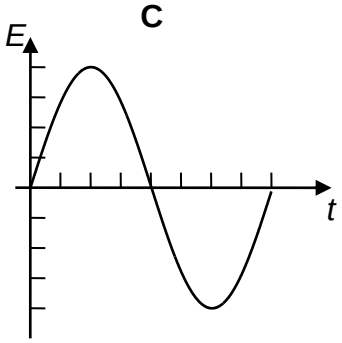
A



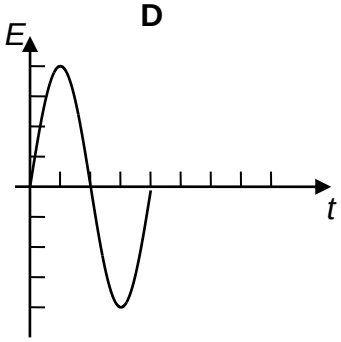
B



C



D

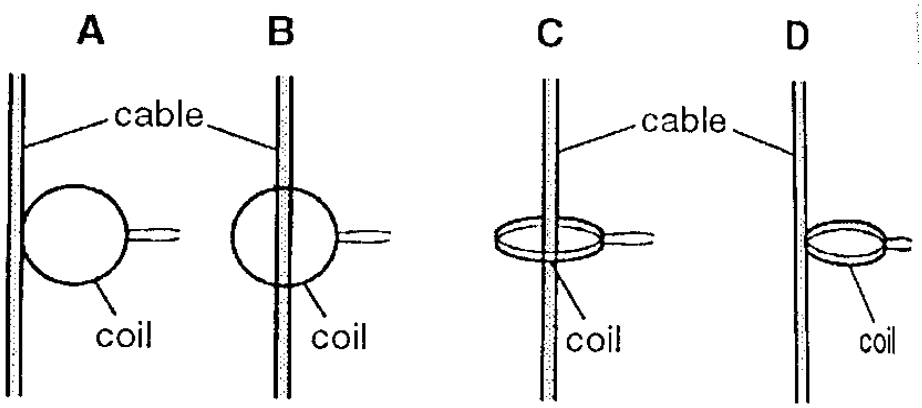


Answers for Self-Practise Qns:

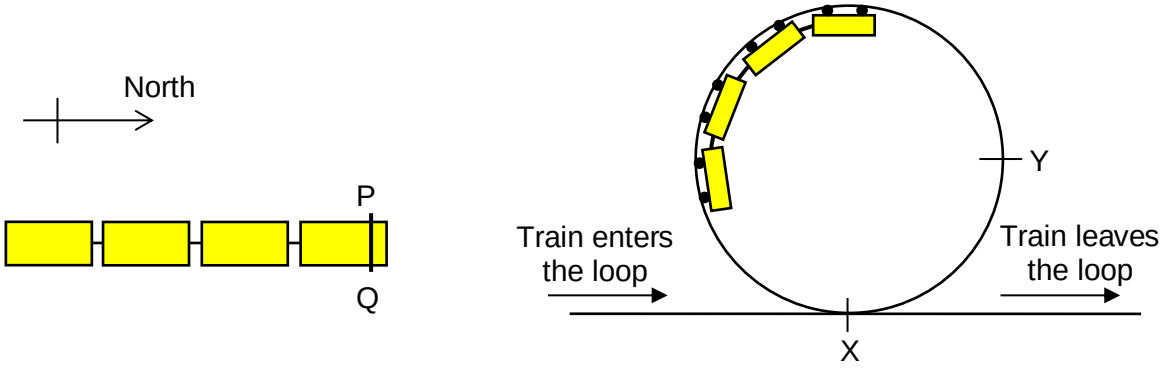
S2	**C**	**S7**	**D**
S3	**0.771 Wb**	**S8**	**C**
S4	**0.18 T**	**S9**	**A**
S5	**D**		

Discussion Questions

D1 Large alternating currents in a cable can be measured by monitoring the e.m.f. induced in a small coil situated near the cable. This e.m.f. is induced by the varying magnetic field set up around the cable. In which arrangement of coil and cable will the peak of the e.m.f. induced be the highest?



D2 A roller coaster train moves in the Northward direction in a region where the Earth's magnetic field is horizontal. It 'loop-the-loop' in a vertical circle as shown. The axle PQ of the front wheels of the train is a conductor and is perpendicular to the Earth's magnetic field.



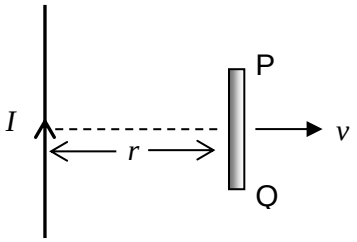
Top view

Side view

At which point, X or Y, of the loop will the induced e.m.f. in the axle PQ be larger, and correspondingly, which end of PQ has a higher potential?

	larger e.m.f.	*higher potential*
A	X	P
B	X	Q
C	Y	P
D	Y	Q

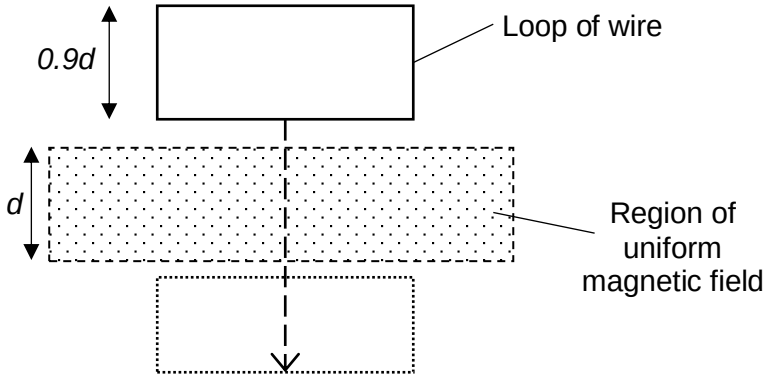
D3 The diagram on the right shows a wire carrying a current I. The magnetic flux density due to the straight wire is given by $B = \frac{\mu_0 I}{2\pi r}$. A straight conductor PQ is placed on the same vertical plane as the wire and is moved at constant speed v away from the wire.



How will the magnitude of the induced e.m.f. in PQ vary and which end will be at a higher potential?

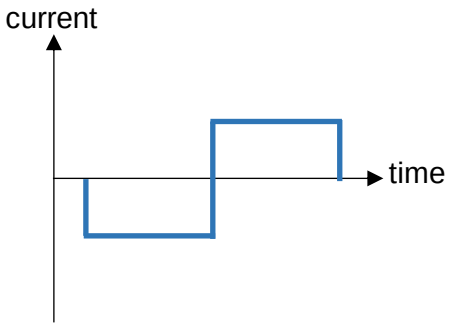
	magnitude of induced e.m.f.	end at higher potential
A	decreases	P
B	decreases	Q
C	increases	P
D	increases	Q

D4 A rectangular loop of wire is moving with a constant velocity through a region of uniform magnetic field which is perpendicular to the plane of loop. The loop has width of *0.9d* and the region of magnetic field has the width *d* as shown in the figure below.

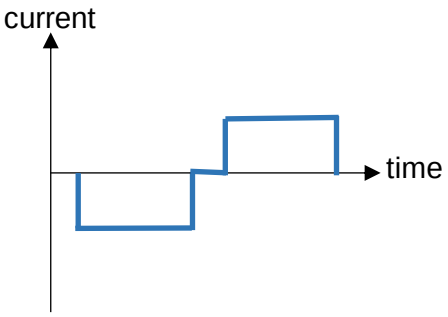


Which of the following graphs shows the flow of the current in the loop of the wire as it passes through the region of magnetic field?

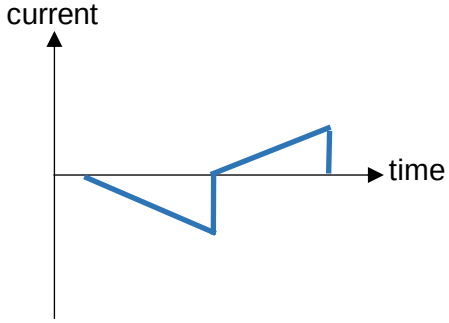
A



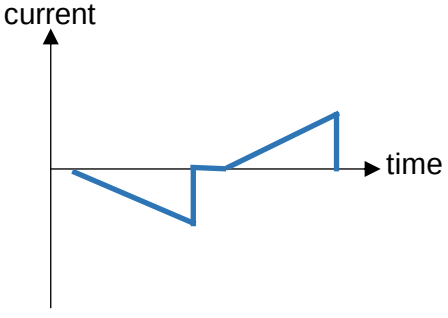
B



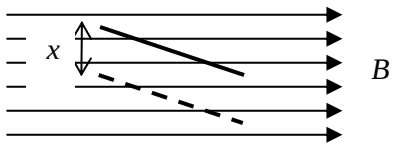
C



D



D5 [J82/II/20] A plane coil of wire containing N turns, each turn of area A is placed so that the plane of the coil makes an angle θ with the direction of a uniform horizontal magnetic field of flux density B. The coil is now moved through a vertical distance x in time t to the position shown dotted. What is the expression of the average e.m.f. induced in the coil?



D6 [N83/II/21] A rectangular coil of wire, initially placed as shown, is rotated with a constant angular velocity in a magnetic field which acts in the direction XX'.

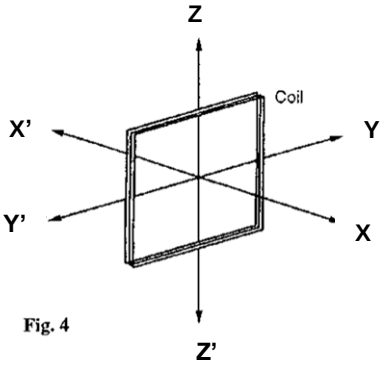
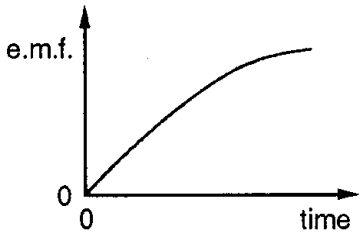


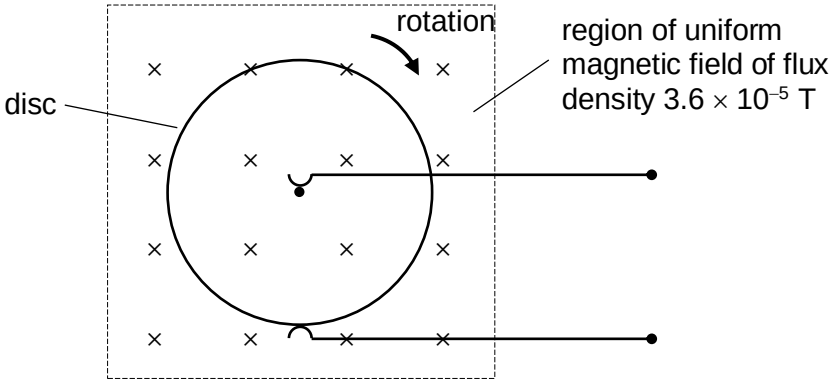
Fig. 4



The sinusoidal e.m.f. represented is observed to be produced across the ends of the coils. Which of the following movements would have given this result?

A rotation of a quarter revolution about axis YY'
B rotation of a half revolution about axis XX'
C rotation of half a revolution about the axis ZZ'
D rotation of a quarter revolution about axis XX'
E rotation of a half revolution about axis YY'

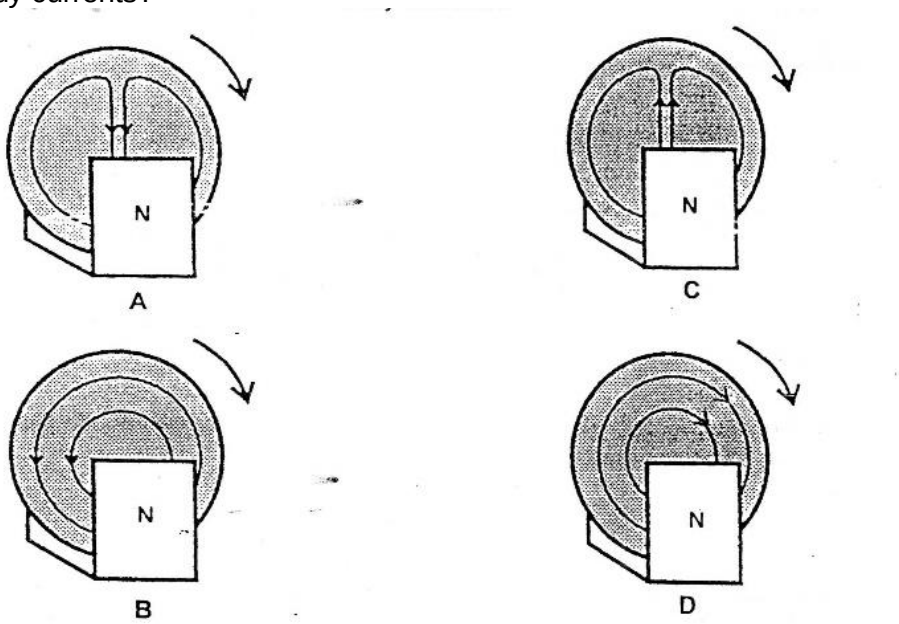
D7 The diagram below shows a metal disc of diameter 2.3 cm situated in a uniform magnetic field of flux density 3.6×10^{-5} T. The plane of the disc is perpendicular to the magnetic field which is directed into the plane of the diagram. The disc is rotated about an axis through its centre at 1500 revolutions per minute.



What is the magnitude of the induced e.m.f. between the centre of the disc and its rim?

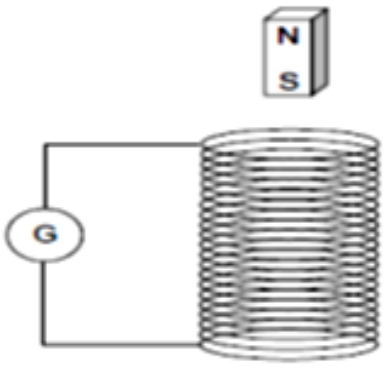
A 0 V **B** 3.74×10^{-7} V **C** 2.35×10^{-6} V **D** 2.24×10^{-5} V

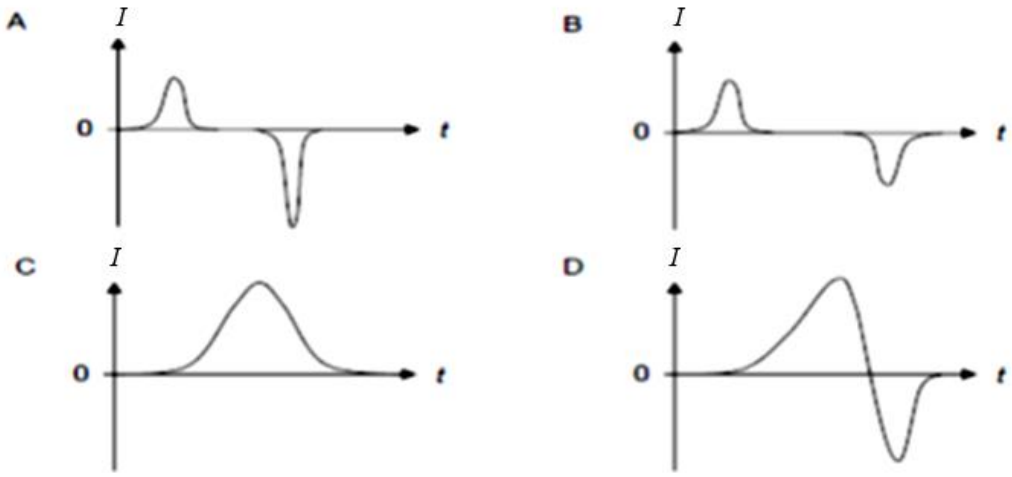
D8 A metal disc is spun clockwise between the poles of a small horse-shoe magnet. Which diagram shows the direction of eddy currents?



D9 [HCI 2010 BT1]

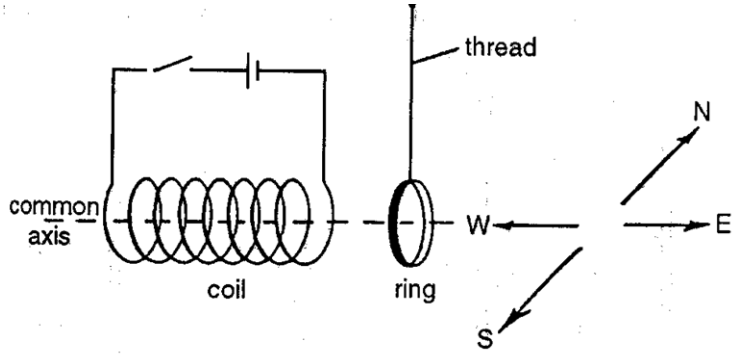
A short bar magnet is dropped vertically through a long solenoid. A galvanometer is connected across the solenoid. Which of the following graphs best represents the variation of current measured with the galvanometer over time?



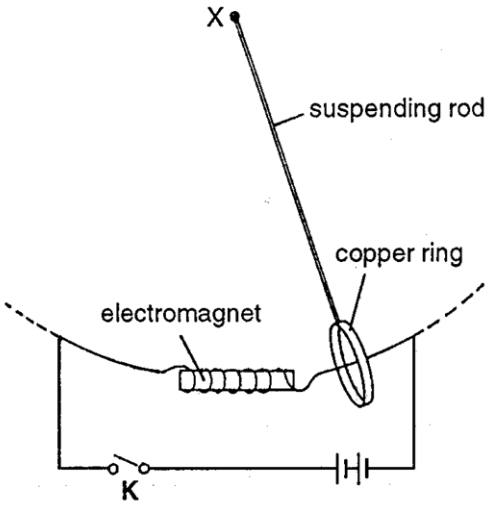


D10 [N86/I/20] An aluminium ring hangs vertically from a thread with its axis pointing east-west. A coil is fixed near the ring, coaxial with it. What is the initial motion of the aluminium ring when the current in the coil is switched on?

A remains at rest
B moves towards S
C moves towards W
D moves towards N
E moves towards E

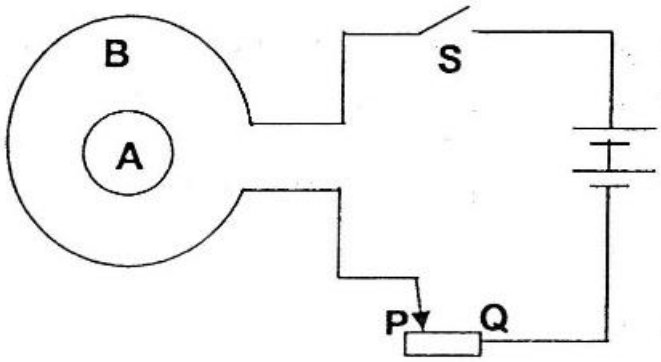


D11 [J86/I/19] A copper ring is suspended by a long, light rod pivoted at X so that it may swing as a pendulum. An electromagnet is mounted so that the ring passes over it as it swings. The ring is set into oscillation when the switch K opens. What happens to the motion of the ring after the switch K closes, ignoring the effect of air resistance?



A The amplitude will decrease initially and remains constant thereafter.
B The amplitude will decrease continuously.
C The amplitude will increase initially and will decrease thereafter.
D The amplitude will remain constant.

D12 A closed smaller circular wire loop A lies in the plane of another larger loop B, which is connected to a battery and rheostat as shown.

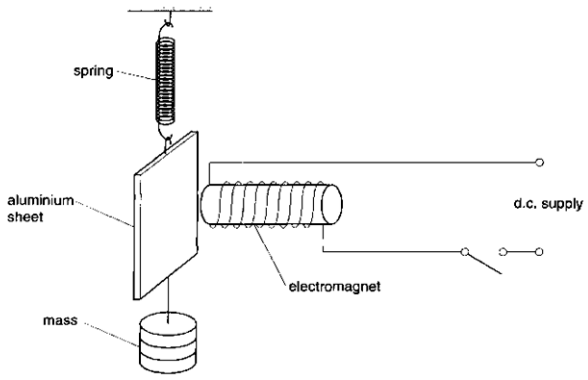


(i) Sketch the directions of the induced currents in coil A
1. when switch S is just closed
2. when switch S is already closed, slider is moving from P towards Q
3. when switch S is opened again after being closed for some time.

(ii) Switch S is closed for some time and coil A is now rotated at a constant angular velocity about its diameter. Draw sketch graphs on the same time axis, to show the variation of
1. the magnetic flux through coil A
2. the current induced in coil A

D13 [N99/II/5]

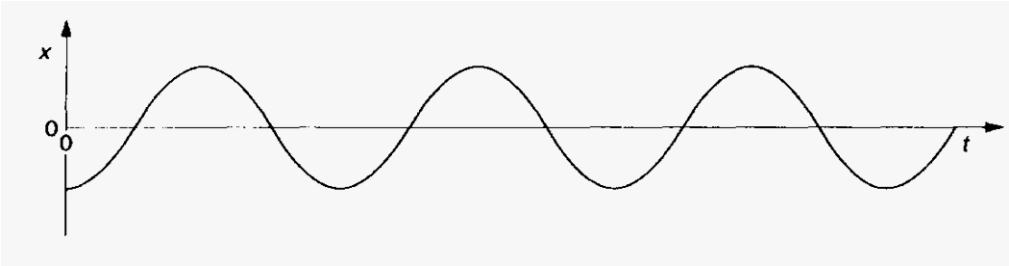
A helical spring is clamped vertically. The free end of the spring is attached to a sheet of aluminium and a mass.



An electromagnet is placed near the centre of the aluminium sheet. The mass is displaced vertically and then released, with the electromagnet switched off. The variation with time t of the displacement x of the mass is shown.

The electromagnet is switched on and the experiment is repeated with the same initial amplitude. The motion of mass is observed.

a) Sketch the new variation with time of the displacement x of the mass on the same graph.



b)

1. State Faraday's law of electromagnetic induction.
2. Describe and explain the new variation of the motion of the mass.

c) Suggest how critical damping could be demonstrated using the apparatus shown in above diagram.

D14 [J93/II/4]

A current carrying solenoid is placed near a search coil as shown below.

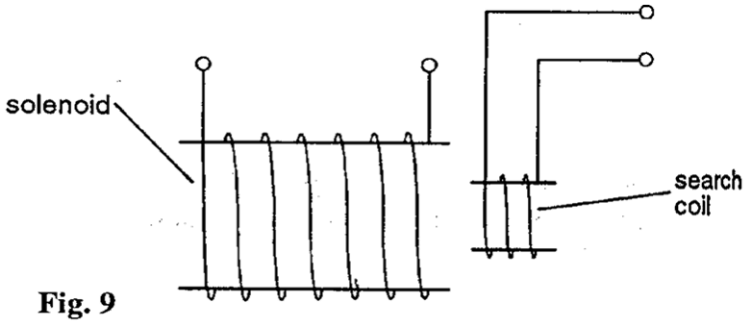
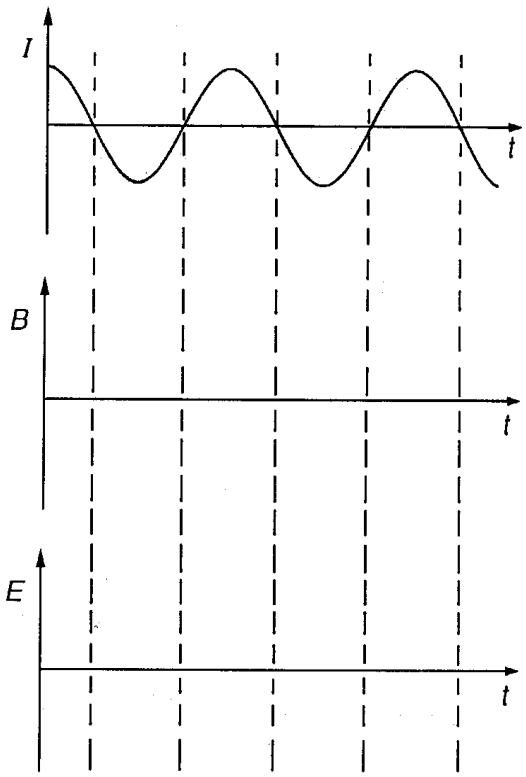


Fig. 9



(a) The variation with time of the current I in the solenoid is shown on the right.

(i) Sketch the variation with t of the magnetic flux density B in the solenoid.

(ii) Sketch the variation with t of the induced e.m.f. E in the search coil.

(b) For the experiment outlined above, briefly describe and explain the effect on the amplitude and frequency of the induced e.m.f. E, if separately,

(i) a ferrous core is slowly introduced into the solenoid;

(ii) the frequency of the current in the solenoid is increased, whilst maintaining the same amplitude.

D15 [2018 H2PH, P3 Q4b]

(a) A coil of wire is wound around one end of a soft-iron core as shown in Fig. 4.1.

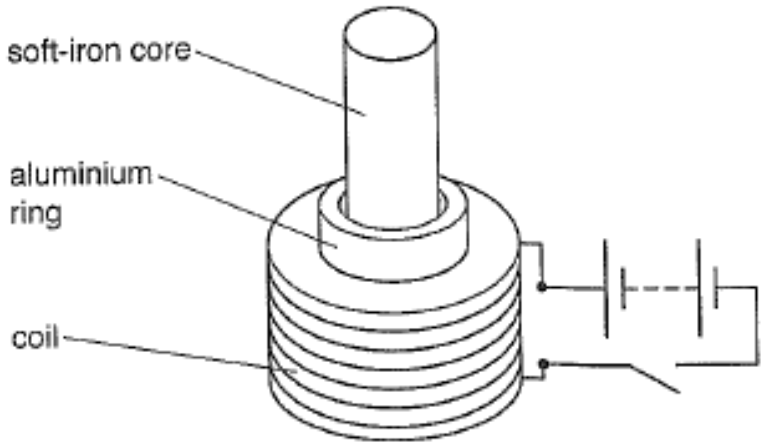


Fig. 4.1

The axis of the coil is vertical.

An aluminium ring is placed over the other end of the soft-iron core. The ring can move freely on the core.

The switch is closed and there is a direct current in the coil.

Explain why the aluminium ring moves vertically when the switch is closed. [4]

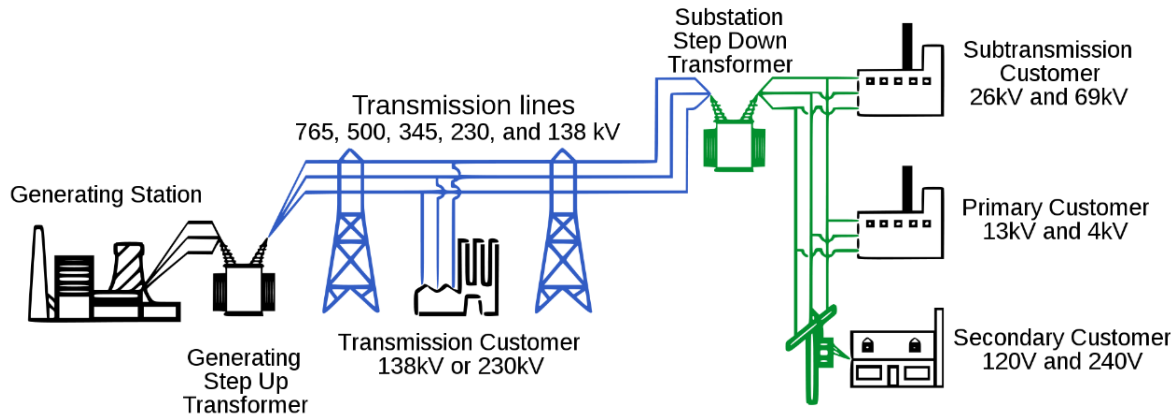
(b) The switch is opened. The aluminium ring is now replaced with a ring made from an electrical insulator. This ring can also move freely on the core.
State and explain whether this ring would move vertically when the switch is closed. [2]

Chapter 16

Alternating Current

Engineers design transmission networks to transport the energy as efficiently as feasible, while at the same time taking into account economic factors, network safety and redundancy. These networks use components such as power lines, cables, circuit breakers, switches and transformers. The transmission network is usually administered on a regional basis by an entity such as a regional transmission organization or transmission system operator.

Transmission efficiency is hugely improved by devices that increase the voltage, and proportionately reduce the current in the conductors, thus keeping the power transmitted nearly equal to the power input. The reduced current flowing through the line reduces the losses in the conductors.



Source:
http://en.wikipedia.org/wiki/Electric_power_transmission

Playlist of lecture examples can be found at https://for.edu.sg/ldx72q

16 Alternating Current

Content (H2 Physics 9749)

- Characteristics of alternating currents
- The transformer
- Rectification with a diode

Learning Outcomes

Candidates should be able to:

a) show an understanding and use the terms period, frequency, peak value and root-mean-square (r.m.s.) value as applied to an alternating current or voltage.

b) deduce that the mean power in a resistive load is half the maximum (peak) power for a sinusoidal alternating current.

c) represent an alternating current or an alternating voltage by an equation of the form $x = x_0 \sin \omega t$.

d) distinguish between r.m.s. and peak values and recall and solve problems using the relationship $I_{rms} = \frac{I_0}{\sqrt{2}}$ for the sinusoidal case.

e) show an understanding of the principle of operation of a simple iron-cored transformer and recall and solve problems using $\frac{N_s}{N_p} = \frac{V_s}{V_p} = \frac{I_p}{I_s}$ for an ideal transformer.

f) explain the use of a single diode for the half-wave rectification of an alternating current.

Table of Contents

16.1 What is alternating current (AC)? …… 4

16.2 Sinusoidal AC …… 5

16.3 What does the root-mean-square (r.m.s.) value mean? …… 5
16.3.1 Root-mean-square (r.m.s.) values for sinusoidal AC …… 7
16.3.2 Root-mean-square (r.m.s.) values for non-sinusoidal AC …… 8

16.4 Transformer …… 9
16.4.1 Induced e.m.f.s in a transformer …… 9
16.4.2 Power transmission in a transformer …… 11
16.4.3 Efficiency …… 12

16.5 Rectification …… 12
16.5.1 Half-wave rectification …… 13

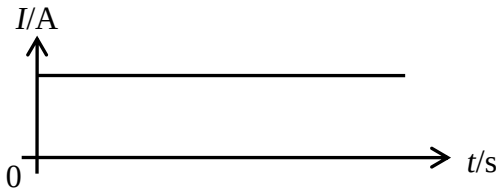
Appendix …… 14

Tutorial 16 …… 16

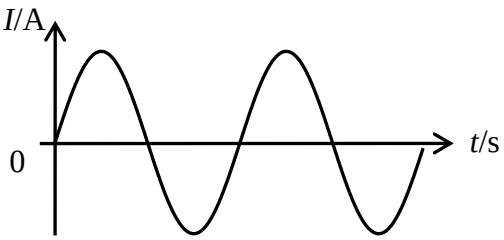
16.1 What is alternating current (AC)?

Direct current (DC) refers to current that does not change in direction. One of the most common DC sources is a battery.

On the other hand, the direction of an alternating current (AC) will vary with time. A common form of AC (e.g. the 230 V, 50 Hz AC power sockets in our homes) is one that varies in a sinusoidal manner.



Constant DC

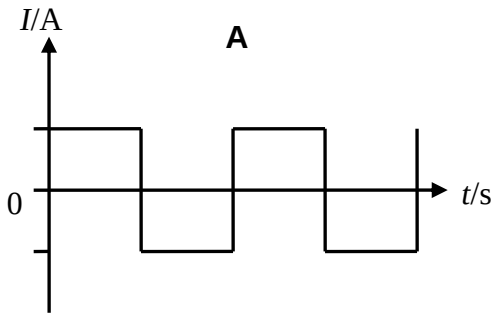


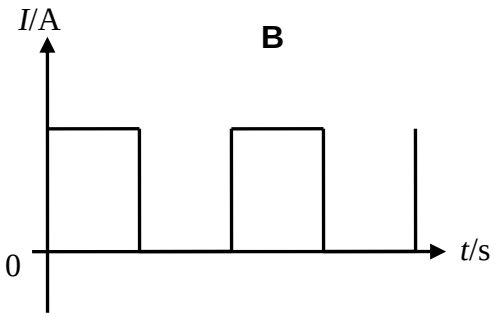
Sinusoidal AC

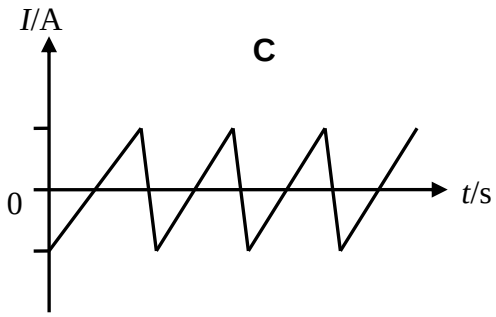
Alternating current occurs when charge carriers periodically reverse their direction of motion.

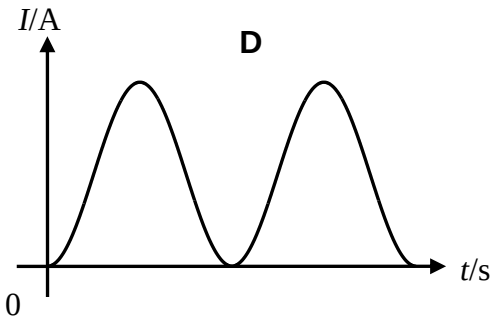
Example 1:

Which of the following is/are alternating current(s)?









16.2 Sinusoidal AC

The electricity supply generated for consumers' use is a sinusoidal AC.

The time-variation of a sinusoidal current or voltage can be represented by the equation

$$x = \pm x_0 \sin \omega t \quad \text{or} \quad x = \pm x_0 \cos \omega t$$

where x_0 = peak (or maximum) value
ω = angular frequency of the supply (rad s^{-1}).

Note that $\omega = 2\pi f$ where f is the frequency of the supply in hertz (Hz). Frequency is also related to the period T through the expression $f = \frac{1}{T}$.

Example 2:

The AC supply found in Singapore homes have peak voltage of 330 V and frequency of 50 Hz. Write down an expression for the voltage and sketch the waveform.

16.3 What does the root-mean-square (r.m.s.) value mean?

If a *steady DC* passes through a resistive load R, the power P dissipated can be calculated by $P = VI$, $P = I^2R$, or $P = \frac{V^2}{R}$. Since V and I are unchanging for a DC, there is no difference between instantaneous and average power dissipated in the resistive load.

If the current through the load is *alternating current (AC)*, the **instantaneous power** dissipated must be changing with time, since V and I are changing with time. In practice, we are often only interested in the **average power** dissipation. Interestingly, the average power can be calculated by the same set of equations ($P = VI$, $P = I^2R$ or $P = \frac{V^2}{R}$), as long as we use the **root-mean-square** values $V_{rms.}$ and $I_{rms.}$.

So, what is the root-mean-square value then?

Suppose a periodic alternating current I passes through an electrical load of resistance R, its instantaneous power P dissipated will be changing with time t. The peak current is I_0.

Instantaneous power $P = VI = I^2R = \frac{V^2}{R}$

Peak power $P_0 = V_0 I_0 = {I_0}^2 R = \frac{{V_0}^2}{R}$

In general,

mean power $< P > = < I^2 R >$

$= < I^2 > R$ (since R is constant)

where $< I^2 >$ is the mean-squared current I.

We define the root-mean-squared current $I_{rms} = \sqrt{< I^2 >}$,
then the mean power can also be expressed as $< P > = (I_{rms})^2 R$

The **root-mean-square** value of an alternating current is that **equivalent constant direct current** that will **dissipate the same average power in a given resistive load**.

In other words, a direct current of 1 A and an alternating current of r.m.s. value of 1 A have **the same heating effect**, i.e. both currents dissipate the same power in a given resistor.

Mean power $< P > = V_{rms} I_{rms} = {I_{rms}}^2 R = \frac{{V_{rms}}^2}{R}$ [1]

Based on the above equation in [1], we can derive that $V_{rms} = I_{rms} R$

Note that the above relationships are not restricted to sinusoidal AC. They are applicable to non-sinusoidal AC as well.

Last but not least, for a periodic current with period T, the mean-squared current $< I^2 >$ is computed using

$$< I^2 > = \frac{\int_0^T I^2 \, dt}{T} = \frac{1}{T} \int_0^T I^2 \, dt .$$

In the next section, you will learn how to compute the r.m.s. values and the mean power for a sinusoidal AC.

16.3.1 Root-mean-square (r.m.s.) values for sinusoidal AC (IMPORTANT!)

Suppose a sinusoidal current of $I = I_0 \sin \omega t$, as shown in **Fig. (a)**, passes through an electrical load of resistance R.

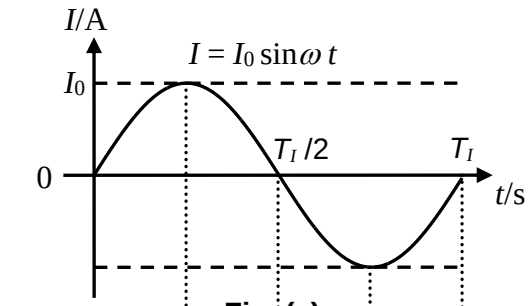


Fig. (a)

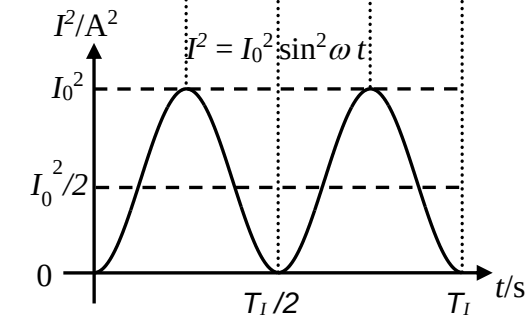


Fig. (b)

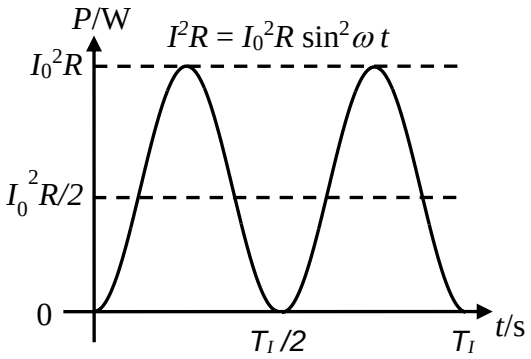


Fig. (c)

Fig. (b) shows I^2 against time and **Fig. (c)** shows power dissipated against time graph for one period. The period T_I in **Fig. (b)** and **(c)** follows the period for the current in **Fig. (a)**.

Using the concept of power dissipation,

mean power $<P> = \dfrac{(area\ under\ P\ vs\ t\ graph)}{duration\ of\ one\ period}$

Because of the symmetrical nature of the sine function, the **Fig. (c)** indicates that

$<P> = \frac{1}{2}(I_0{}^2 R)$ which is equal to $\frac{1}{2} P_{max}$

$= \left(\frac{I_0}{\sqrt{2}} \right)^2 R$

From the previous page in equation [1], which shows that

$<P> = (I_{rms})^2 R$,

we can deduce that $I_{rms} = \dfrac{I_0}{\sqrt{2}}$ for a sinusoidal current.

In summary, for a **sinusoidal** AC,

$$\boxed{I_{rms} = \frac{I_0}{\sqrt{2}} = 0.707 I_0}$$,

$$\boxed{V_{rms} = \frac{V_0}{\sqrt{2}} = 0.707 V_0}$$ and

$$\boxed{<P> = \frac{1}{2} P_{max}}$$

Note that the above 3 formulae are ***ONLY applicable to sinusoidal AC***.

For other forms of AC, you would need to derive the corresponding relationship between its r.m.s. value and its peak value.

Helpful Tips

For the A-level exams, AC supply is assumed sinusoidal unless otherwise stated. The value of the AC power quoted in an exam question usually refers to the mean power.

A.c. current and voltage values are *often quoted in r.m.s. values* unless otherwise stated. The r.m.s. values and the mean power are more meaningful and useful rather than the peak values (which are just instantaneous values).

Example 3:

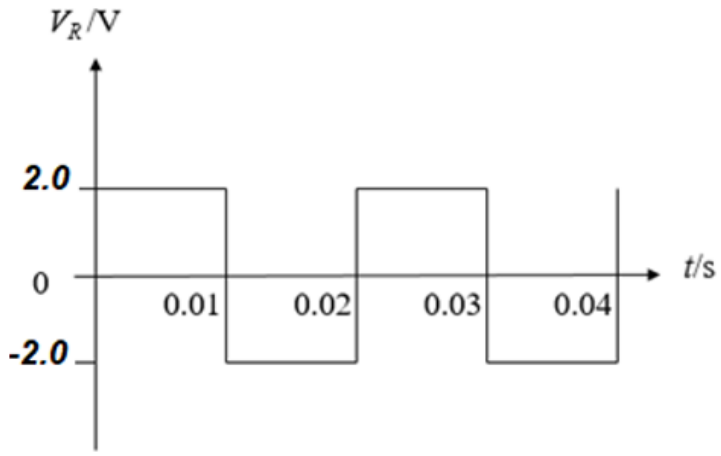
A 3.0 Ω resistor is connected to a sinusoidal AC supply of peak voltage 5.0 V. Calculate
(a) the r.m.s. voltage;
(b) the r.m.s. current;
(c) the mean power;
(d) a D.C. supply is used instead. What must the D.C. voltage be to produce the same power (rate of heating) in the 3.0 Ω resistor?

16.3.2 Root-mean-square (r.m.s.) values for non-sinusoidal AC

Example 4:

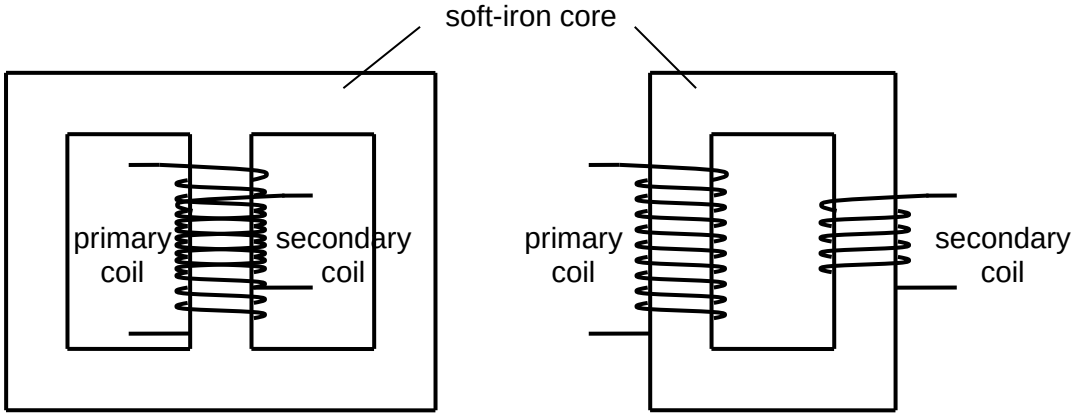
The waveform of potential difference V_R across a 3.0 Ω resistor is shown below. Determine

(a) the mean potential difference
(b) the r.m.s. potential difference and
(c) the average power dissipated in the resistor.

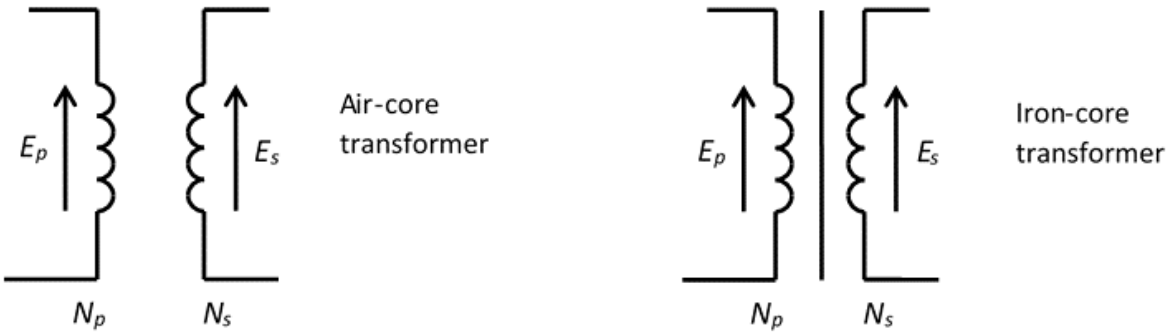


16.4 Transformer

The transformer is a device which makes use of electromagnetic induction to produce a large alternating e.m.f. from a small one, or a small alternating e.m.f. from a large one. The primary and secondary coils are formed by winding insulated wires on a soft iron core, either one on top of the other or on separate limbs (as shown in the figures below). There is no electrical connection between the coils but they are linked magnetically via the soft iron core; i.e. the flux is the same in both coils.

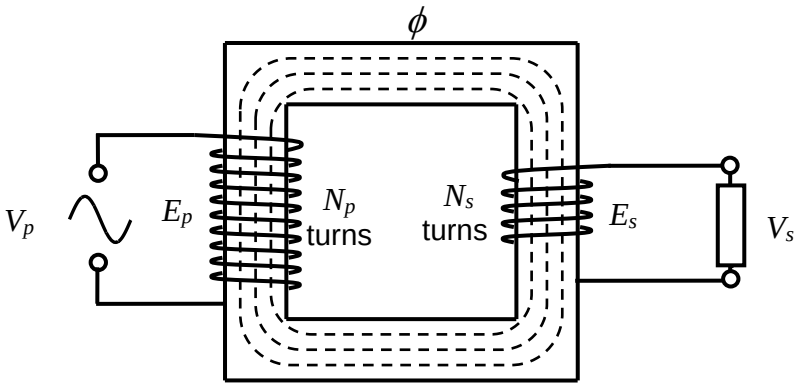


Below are circuit symbols of transformers.



16.4.1 Induced e.m.f.s in a transformer

When an alternating e.m.f. V_p is applied to the primary coil, alternating currents in both primary coil and secondary coil set up a resultant alternating flux ϕ in the core.



If N_p and N_s are the number of turns in the primary and secondary coils, their linkages with the flux ϕ are

$(\Phi_{linkage})_p = N_p \phi$ and $(\Phi_{linkage})_s = N_s \phi$

The alternating flux induces an alternating e.m.f. E_p in the primary coil as well as an alternating e.m.f. E_s in the secondary coil. By Faraday's law, the magnitudes of E_p and E_s are given by

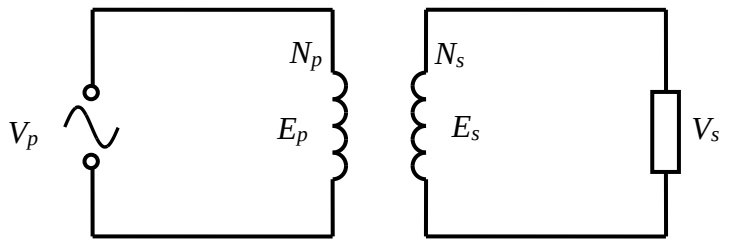
$$E_p = \frac{d(\Phi_{linkage})_p}{dt} = N_p \frac{d\phi}{dt} \quad \text{and} \quad E_s = \frac{d(\Phi_{linkage})_s}{dt} = N_s \frac{d\phi}{dt}$$

Dividing the above two equations, we have $\dfrac{\textit{e.m.f. induced in secondary}}{\textit{e.m.f. induced in primary}} = \dfrac{E_s}{E_p} = \dfrac{N_s}{N_p}$

Thus the transformer steps voltage up or down according to its 'turns-ratio' $\frac{N_s}{N_p}$.

If V_s is the output p.d. across the secondary coil, then for an ***ideal transformer*** that has <u>*zero-resistance coils*</u>, $V_p = E_p$ and $V_s = E_s$, giving

$$\boxed{\frac{V_s}{V_p} = \frac{N_s}{N_p}}$$



Note that

a) When the secondary coil is open-circuited, the same analysis can still be applied and resultant alternating flux ϕ in the core will be due to the current in the primary coil.

b) V_p must be an alternating e.m.f.. If not, the flux in the primary coil would not change (except for a short time immediately after the e.m.f. is first applied) and there would be no induced e.m.f. in the secondary coil.

c) The induced e.m.f. E_s has the same frequency as the applied voltage V_p.

d) If $N_s > N_p$, the transformer is called a *step-up transformer* because $V_s > V_p$.
A step-down transformer has $N_s < N_p$.

16.4.2 Power transmission in a transformer

If the transformer is ideal, i.e. *100% efficient,*

output power = input power

$$V_s I_s = V_p I_p$$

Thus, turns ratio $\boxed{\frac{N_s}{N_p} = \frac{V_s}{V_p} = \frac{I_p}{I_s}}$

The above relationship is *only applicable to an ideal transformer.* In the A levels, transformers are assumed to be ideal unless otherwise stated.
Note that the relationship is obtained from the Principle of Conservation of Energy.

Example 5:

A step-up transformer has 200 turns on its primary coil and 3000 turns on its secondary coil. The primary coil is supplied with a 30 A alternating current at 90 V.

(a) What is the voltage in the secondary circuit?
(b) What current flows in the secondary circuit?
(c) What is the power in the primary circuit?
(d) What is the power in the secondary circuit?

16.4.3 Efficiency

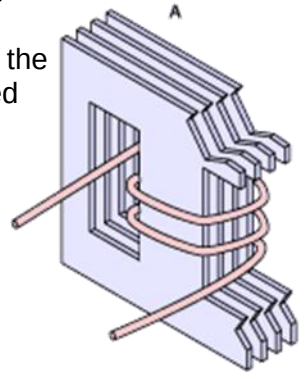
The efficiency of commercial transformers are very high – typically in the range of 95% to 99%. The four most important sources of power loss are listed below.

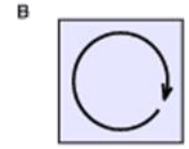
(a) Resistance of windings
Some energy is dissipated as heat in both coils (I^2R). This can be reduced to an acceptable level by using suitably thick wires. (The coil which has the smaller number of turns, carries the larger current. And thus, it usually requires thicker wires.)

(b) Eddy Currents
The alternating magnetic flux induces eddy currents in the iron core and causes heating. The effect is reduced by having a laminated core.

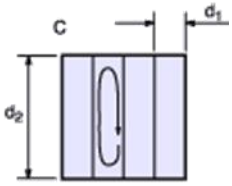
Figure PLM: Lamination

A: Cross section showing the laminated core





B: Eddy current in an unlaminated core (viewed from the top)



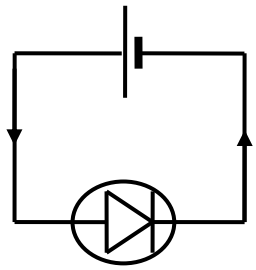
C: Eddy current in a laminated core (viewed from the top)

(c) Hysteresis
Each time the direction of magnetization of the core is reversed, some energy is wasted. This is known as hysteresis loss and it produces heating in the core. It is minimized by using special alloys (e.g. Permalloy which is an alloy of nickel and iron) for the core material.

(d) Flux Leakage
The flux due to the primary coil may not link to the secondary coil completely if the core is badly designed.

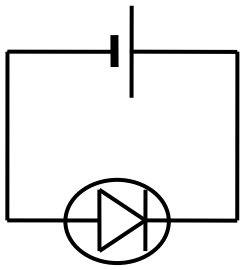
16.5 Rectification

The I-V characteristic of rectifiers (or diodes) are such that they present a very low resistance to current flow in one direction (when forward biased), and a very high resistance in the opposite direction (when reverse biased).



I

In the forward biased connection, the diode offers negligible resistance and current flows through it easily.

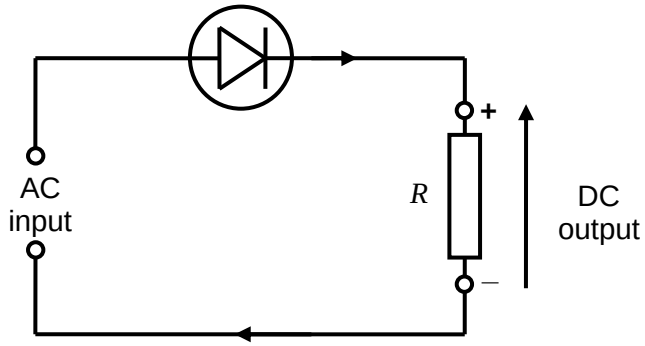


In the reverse biased connection, the diode offers infinite resistance and no current flows in the circuit.

16.5.1 Half-wave rectification

Rectification is the conversion of AC to DC. A simple half-wave rectifying circuit is shown below. For simplicity, the resistor R is used to represent a load which requires a DC power source.

As the applied potential difference to the rectifier and load is a sinusoidal wave, the diode is forward biased during one half-cycle, and reverse biased during the next half-cycle.



During the half-cycle when the rectifier is forward biased, current flows round the circuit, creating a potential difference across R. This p.d. will have almost the same value as the applied p.d. if the forward resistance of the diode is small compared with R.

During the half-cycle when the diode is reverse biased, little or no current flows and the potential difference across R is zero.

This is repeated for each cycle of AC input. The current flow is unidirectional and so is the potential difference across R. Although it fluctuates, it never changes direction.

Example 6:

Consider the half-wave rectifying circuit above.

The variation with time *t* of the potential difference from the AC supply is

$$V = V_0 \sin(\omega t)$$

a) Sketch the waveform for the potential difference across the load *R* after the half-wave rectification.

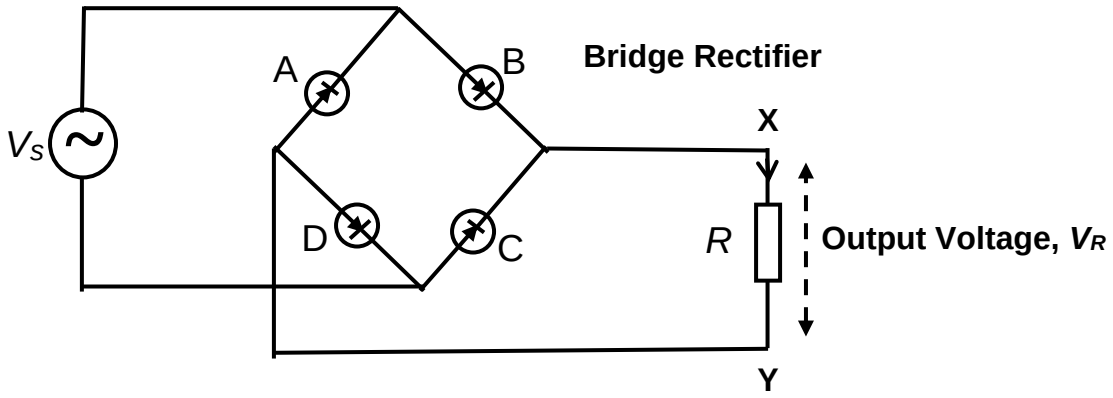
b) Determine the r.m.s voltage across the load in terms of V_0. (Ans: $\frac{V_0}{2}$)

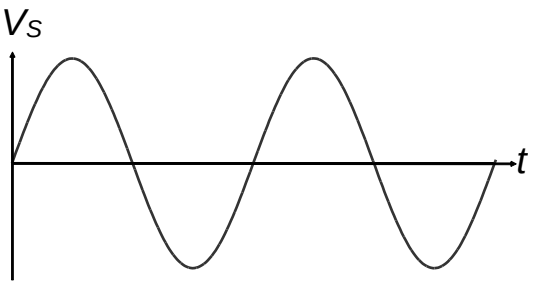
c) Hence, find the power dissipated in the load in terms of V_0 and *R*. (Ans: $\frac{V_0^2}{4R}$)

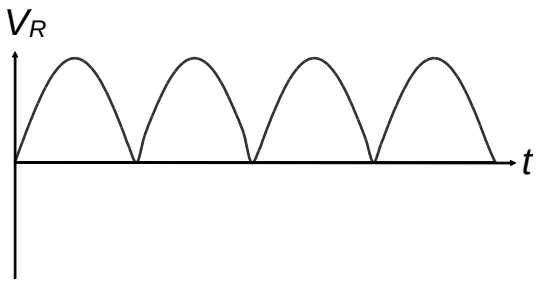
Appendix

1 Full-wave Rectification

Full-wave rectification can be accomplished with a bridge rectifier consisting of 4 diodes. During any half cycle, two of the four diodes are forward biased.



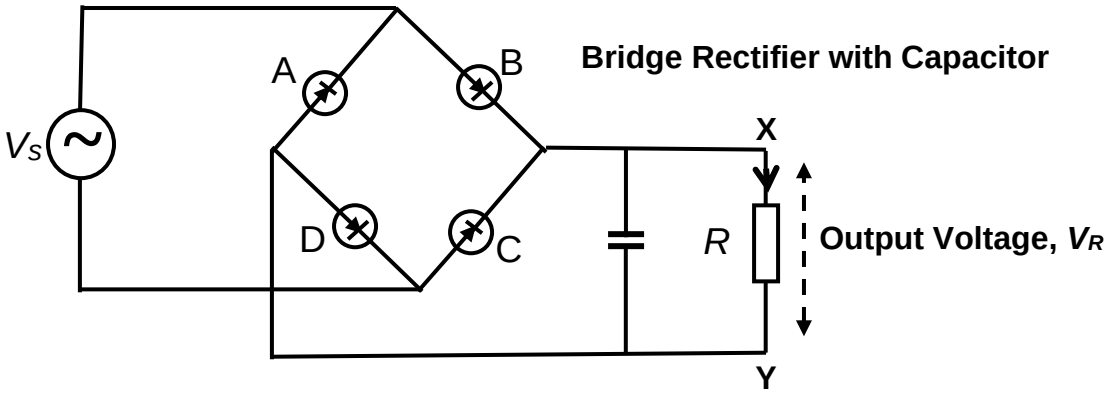


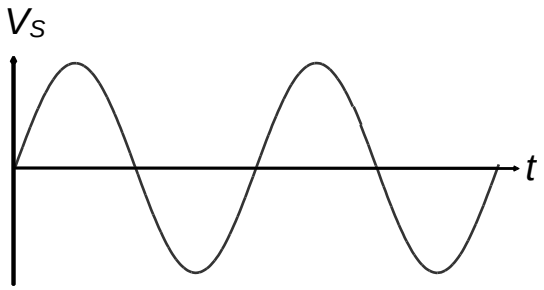


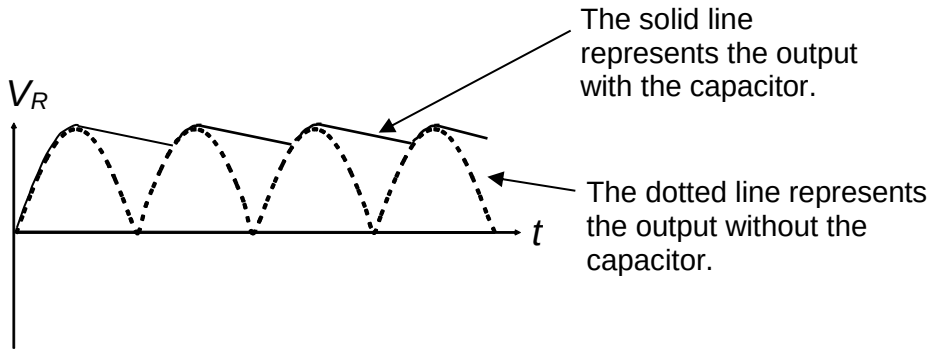
- When the upper terminal of V_S is positive with respect to the lower terminal, diodes B and D are forward biased, while diodes A and C are reverse biased.
- When the lower terminal of V_S is positive with respect to the upper terminal, diodes A and C are forward biased, while diodes B and D are reverse biased.
- Hence, current will always flow in the direction from X to Y through the load resistor R, and the output voltage V_R is unidirectional.

2 Smoothing output voltage

Smoothing is the process of reducing the fluctuations of the rectified DC output. In other words, it is to make the output more constant. This is done by connecting a special device, known as a capacitor, in ***parallel*** with the load R .







The output of the rectifier is 'smoothed' by connecting a capacitor across the output.

Tutorial 16 Alternating Current

Self-Review Questions

S1 A fuse is rated (to fuse) at 3 A. The fuse is connected in series with a 100 Ω resistor. Will the fuse break if the resistor (and fuse) is connected to a

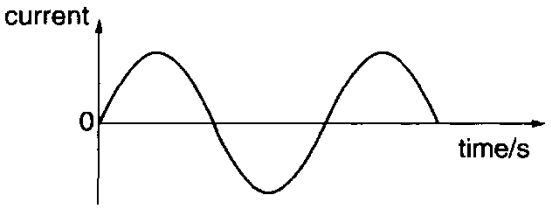
a) 220 V DC supply
b) 220 V AC supply

S2 An electric kettle, designed for travellers, can be used with different supply voltages. It is rated at 700 W for a 240 V r.m.s. alternating supply. Assuming its resistance is constant, what will be its power output if it is connected to

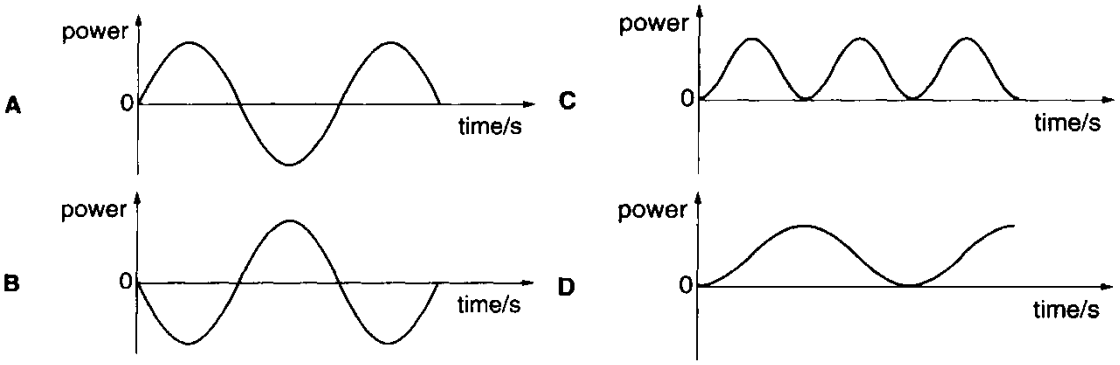
a) a 120 V r.m.s. alternating supply,
b) a 120 V direct supply?

S3 N96/I/21

An alternating current flows through a resistor. The variation with time of this current is shown.



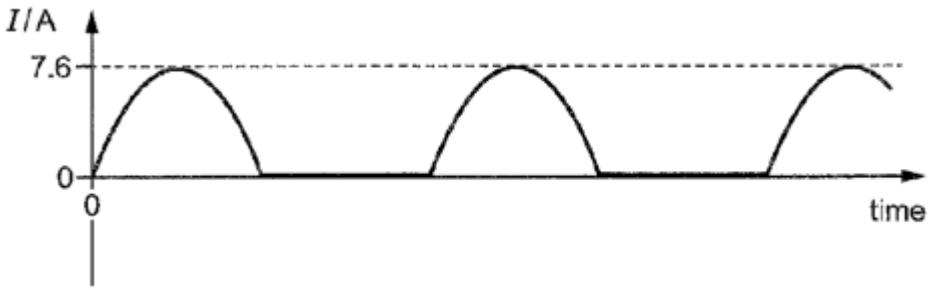
Which graph shows the variation with time of the power dissipated in the resistor?



S4 A student tried to step up a 60 V DC supply to 120 V using a transformer which has a step-up factor of 2. Will he succeed?

S5 N18/I/27

An alternating voltage is connected in series to a diode and a resistor of resistance 9.4 Ω. The current in the circuit varies as shown.



What is the mean power dissipated in the resistor?

A 34 W **B** 68 W **C** 140 W **D** 270 W

S6 N17/I/26

A 16:1 step-down transformer is connected to a load resistor *R*. The resistor dissipates power at a rate of 32 W when the peak alternating voltage of the supply is 180 V.

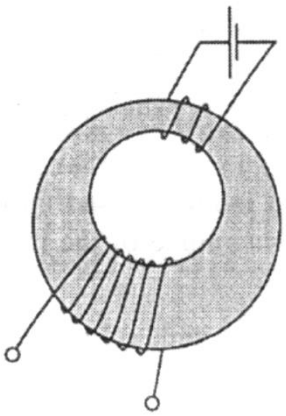
What is the value of *R*?

A 1.0 Ω **B** 2.0 Ω **C** 4.0 Ω **D** 32 Ω

Discussion Questions

D1 N09/I/33

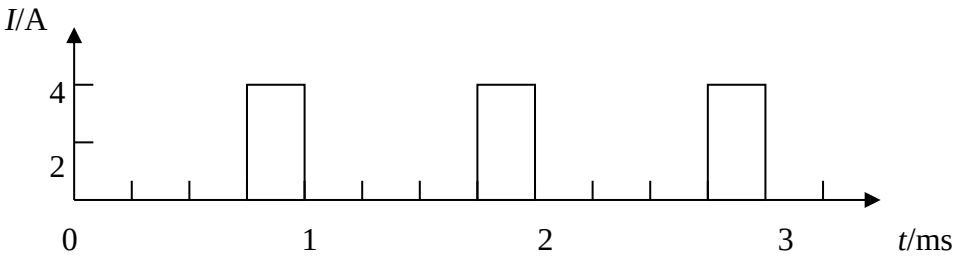
A soft iron ring of variable cross-section has two coils wound round it. One coil has 7 turns. The other coil which has 3 turns, is connected to a d.c. supply.



Which quantity is the same inside each coil?

A magnetic field
B magnetic flux
C magnetic flux density
D magnetic flux linkage

D2 J83/I/10



The figure above shows (in part) the variation with time of a periodic current.

a) What is the average value of the current?
b) What the root-mean-square current?

The periodic current passes through a resistor, producing heat at a certain rate.
c) What steady current, passing through the same resistor, would have an identical heating effect?

D3 J86/III/7

A signal generator produces either (a) a sinusoidal or (b) a square wave with the same peak value e.m.f.

Showing your working, determine for a pure resistive load, the ratio of the mean power in case (a) to the mean power in case (b).

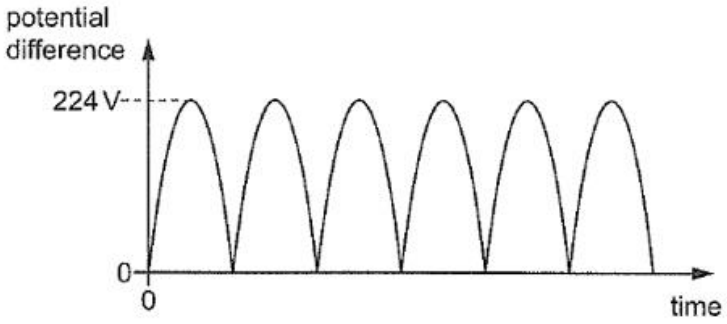
D4 N90/I/20

A steady current I dissipates a certain power in a variable resistor. The resistance has to be halved to obtain the same power when a sinusoidal alternating current is used. What is the r.m.s. value of the alternating current?

A $\frac{1}{2}\ I$ **B** $\sqrt{2}\ I$ **C** $\sqrt{\frac{1}{2}}\ I$ **D** $2\ I$

D5 N17/I/25

An alternating current is supplied to a rectifier. The output from the rectifier has the shape shown in the diagram, with a maximum value of 224 V.



This output is used as the supply to a resistor of resistance 70.0 Ω.

What are the r.m.s values of the current and the power to the resistor?

	r.m.s. current / A	power / W
A	2.26	358
B	2.26	507
C	3.20	507
D	3.20	717

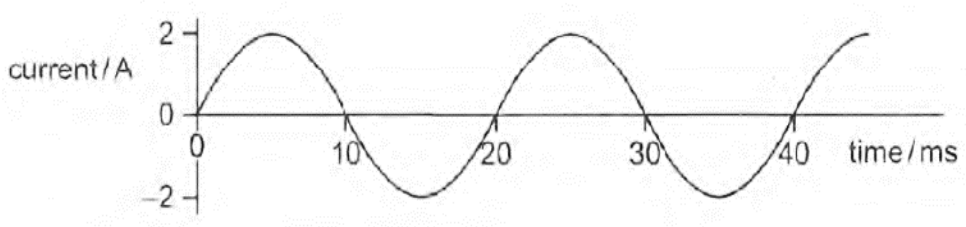
D6 N11/I/34

A sinusoidal alternating supply of peak voltage 100 V is connected to a diode and a 100 Ω resistor in series. The diode is ideal with a zero forward resistance and an infinite reverse resistance. What is the value of the mean current in the resistor?

A less than 0.5 A **B** 0.5 A **C** 0.7 A **D** 1 A

D7 N22/I/27

An a.c. generator is connected to a resistor of resistance 20 Ω. The graph shows the variation of the current in the resistor with time.



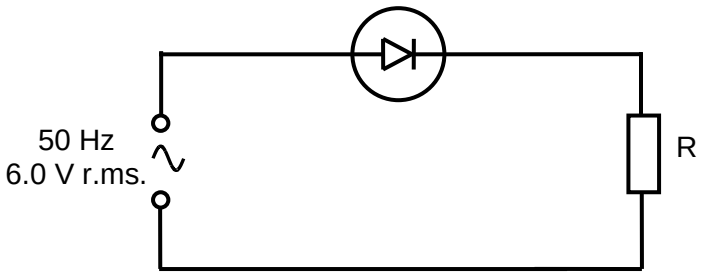
The rotational speed of the a.c. generator is halved.

What is now the mean power dissipated in the resistor?

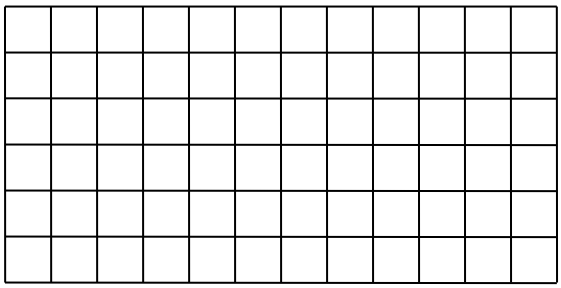
A 5 W **B** 10 W **C** 20 W **D** 40 W

D8 N98/II/4 (modified)

The diagram below shows the circuit diagram for a half-wave rectifier. The supply to the rectifier is rated at 50 Hz, 6.0 V r.m.s.



a) A cathode-ray oscilloscope (c.r.o.) has its Y-plates connected across the load resistor R. Given that the Y-plate sensitivity of the c.r.o. is 3.0 V division^{-1}, and the time base is 5 ms division^{-1}, sketch the trace seen on the c.r.o. Each box is 1 cm x 1 cm.



b) On the same diagram, sketch the trace seen when the connections to the diode are reversed.

D9 a) N2017/III/9 (part)

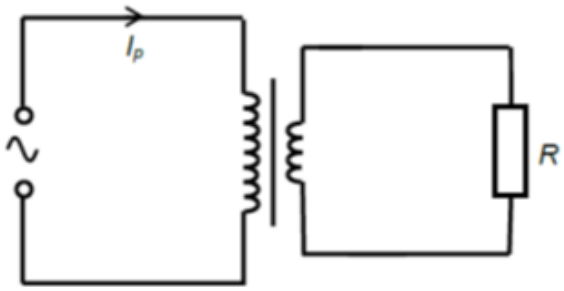
(i) Use Faraday's law to explain why an alternating current in the primary coil gives rise to an alternating e.m.f. in the secondary coil.
(ii) Explain what is meant by an ideal transformer.
(iii) In practice, the core of some transformers is made of laminated soft iron. State two reasons why soft iron, which is easily magnetised and demagnetised, is used as the core.

b) J96/II/4 (part)

In a transformer, the core on which the primary and secondary coils are wound is laminated, i.e. the core is made up of a large number of strips, rather than being one solid piece. This reduces energy losses due to currents induced in the core. Explain
(i) how these currents arise in the core,
(ii) why laminating the core reduces energy losses due to these currents.

D10 HCI Prelim/14/I/29

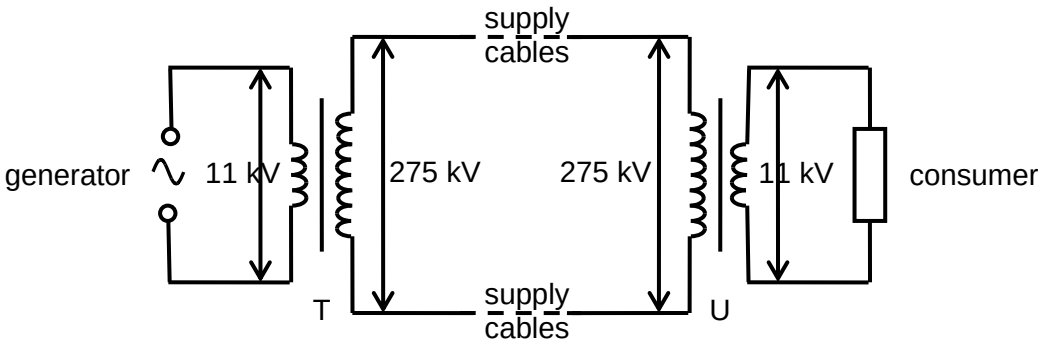
An ideal transformer is used to step down the a.c. voltage supply to a resistive load *R*. If the number of turns in the primary coil is doubled, what is the new current in the primary coil?



A Twice the original current in the primary coil
B Same as the original current in the primary coil
C Half the original current in the primary coil
D One quarter of the original current in the primary coil

D11 J98/II/5

Electrical power of 4400 kW is supplied to an industrial consumer at a considerable distance from a generating station. This is represented in the figure below.



In order to do this, the electricity supply company makes use of a circuit containing two transformers, T and U. The transformers can be considered to be ideal and the supply cables can too be considered to have negligible resistance.

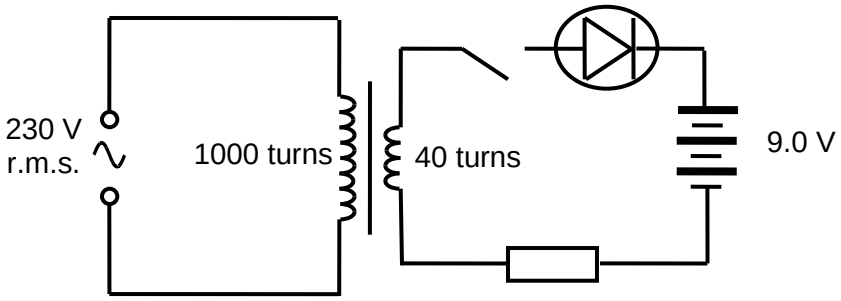
a) The power is generated at 11 kV r.m.s. and is supplied to the consumer at 11 kV r.m.s. Calculate the r.m.s. current supplied to the consumer.

b) There is a p.d. of 275 kV r.m.s. between the supply cables. Calculate
   i. the ratio N_s / N_p required for each transformer.
   ii. the current in the supply cables.

c) Explain why, when the resistance of the supply cables cannot be neglected, this arrangement is preferable to a system which generates and transmits the power at the same voltage of 11 kV r.m.s.

D12 J00/II/5 (modified)

The primary coil of a transformer has 1000 turns and is connected to a 230 V r.m.s. supply. The secondary coil has 40 turns and may be connected, through a switch and a diode, to a 9.0 V rechargeable battery.



a) Initially the switch is open. Assume the transformer and the diode are ideal, calculate
   i. the r.m.s. potential difference across the secondary coil,
   ii. the peak potential difference across the secondary coil.

b) The switch is now closed so that the battery is being recharged.
   i. Suggest why the diode is necessary in the secondary circuit.
   ii. Suggest why the resistor is necessary in the secondary circuit.

c) Sketch the p.d. across the resistor.

Numerical answers to Self-Review Questions

S2 i) 175 W ii) 175 W
S3 C
S5 C
S6 B

Numerical answers to Discussion Questions

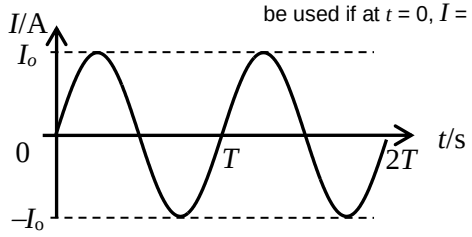
D2 a) 1 A b) 2 A c) 2A
D3 1:2
D11 a) 400 A b(i) T 25:1, U 1:25 b(ii) 16 A
D12 a(i) 9.2 V a(ii) 13 V

ALTERNATING CURRENT (AC)

definition → Alternating current occurs when charge carriers periodically reverse their direction of motion.

Sinusoidal AC

(The cosine form can also be used if at t = 0, $I = I_o$.)



An equation to represent the above current-time variation is $I = I_o \sin \omega t$

where I = instantaneous value of current

I_o = peak (or maximum) current

ω = angular frequency of the supply in $rad\ s^{-1}$

= 2 πf (f is the frequency of supply in Hz.)

Root-mean-square values (r.m.s) values

$$I_{r.m.s} = \frac{I_o}{\sqrt{2}} \quad \textbf{and} \quad V_{r.m.s} = \frac{V_o}{\sqrt{2}}$$

NOTE: These 2 formulae can only be used for sinusoidal waveforms.
Commercially, AC supply is quoted in terms of r.m.s. value.

for the specific case of a sinusoidal AC

find r.m.s value by

Steps to determine r.m.s values from AC graph

1) Square the $I-t$ or $V-t$ graph.
2) Find area under graph in 1 period.
3) Divide the area by the 1 period to get the mean-square value.
4) Square-root the mean-square value to obtain the root-mean-square value.

is related to DC by

Relationship between AC and DC

The root-mean-square value of an alternating current is that *equivalent constant direct current* that will *dissipate the same power in a given resistive load*.

$$\Rightarrow I_{r.m.s} = I_{d.c.}$$

Power dissipation →

Maximum power

$$P_{max} = I_o V_o$$

Average power

$$\langle P \rangle = I_{r.m.s} V_{r.m.s} = I_{r.m.s}^2 R = \frac{V_{r.m.s}^2}{R}$$

And for sinusoidal waveforms only,

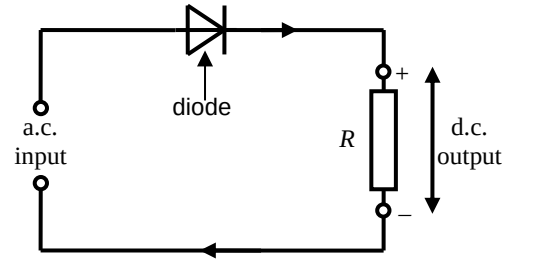
$$\langle P \rangle = \frac{I_o}{\sqrt{2}} \frac{V_o}{\sqrt{2}} = \frac{1}{2} P_{max}$$

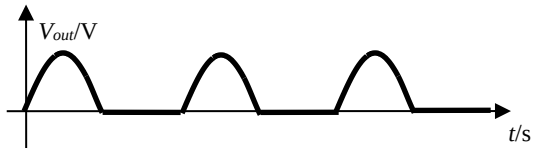
convert to d.c →

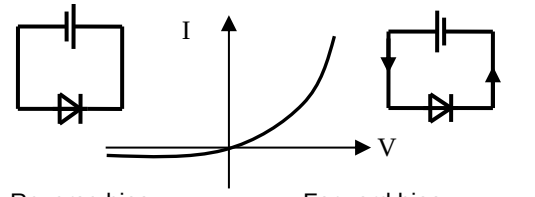
Rectification

Rectification is the conversion of a.c. to d.c.

Half-wave rectification using a diode





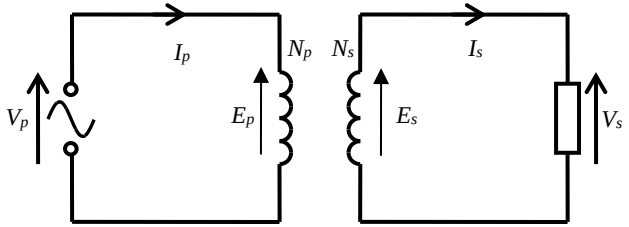


NOTE: The above graph is for the real case.
For ideal diode, we assume zero resistance in forward bias (no p.d. across diode and the entire input exists across the load) and infinite resistance in reverse bias (the circuit is 'open' and no current flows).

voltage can be stepped up or down with →

Transformer

A device that converts an alternating potential difference (or alternating current) from one value to another value. It works on principles of electromagnetic induction.



With no flux leakage, $\frac{N_s}{N_p} = \frac{E_s}{E_p}$

With no power loss,

$V_p = E_p$ and $V_s = E_s$

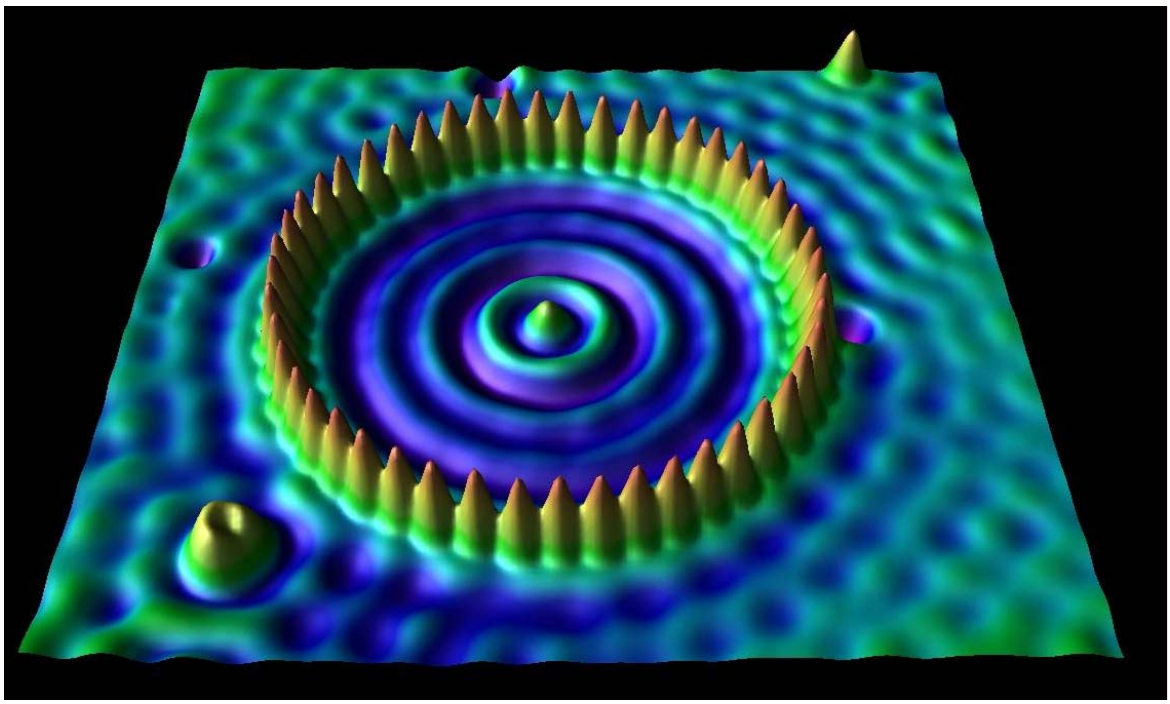
Thus, $\frac{N_s}{N_p} = \frac{V_s}{V_p}$. Also $V_p I_p = V_s I_s$ and so $\frac{N_s}{N_p} = \frac{I_p}{I_s}$

Non-ideality in transformers is due to:

a) Energy dissipated as heat due to resistance of windings in coils. Can be minimised by using thick wires.
b) Alternating magnetic flux induces eddy currents in iron core and causes heating. Can be minimized by using laminated core.
c) Hysteresis loss occurs whenever the direction of magnetic flux is reversed thus causing some energy wasted. Can be minimized by using soft iron core.
d) Flux leakage if core is badly designed.

Chapter 17

Quantum Physics



The Quantum Corral

Scientists discovered a new method for confining electrons to artificial structures at the nanometre length scale. Surface state electrons on Cu(111) were confined to closed structures (corrals) defined by barriers built from Fe adatoms. The barriers were assembled by individually positioning Fe adatoms using the tip of a low temperature Scanning Tunnelling Microscope (STM). A circular corral of radius 71.3 angstrom was constructed in this way out of 48 Fe adatoms.

This STM image shows the direct observation of standing-wave patterns in the local density of states of the Cu (111) surface. These spatial oscillations are quantum-mechanical interference patterns caused by scattering of the two-dimensional electron gas off the Fe adatoms and point defects.

Picture Source : https://www.nisenet.org/catalog/scientific-image-quantum-corral-top-view

Playlist for Lecture Examples : https://for.edu.sg/hvt0qx

Contents

Quantum Physics Syllabus **3**

17.1 The Photoelectric Effect
17.1.1 Background 4
17.1.2 The Photoelectric Effect Experiments 6
17.1.3 Einstein's Explanation on Photoelectric Effect 12
Tutorial 17A: Photoelectric Effect 19

17.2 Atomic Structure and Line Spectra
17.2.1 Bohr's Model of Atomic Quantization 25
17.2.2 Emission Line Spectra 28
17.2.3 Absorption Line Spectra 30

17.3 X-ray Spectra
17.3.1 X-ray Production 32
17.3.2 Explanation of X-ray Emission Spectra 33
Tutorial 17B: Line Spectra & X-ray Spectra 36

17.4 Wave-Particle Duality and The Uncertainty Principle
17.4.1 Introduction 42
17.4.2 de Broglie's Hypothesis – Matter Waves 42
17.4.3 Evidence of Wave-like Nature of Particles 43
17.4.4 The Heisenberg Uncertainty Principle 45
Tutorial 17C: Wave-Particle Duality and The Uncertainty Principle 46

Syllabus 9749 (H2 Physics)

Content

- Energy of a photon
- The photoelectric effect
- Wave-particle duality
- Energy levels in atoms
- Line spectra
- X-ray spectra
- The uncertainty principle

Learning Outcomes

Candidates should be able to:

(a) show an appreciation of the particulate nature of electromagnetic radiation

(b) recall and use the equation $E = hf$ for the energy of a photon

(c) show an understanding that the photoelectric effect provides evidence for the particulate nature of electromagnetic radiation while phenomena such as interference and diffraction provide evidence for the wave nature

(d) recall the significance of threshold frequency

(e) recall and use the equation $\frac{1}{2} mv_{max}^2 = eV_s$, where V_s is the stopping potential

(f) explain photoelectric phenomena in terms of photon energy and work function energy

(g) explain why the stopping potential is independent of intensity whereas the photoelectric current is proportional to intensity at constant frequency

(h) recall, use and explain the significance of $hf = \Phi + \frac{1}{2} mv_{max}^2$

(i) describe and interpret qualitatively the evidence provided by electron diffraction for the wave nature of particles

(j) recall and use the relation for the de Broglie wavelength $\lambda = h/p$

(k) show an understanding of the existence of discrete electron energy levels in isolated atoms (e.g. atomic hydrogen) and deduce how this leads to the observation of spectral lines

(l) distinguish between emission and absorption line spectra

(m) recall and solve problems using the relation $hf = E_2 - E_1$

(n) explain the origins of the features of a typical X-ray spectrum

(o) show an understanding of and apply $\Delta p \Delta x \gtrsim h$ as a form of the Heisenberg position-momentum uncertainty principle to new situations or to solve related problems.

Introduction

In the mid-19th century, classical physics (Newtonian mechanics and electromagnetic theory) dominated scientific understanding, confidently explaining many phenomena. However, by the late 19th century, it became evident that classical physics struggled to account for certain observed phenomena, including:

(1) blackbody radiation and the ultraviolet catastrophe (not in the syllabus)
(2) the photoelectric effect
(3) the Compton effect (not in the syllabus)
(4) the spectrum of bright lines from gas discharge tubes.

In response, several eminent scientists embarked on a quest to understand these anomalies which lead to the development of quantum theory. This revolutionary theory, spearheaded by eminent figures like Max Planck, Albert Einstein, and Niels Bohr, introduced the concept of discrete energies and fundamental constants such as the Planck constant. Their seminal contributions not only earned them Nobel Prizes[1] but also lay the groundwork for the development of quantum mechanics.

The impact of quantum theory extends far beyond theoretical physics, underpinning modern technologies like lasers and nanotechnology. As we delve into topics in quantum mechanics, it is crucial to appreciate the profound influence of these concepts on our understanding of the universe and our technological advancements.

17.1 The Photoelectric Effect

17.1.1 Background

In 1887, Heinrich Hertz discovered the **photoelectric effect** – electrons were emitted when he shone UV light on a freshly polished zinc plate.

The **photoelectric effect** is a phenomenon in which **electrons** are liberated from a **cool metal surface** when **electromagnetic radiation** of sufficiently high **frequency** is incident upon it.

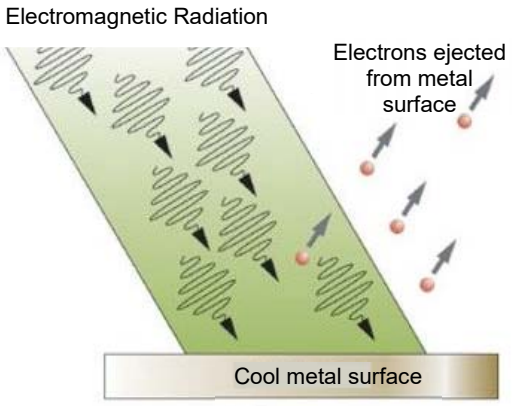


Fig. 1

[1] Max Plank received his Nobel Prize in 1918, for resolving the issue of blackbody radiation and the ultraviolet light catastrophe by introducing the concept of discrete energies (quantum theory) and a crucial constant *h*, now known as Planck constant. Albert Einstein received his Nobel Prize in 1921 for using Planck's proposal of discrete energies for electromagnetic radiation to successfully explain the photoelectric effect. Niels Bohr was honoured with a Nobel Prize in 1922 for explaining the spectrum of bright lights using the quantum theory of light.

The mere existence of the photoelectric effect is not, as is sometimes assumed, a difficulty for classical physics.

Electrons that are the charge carries in the metals which are "able to move freely" are still bound inside the metal and do not spontaneously spill out at room temperature. A minimum amount of energy ϕ is still needed to free an electron from the metal's crystal lattice. Some electrons are deeper than others in the lattice structure of the metal and may require more energy than ϕ to escape from the surface of the metal, but all will require *at least* ϕ. This minimum amount of energy ϕ is known as the **work function** of the metal. Different metals have different work functions and work functions are characteristic of the metal.

Work Function ϕ :

The **work function** is the minimum amount of energy required to remove an electron from the metal surface (i.e., the energy required to remove the least strongly held electron).

One way to liberate electrons is through a process called *thermionic emission* where the metal is heated to high temperatures. Heating a metal, increases the thermal energy of the electrons. At a sufficiently high temperature, the kinetic energy of a small percentage of the electrons may exceed the work function. These electrons can then leave the metal.

Classical physics explains photoelectric effect in a way that is similar to thermionic emission. When light is shone on the surface, the electromagnetic waves spread out across the area of illumination and heat the electrons and increases their kinetic energy which is transferred to the crystal lattice, via collisions. But if the light is sufficiently intense, the electron's kinetic energy may be sufficiently high to cause the emission of the electrons. Hence in the 1900s, scientists thought it was plausible that an intense light source might be able to cause the thermal emission of the electrons.

However, in 1902, Philipp Lenard conducted a series of detailed experiments to explore the photoelectric effect further. It was his findings that presented a significant challenge to classical physics and the prevailing wave theory of light. It was not until 1905 that Albert Einstein provided a groundbreaking explanation to the observations. Einstein's revolutionary proposal, which introduced the particulate nature of light, not only resolved the discrepancies observed in the photoelectric effect experiments but also laid the foundation for the development of quantum mechanics.

17.1.2 The Photoelectric Effect Experiments

17.1.2.1 The Experimental Setup

We start off our exploration on photoelectric effect by looking at a typical experimental setup (Fig. 2) which was used by Lenard in his experiments.

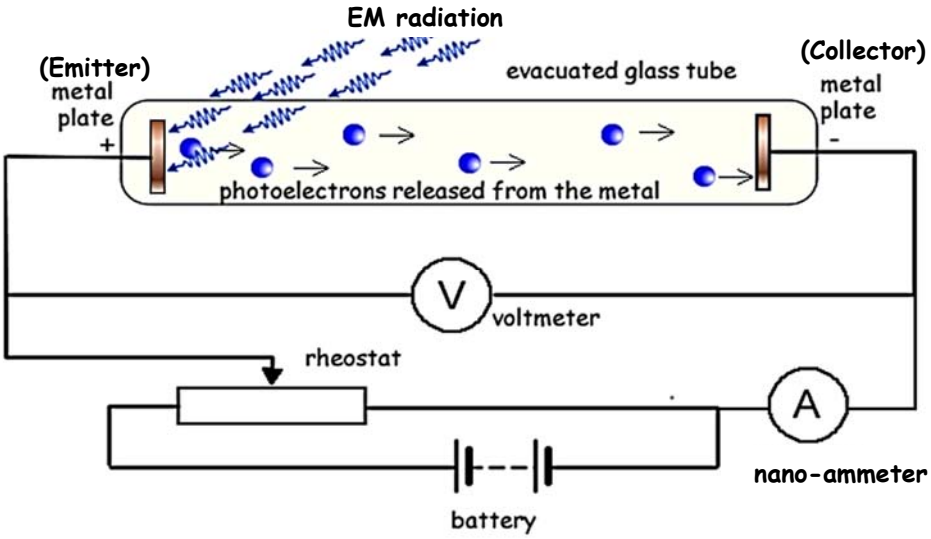


Fig. 2 Experimental set-up

Two metal plates are located at two ends of an evacuated tube. The potential difference between the plates can be varied by the rheostat, while the polarity can be changed by changing the connections across the battery.

Light of various **frequencies** and **intensities** is shone on one of the metal plates - the emitter which is often also called the cathode. In doing so, electrons may be ejected from the emitter and travel to the other metal plate - the collector which is often called the anode.

If electrons reach the collector, a current (called the photocurrent) will be registered and measured by the ammeter as shown in the diagram.

The ejected electrons are called **photoelectrons**.

An explanation of the photoelectric effect can also be found in the following video:
Photoelectric Effect Explanation https://www.youtube.com/watch?v=Ef3f3AWXMmk

17.1.2.2 Experiment 1: Varying the Applied Voltage

With the understanding of the setup, we can now explore the experiments that were conducted by Lenard.

In this first experiment, a monochromatic source of e.m. radiation of a sufficiently high frequency and constant intensity is shone on the metal plate (emitter) such that photoelectrons are liberated from the emitter. The photoelectrons will liberate from the emitter with a range of kinetic energies and directions.

Case 1 : When V =0

When the potential difference V across the collector and emitter is set to zero. A current is detected by the ammeter.

This current may not be the maximum possible current as the photoelectrons are emitted with various speeds in all directions, and some reaches the collector but some will not.

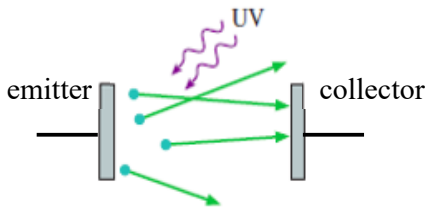


Fig. 3.1. V= 0 : The photoelectrons leave the emitter in all directions. A few reach the anode.

Case 2 : When V > 0

When the collector is set at a **higher** potential than the emitter, the photoelectrons will feel an electric force towards the collector (Fig. 3.2 (a)), more electrons will reach the collector - the current is higher than before.

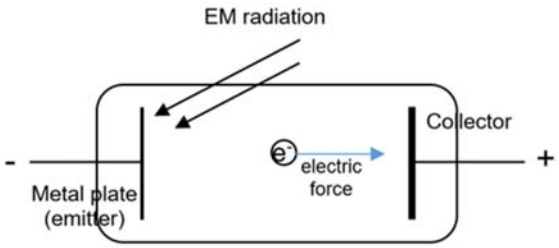


Fig. 3.2 (a)

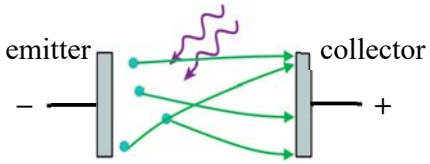


Fig. 3.2 (b). V > 0 : A higher potential at the collector attracts the photoelectrons to the collector

- **Saturation Current, I_s**

  As V is increased further, there will come a time where the current I reaches a maximum. This is because all the emitted photoelectrons are collected at the collector i.e. rate of emission of photoelectrons is equal to rate of photoelectrons reaching the collector. Increasing V further does not increase I.

  This maximum current is known as the **saturation current** I_s.

  Note: In many A-Level questions, the saturation current is reached at V = 0. This is possible - it happens when all the emitted photoelectrons are reaches the collector even without a potential difference applied across the plates.

- **More on Saturation Current:**

  Since saturation current happens when **all the electrons emitted are collected at the collector**,

  I_s = Charges emitted per unit time = (No. of electrons emitted per unit time)(e)

  Hence, the no. of electrons emitted per unit time, $\frac{N_{electrons}}{t} = \frac{I_s}{e}$

Case 3 : When $V < 0$

When the potential difference is adjusted such that collector is at a **lower** potential than the emitter, the photoelectrons will feel an electric force away from the collector (Fig. 3.3 (a)). However, the more energetic electrons can still reach the collector. Hence, a current can still be detected by the ammeter if the potential difference is not large enough.

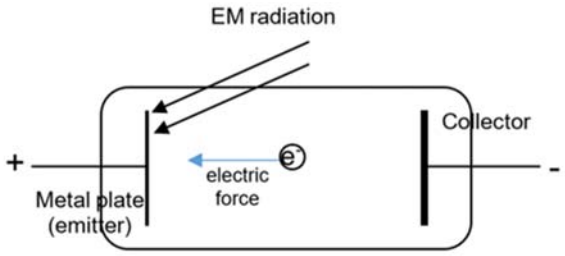


Fig. 3.3 (a). Photoelectrons experience an electrical force that repels them from the collector when the collector is set at a lower potential than the emitter.

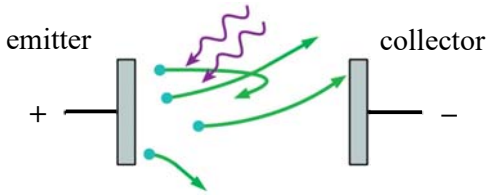


Fig. 3.3 (b). $V < 0$: A lower potential at collector repels the electrons from the collector. Only the electrons which are emitted with greater kinetic energies are able to reach the collector.

- ### Stopping Potential V_s

  If, however, the potential of the collector is continually lowered, there comes a point where the current will fall to zero. This corresponds to the potential difference whereby even the **most energetic** photoelectrons cannot reach the collector.

  This magnitude of potential difference across the collector and the emitter at this point is called the **stopping potential** V_s.

  Lowering the potential of the collector further therefore yields zero current as there are no more photoelectrons reaching the collector.

- ### Results of the Three Cases Represented in *I-V* Graph

  Fig.4 summarises the above observations in an *I-V* graph

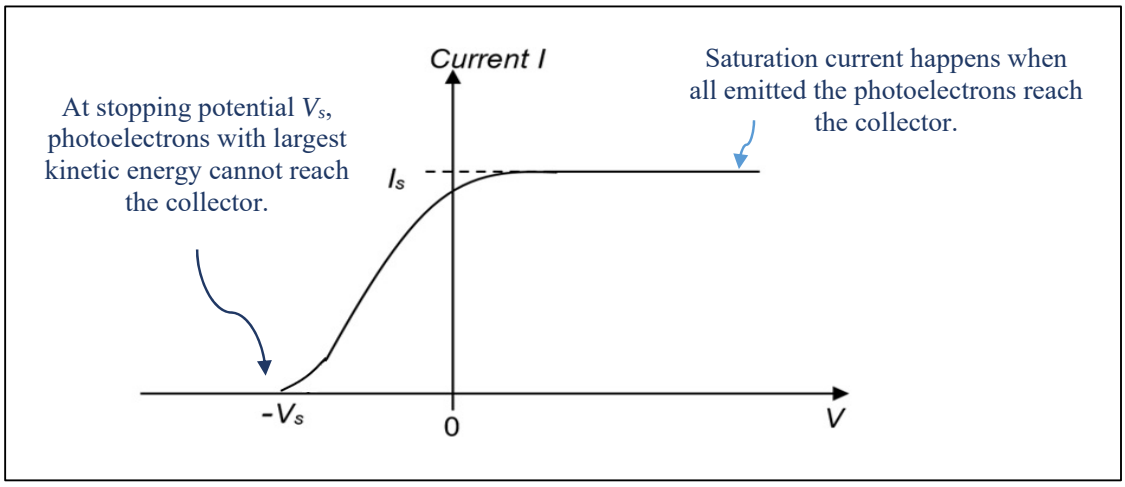


Fig. 4

- **Relationship between Stopping Potential V_s and the Maximum Kinetic Energy of the Photoelectrons**

Recall that at the stopping potential V_s , even the photoelectrons that are liberated with the most kinetic energy (KE_{max}) will not be able to reach the collector plate. This happens because the kinetic energy of these electrons converts to electrical potential energy just before the electrons reaches the collector.

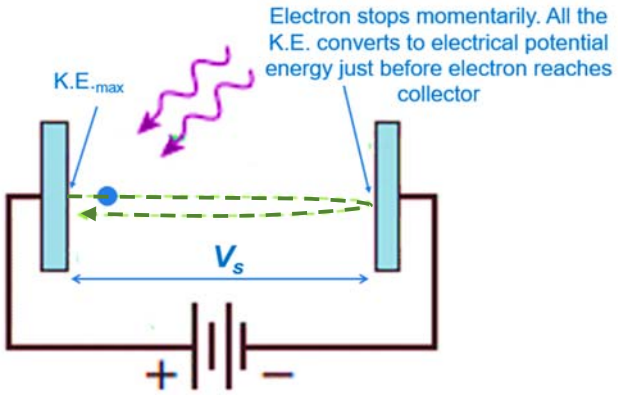


Fig. 5

Applying conservation of energy to the system of the field and these photoelectrons, we can write

$\Rightarrow$ Initial kinetic energy + Initial electrical potential energy

= Final kinetic energy + Final electrical potential energy

$\Rightarrow KE_{max} + (- e V_i) = 0 + (- e V_f)$

$\Rightarrow KE_{max} = e (V_i - V_f) = eV_s$

since $(V_i - V_f)$ is the potential difference between the collector and emitter plates and it is at a value of the stopping potential V_s.

Hence, we see that the stopping potential V_s is a measure of the maximum kinetic energy KE_{max} of the emitted photoelectrons and relationship between them is

$$KE_{max} = eV_s$$

In the setup, the adjustable DC supply is used as a means to determine the *KE* possessed by the most energetic photoelectrons (i.e., KE_{max}) when they are first ejected from the metal surface. The applied potential difference *V* does not alter the KE_{max} in any way.

Now that we have a way of measuring KE_{max} of the photoelectrons, we can investigate how KE_{max} changes with other factors like the intensity and frequency of the incident light. This will be the focus of the next two experiments.

17.1.2.3 Experiment 2: Varying the Intensities of Light

One of the experiments Lenard did was to repeat experiment 1 for different intensities of the incident light of same frequency. The following results were obtained. For each result, we will compare the predictions made by classical approach, using the wave theory of light, with the actual experimental observation.

(1) Saturation current I_s increases with light intensity.

- **Classical Prediction:** With more intense light falling on the emitter, more power per unit area of light is supplied to the electrons on the emitter plate. Therefore, more electrons can be emitted per unit time and the saturation current should increase.
- **Experimental Result:** With more intense light falling on the emitter, indeed, more photoelectrons were emitted per unit time and the saturation current was increased. If the light intensity was doubled, the saturation current I_s doubled.

(2) Maximum kinetic energy of photoelectrons is independent of the intensity of light.

- **Classical Prediction:** Electrons should absorb energy continuously from the electromagnetic waves. As the light intensity incident on a metal is increased, power per unit area incident on the electrons in the metal and the electrons should be ejected with greater maximum kinetic energy KE_{max}. This should manifest as a larger stopping potential V_s on the *I-V* graph.
- **Experimental Result:** The maximum kinetic energy of photoelectrons is found to be independent of light intensity. No matter now weak or how strong the intensity of light was, the stopping potential V_s was observed not change with an increased in the intensity of light.

Fig. 6.1 and Fig. 6.2 compares the actual experimental results with the classical prediction in an *I-V* graph.

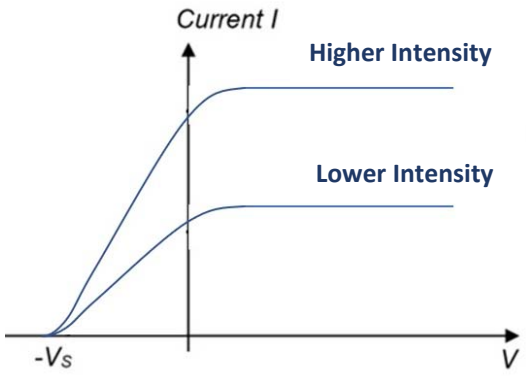


Fig. 6.1. Actual Experimental Results. When light intensity increases, the saturation current increases but stopping potential stays the same.

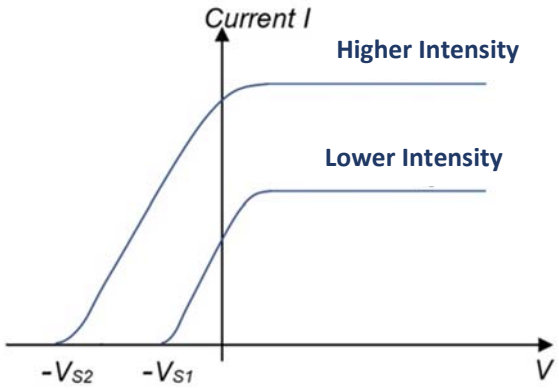


Fig. 6.2. Classical Prediction. Wave theory predicts that when light intensity increases, the saturation current increases but stopping potential should also have increased.

17.1.2.4 Experiment 3: Varying the Applied Voltage for Different Frequencies of Light

Lenard also performed experiments by shining lights of different frequencies on the emitter plate. The following results were obtained.

(3) Almost instantaneous emission of photoelectrons.

- **Classical Prediction:** Because the light energy is distributed over the area where the light is incident on, at very low light intensities, a measurable time interval should pass between the instant the light is incident on the emitter and the time an electron is emitted from the metal.

  This time interval is required for the electron to absorb the incident radiation before it acquires enough energy to escape from the metal.

- **Experimental Result:** Electrons are emitted from the surface of the metal almost ***instantaneously*** (within ~ 1 ns) after the surface is illuminated, even at very low light intensities.

(4) Existence of Threshold Frequency of Light.

- **Classical Prediction:** Electrons should be ejected from the metal at any incident light frequency, as long as the light intensity is high enough, because energy is transferred to the metal regardless of the incident light frequency.

- **Experimental Result:** For any given metal, electrons are emitted only when the frequency of incident light is above some minimum value f_o. It is found that If the frequency of light is below f_o, regardless of how intense the light is no electrons are emitted. This frequency is know as the **threshold frequency** and is characteristic of the metal being illuminated.

Threshold frequency f_o :

The threshold frequency is the **minimum frequency** of the electromagnetic radiation required to liberate photoelectrons from a metal surface.

(5) Dependence of maximum kinetic energy of photoelectrons on light frequency

- **Classical Prediction:** There should be *no relationship* between the frequency of the light and the maximum kinetic energy of the photoelectrons. The maximum kinetic energy should be related to the intensity of the light instead.

- **Experimental Result:** The stopping potential V_s and hence the maximum kinetic energy KE_{max} of the electrons was found to increase with increasing light frequency. (Fig. 7)

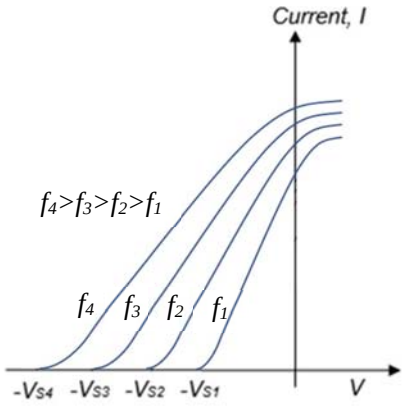


Fig. 7. The stopping potentials and hence KE_{max} of the photoelectrons increase with with higher frequencies of light.

From the experimental results above, you can see that although classical physical was able to offer an explanation for the existence of the photoelectrons, it was unable to explain the observed behaviors of the photoelectrons in the experiments.

17.1.3 Einstein's Explanation for Photoelectric Effect

The Quantum Theory of Light – the Photon

A successful explanation of the observations of the photoelectric effect was finally given by Einstein in 1905, the same year he published his special theory of relativity.

As part of a general paper on electromagnetic radiation, for which he received a Nobel Prize in Physics in 1921, Einstein extended Max Planck's concept of quantization to electromagnetic waves.

Einstein suggested that light is not really a continuous wave but, instead, arrives in small packets or bundles of energy. Einstein called each packet of energy a **light quantum** and he postulated that the energy of one light quantum to be directly proportional to the frequency of the light. We now call these light quanta **photons**.

Photon :

A **photon** is a quantum of electromagnetic radiation. Its energy E is given by

$$E = hf$$

where h is the Planck constant ($h = 6.63 \times 10^{-34}$ J s) and f the frequency of the light.

Since each photon travels at the speed of light $c = f\lambda$, where λ is the wavelength. E can also be written as

$$E = h\frac{c}{\lambda}$$

An Alternative Unit for Energy - eV

The magnitude of the energies involved is in this topic are often on the atomic scale and very small compared to everyday applications. Hence, we find it convenient to use an alternative unit of energy, called the **electron volt** (eV).

The electron volt is defined as the energy the charge-field system gains or loses when a charge of magnitude e is moved through a potential difference of one volt. In other words, when a particle of charge e moves through a potential difference of 1 V, the change in the potential energy is 1 eV.

$$1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$$

An introduction to the Quantum Theory of Light can be found in the following YouTube video:
The Origin of Quantum Mechanics (feat. Neil Turok)
https://www.youtube.com/watch?v=i1TVZIBj7UA

The following video attempts to explain what exactly a photon is, besides a packet of energy.
What the HECK is a Photon? https://www.youtube.com/watch?v=73XdlgzSKws

Example 1

The eye is most sensitive to light of wavelength 550 nm. What is the energy of photons with this wavelength? Express your answer in electron volts (eV).

Example 2

A source emits monochromatic light of wavelength 550 nm at a power of 1.25 W. How many photons are emitted in one second?

Einstein's Postulates

To explain the findings in the photoelectric effect experiments by Lenard, Einstein postulated the following:

(1) Light of frequency *f* consists of discrete quanta, known as photons, each of energy *E = hf.*

Each photon travels in vacuum at the speed of light *c*.

(2) A photon of the incident light gives all its energy *hf* to a single electron in the metal.

Therefore, the absorption of energy by the electrons is not a continuous absorption process as envisioned in the wave model; rather; it is a discontinuous process in which energy is delivered to electrons in discrete bundles. The energy transfer is accomplished via a one photon-one electron event.[2] An electron in the metal must absorb the entire quantum of light at once, it cannot absorb two or more photons. Nor can it absorb a fraction of a quantum.

Einstein's Photoelectric Effect Equation

Electrons ejected from the surface of the metal and not making collisions with the other metal atoms before escaping will possess the maximum kinetic energy KE_{max} upon liberation.

Recall that the minimum energy required to free an electron from its bond to the metallic lattice is known as the work function ϕ. Hence, by conservation of energy, for those photoelectrons that are liberated with maximum kinetic energy KE_{max}

Energy of photon absorbed = Work function of metal + maximum kinetic energy of photoelectrons

$\Rightarrow hf = \phi + KE_{max}$

Rearranging this gives **Einstein's photoelectric effect equation:**

$$KE_{max} = hf - \phi$$

- This equation is **only valid for the photoelectrons that are liberated with maximum kinetic energy**. They come from the surface of the metal. Due to collisions with other atoms, those below the surface emerge with smaller kinetic energy.

- In words, the maximum kinetic energy KE_{max} of the photoelectrons is equal to the energy gained by absorbing a single photon hf minus the work function energy ϕ (i.e. work done to escape the metal).

[2] In principle, an electron can absorb two photons. However, the probability of it happening is very low.

How Einstein's "Photon" Model Explains the Observations of the Photoelectric Effect

Using the photon model of light, one can now explain the observed features of the photoelectric effect that are incorrectly predicted by classical concepts. We will now relook at each of the observations above that wave theory of light failed to explain.

- **Observation 1 : Maximum kinetic energy of photoelectrons is independent on the intensity of light.**
  - **Wave Theory:** When light intensity increases, there is more energy incident on the electrons, the maximum kinetic energy of the photoelectrons should have been higher too.
  - **Photon Model:** A more intense light means *more photons,* but the energy of each photon is still the same. These additional photons eject a greater number of photoelectrons and cause the larger saturation current. However, since the energy of each photon is still the same and the interaction between the photon and electron is one to one, hence the maximum kinetic energy of the photoelectrons remains the same.

- **Observation 2 : Almost instantaneous emission of photoelectrons.**
  - **Wave Theory:** For low intensities of light, a measurable time interval should have passed between the light is incident on the emitter and the time an electron is emitted as time is need for the electron to absorb the incident radiation.
  - **Photon Model:** The near instantaneous emission of photoelectrons observed is consistent with the photon model of light. Since each photon transfers its energy to just one electron, if that electron immediately has energy to escape, the current should begin instantaneously. If the incident light has very low intensity, it just means that there are very few photons arriving per unit time. The interaction between the photon and electron is still one-to-one and instantaneous.

- **Observation 3 : Existence of threshold frequency of light.**
  - **Wave Theory:** If the intensity of the incident light is sufficiently high, electrons should have been ejected regardless of the frequency. There should not be a threshold frequency.
  - **Photon Model:** To be liberated from a metal, an electron needs to absorb a photon of energy that is at least the work function ϕ of the metal. Because the energy of each photon is given by hf where f is the frequency of the photon, the photoelectric effect therefore cannot be observed if the frequency of the photon is below a minimum value – the threshold frequency f_o . If the frequency of light is below this threshold frequency, an electron cannot be ejected from the surface, even if many photons per unit time are incident in a very intense light beam.

Relationship Between threshold frequency f_o and work function ϕ of the emitter metal :

$$\phi = hf_o$$

This also explains why for different metals (hence different work functions), the threshold frequency of the incident light differs.

- **Observation 4 : Dependence of maximum kinetic energy of photoelectrons on light frequency**
  - **Wave Theory:** There should be *no relationship* between the frequency of the light and the maximum kinetic energy of the photoelectrons.
  - **Photon Model:** A photon of higher frequency carries more energy and therefore ejects a photoelectron with greater maximum kinetic energy when incident on the same metal (hence, same work function energy) than does a photon of lower frequency.

More on Einstein's Photoelectric Effect Equation – Graphical Representation

Not only do Einstein's postulates explain all of Lenard's observations, they also made a new prediction – there is a linear relationship between the maximum kinetic energy of the photoelectrons with the frequency of light the metal is illuminated with since,

$$KE_{max} = hf - \phi$$

where ϕ is the work function of the metal that is illuminated.

Lenard had not verified this relationship for different light frequencies, so Einstein offered this as an untested prediction of his postulates. Robert Millikan subsequently took up the challenge and measured the stopping potentials for various metals and validated this relationship.

Fig. 8 shows the graphs of KE_{max} for photoelectrons *vs.* frequency f of incident light for a typical photoelectric effect experiment with different metals.

Key Features of the Graphs

From the equation, $KE_{max} = hf - \phi$,

1. **Gradient:**

   Gradient of each graph = Planck constant h. This is why all the graphs are parallel to each other.

2. **Vertical intercept:**

   Vertical intercept = $-\phi$ where ϕ is the work function of each metal.

   Since the work function is a characteristic of the metal illuminated, it would be unique for each metal.

   Note that the work function is usually given as a positive number. So, for example, if the vertical intercept is -2.80 eV, the work function = 2.80 eV.

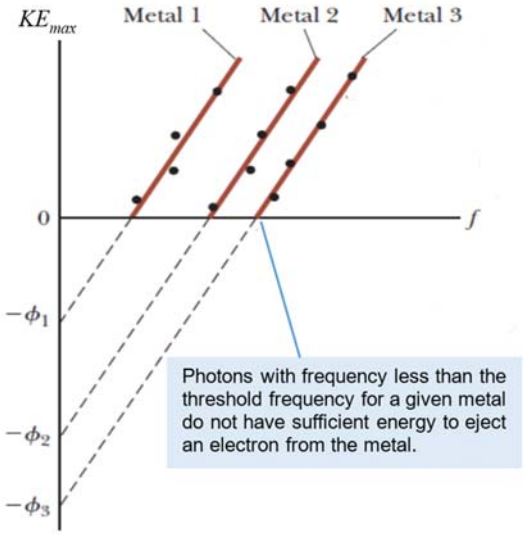


Fig. 8. KE_{max} vs. f graph for different metals.

3. **Horizontal intercept:**

   The horizontal intercept happens when the maximum kinetic energy KE_{max} of the photoelectrons is zero, the energy hf_0 of the incident photon must be equal to the work function ϕ of the metal:

$$hf_o = \phi \Rightarrow f_o = \frac{\phi}{h}$$

   Hence, the horizontal intercept of each graph gives the threshold frequency of the metal. This threshold frequency is also equals to $\frac{\phi}{h}$. For different metals, the work functions are different, hence the threshold frequency is different.

Hence, although Einstein's idea that light comes in discrete light quanta that cannot be divided, and that the energy of one quantum is delivered to one electron seems trivial. They were not only able to explain the observations of the photoelectric experiments that classical physics failed to explain, they are crucial for understanding the new ideas that will lead to quantum physics for they are completely at odds with the concepts of classical physics, where energy can be continuously divided and shared.

Example 3 (N84/II/14)

A student connects a photocell to a supply and finds that when the cell is exposed to monochromatic radiation, a current flows only when the potential difference across the cell is less than 1.6 V. What is the maximum energy of the emitted electrons?

Example 4 (N90/I/29)

Light quanta of energy 3.5×10^{-19} J fall on the cathode of a photocell. The current through the cell is just reduced to zero by applying a stopping potential of 0.25 V. What is the work function energy of the cathode?

Example 5

In photoemission from a given material, when the wavelength of the incident light is increased, the

A maximum kinetic energy of the emitted electrons increases.

B minimum kinetic energy of the emitted electrons deceases.

C minimum kinetic energy of the emitted electrons increases.

D maximum kinetic energy of the emitted electrons decreases.

Example 6 (N89/I/29)

In an experiment on the photoelectric effect, different metals were illuminated with monochromatic light of variable frequency and variable intensity. It was found that, for each metal, electrons were emitted with a spread of kinetic energies up to a maximum value. This maximum kinetic energy depends on

A the metal but not on the frequency or the intensity.

B the metal and the intensity but not on the frequency.

C the metal and the frequency but not on the intensity.

D the frequency and the intensity but not on the metal.

E the metal, the frequency and the intensity.

Example 7 (N91/I/28)

An ultra-violet source of radiation causes the emission of electrons from a zinc plate. What would be the effect of a more intense source of the same wavelength on the maximum energy per electron, and on the number of electrons emitted per second?

	Max. energy per electron	*No. of electrons emitted per second*
A	less	more
B	the same	the same
C	the same	more
D	more	the same
E	more	more

Tutorial 17A Photoelectric Effect

Self-Review Questions

S1 Three energies are listed below.

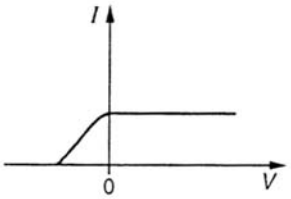
1. The energy of a photon of a 3 m wavelength radio wave.
2. The energy of an X-ray photon.
3. The energy of a photon of yellow light from a sodium lamp.

Arrange the above energies in the order of increasing magnitude.

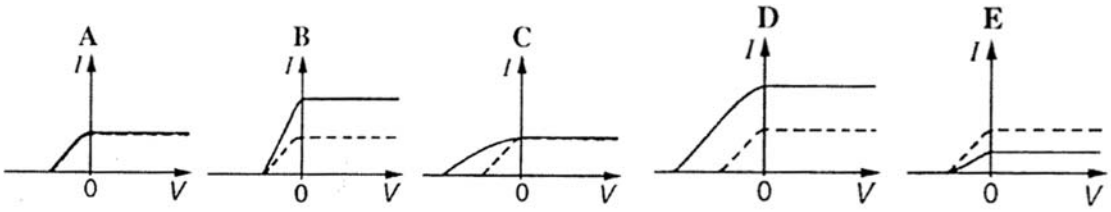
S2 **(a)** State what is meant by the *photoelectric effect*.

(b) Give three of the experimental observations associated with this effect. How do these observations support the idea that light is quantised?

S3 A metal surface in an evacuated tube is illuminated with monochromatic light causing the emission of photoelectrons, which are collected at an adjacent electrode. For a given intensity of light, the way in which the photocurrent *I* depends on the potential difference *V* between the electrodes is as shown in the diagram on the right.



If the experiment were repeated with light of twice the intensity but the same wavelength, which of the graphs below would best represent the new relation between *I* and *V*? (In these graphs, the result of the original experiment is indicated by a broken line.)



UCLES N86/I/28

S4 A beam of monochromatic radiation falls on to a metal X and photoelectrons are emitted. The rate of emission of electrons will be doubled if

A a beam of double intensity is used.

B radiation of double the frequency is used.

C radiation of double the wavelength is used.

D a metal with a work function half that of X is substituted for X.

CIE J82/II/4

S5 If photons that cause photoelectrons to be released all have the same energy, how can the photoelectrons have different energies?

S6 A student claims that he is going to eject electrons from a piece of metal by placing a radio transmitter antenna adjacent to the metal and sending a strong radio signal into the antenna. The work function of a metal is typically a few electron volts.

Will this work?

S7 The equation $hf = \phi + \frac{1}{2}mv_{max}^2$ is used when studying the photoelectric effect.

What is the meaning of each term in this equation?

	hf	ϕ	$\frac{1}{2}mv_{max}^2$
A	the energy of a photoelectron	the least energy required to release an electron	the maximum kinetic energy of a photon
B	the energy of a photoelectron	the work done by an incoming photon	the maximum kinetic energy of a photoelectron
C	the energy of a photon causing the photoelectric effect	the least energy required to release an electron	the maximum kinetic energy of a photoelectron
D	the energy of a photon causing the photoelectric effect	the work done by an incoming photon	the maximum kinetic energy of a photon

UCLES N2012 P1 Q37

S8 Electrons are only just emitted from a metal surface when illuminated with light of frequency 4.0×10^{14} Hz. When illuminated with light of frequency f, the stopping potential is found to be 1.5 V.

What is the value of f?

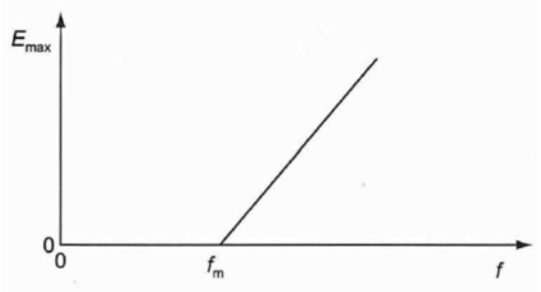
A 3.8×10^{13} Hz **B** 3.6×10^{14} Hz **C** 4.2×10^{14} Hz **D** 7.6×10^{14}Hz

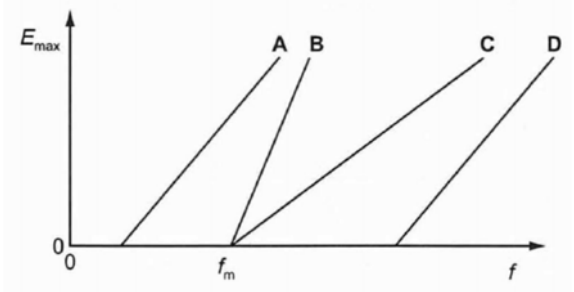
UCLES H1 2010 P1 Q30

S9 The graph on the left shows the variation of the maximum kinetic energy E_{max} with frequency f, for photoelectrons emitted from the surface of a metal illuminated with electromagnetic radiation.

The metal is replaced with one that has a smaller work function.

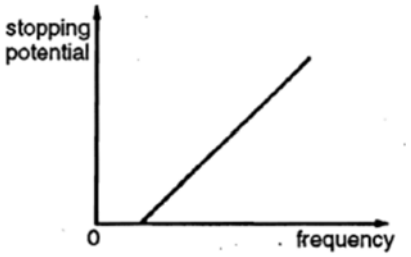
Which line shows the variation of the maximum kinetic energy with frequency for the new metal?





UCLES 2009 P1 Q36

S10 The result of an experiment to investigate the energy of photoelectrons emitted from a metallic surface is represented by the figure.



The gradient of the graph depends on the

A intensity of the incident radiation.

B wavelength of the incident radiation.

C work function of the irradiated surface.

D ratio of the Planck constant to the electronic charge.

CIE J76/II/14

Discussion Questions

D1 A laser emits light of power *P*. The light consists of photons of frequency *f*. The Planck constant is *h* and the speed of light is *c*.

How many of these photons are contained in one metre length of the laser beam?

A $\frac{P}{c}$ **B** $\frac{P}{hf}$ **C** $\frac{Pc}{hf}$ **D** $\frac{P}{chf}$

UCLES H1 N2008 P1 Q30

D2 A vacuum photo-emissive cell is connected in the circuit shown in Fig. 3(a).

The emitter is illuminated with monochromatic radiation of wavelength 365 nm and the current *I* in the circuit is measured for various values of the applied potential difference *V* between collector and emitter. The results are shown in the graph in Fig. 3(b).

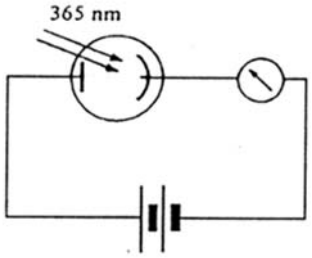


Fig. 3(a)

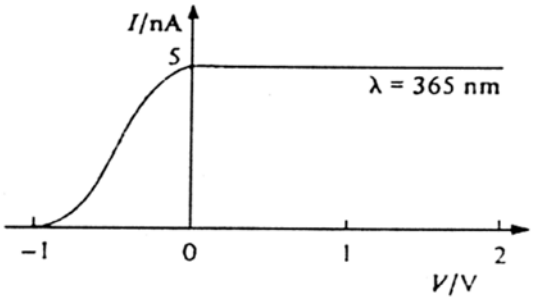


Fig. 3(b)

(a) **(i)** The graph eventually becomes horizontal for larger value of $V > 0$ V. Shouldn't the current increase as the potential difference increases? Explain.

(ii) The current does not immediately fall to zero for $V < 0$ V. Shouldn't $V < 0$ V prevent the electrons from reaching the collector? Explain.

(iii) The current is zero for $V < -1$V. Where do the photoelectrons go? Are no photoelectrons emitted? Or if they are, why is there no current? Explain.

(b) The intensity of the incident radiation is 8.2×10^{-3} W m^{-2}. The area of the metal surface illuminated is 2.0 cm^2. Calculate

(i) the rate (in s^{-1}) at which the photons are incident on the emitter.

(ii) the maximum rate (in s^{-1}) at which electrons leave the emitter.

(c) Comment on, and suggest an explanation for, any difference between your answers to **(b)(i)** and **(b)(ii)**.

(d) Find the maximum kinetic energy of the photoelectrons.

(e) Deduce the work function of the emitter. (Express your answer in electron volts.)

(f) Copy Fig. 3(b), then draw and label graphs for the following three situations:

(i) the intensity of electromagnetic radiation is halved but its wavelength is kept constant.

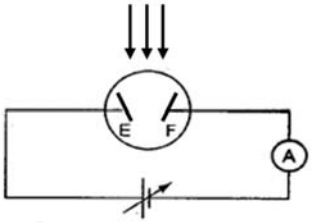
(ii) the experiment was repeated using monochromatic radiation of wavelength 313 nm.

(iii) the emitter work function is increased.

In each case, no other parameters of the experiment are changed. Explain the shape of each of your new graph.

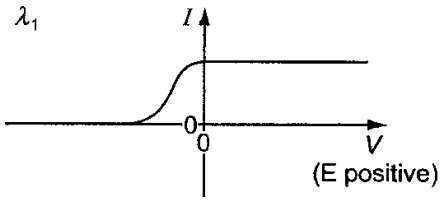
UCLES N83/I/6 (modified), J97/II/6 (modified); Knight (4E) Workbook

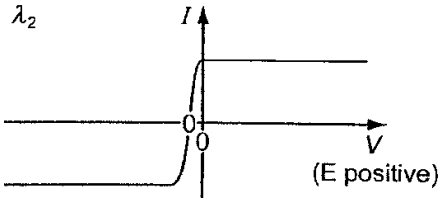
D3 The diagram on the right shows a circuit used for photoelectric emission experiments.



The two electrodes E and F are made of different metals. The work function of electrode E is ϕ_E and the work function of electrode F is ϕ_F.

Current-voltage (*I*-*V*) characteristics are obtained when both electrodes are separately illuminated with monochromatic light of different wavelengths, λ_1 and λ_2, as follows:





Explain whether

(a) λ_2 is greater or less than λ_1,

(b) ϕ_E is greater or less than ϕ_F.

UCLES Nov 2007 P1 Q36 modified

D4 The graph drawn in Fig. 6.1 shows how the maximum kinetic energy E_k of the photoelectrons from a particular material varies with the frequency f of the electromagnetic radiation that causes the emission of photoelectrons for an experiment.

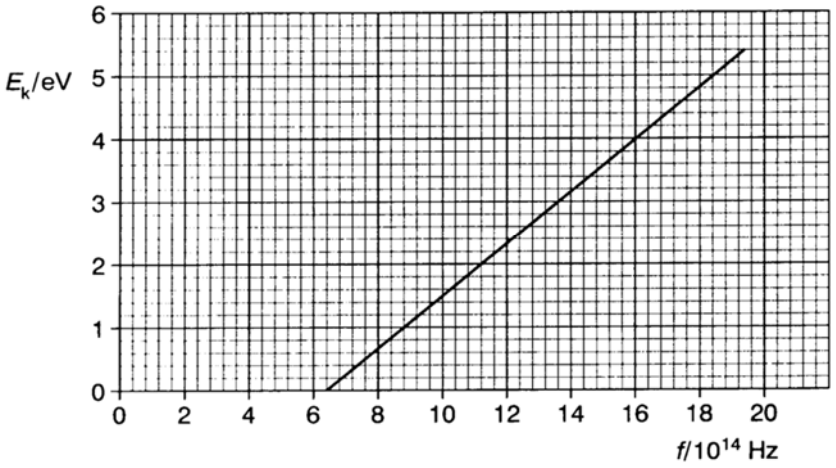


Fig. 6.1

(a) Use the graph to determine

(i) the threshold frequency for this material,

(ii) the maximum kinetic energy of photoelectrons from this material when it is illuminated with electromagnetic radiation of frequency 18.0 x 10^{14} Hz. Give your answer in joules.

(iii) the experimental value of the Planck constant for this experiment.

(b) On Fig. 6.1, draw and label graphs for the following two situations:

(i) the light intensity is doubled,

(ii) the material is replaced by another one with a smaller work function energy

In each case, no other parameters of the experiment are changed. Explain the shape of your new graphs.

(c) As the maximum kinetic energy E_K is related to the stopping potential V_s and the frequency f of the incident photons is also related to the wavelength λ of the photon, there can be a few variations of the above graph that can be plotted.

For each of the graph below, state the significance of the gradient and vertical intercept.

(i) Graph of V_s vs. f.

(ii) Graph of E_K vs. $1/\lambda$.

(iii) Graph of V_s vs. $1/\lambda$.

UCLES Nov 2000 P3 Q6 modified. Knight (4E) Workbook.

D5 Europium is a metal with a work function energy of 2.5 eV.

(a) Light of wavelength 490 nm is incident on the europium surface.

Calculate the maximum speed of the emitted electrons. [4]

(b) The europium is placed on a gold-leaf electroscope as illustrated in Fig. 4.2.

(i) When the europium is charged negatively, the leaf rises.

Explain why the leaf falls when the europium is illuminated with blue light but not when illuminated with red light. [1]

(ii) When the europium is charged positively, the leaf rises.

Explain why the leaf will not fall no matter what wavelength of light is used to illuminate the surface of the europium. [2]

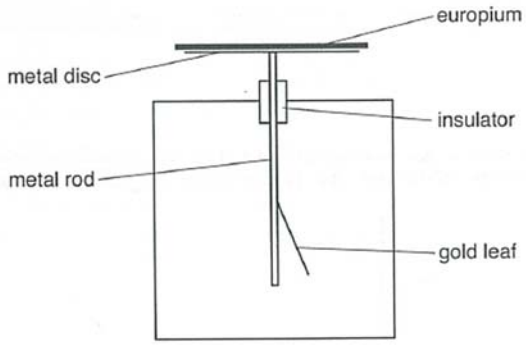


Fig. 4.2

UCLES N2018 P2 Q4(c)

D6 One of the most useful applications of the photoelectric effect is the charge-coupled device (CCD). An array of these devices is used in digital cameras instead of film to capture pictures by emitting photoelectrons. Each pixel captures a small part of a picture.

(a) Calculate the energy of a photon of visible light of wavelength 500 nm.

(b) When visible light of wavelength 500 nm is incident on a pixel, electrons of maximum kinetic energy 1.99 eV are released. Determine the minimum frequency of light necessary for electrons to be released.

(c) Visible light of 500 nm was captured by a pixel array over an exposure time of 0.34 ms. Given that the intensity of light was 1.0 W m^{-2} and a pixel is 20 μm x 20 μm, estimate the rate of photons incident on each pixel.

HCI 2003 C2 BT2 P2, Q4 modified

D7 An isolated copper sphere of radius 5.00 cm, initially uncharged, is illuminated by ultraviolet light of wavelength 200 nm.

Given that the work function for copper is 4.70 eV, what charge will the photoelectric effect induce on the sphere?

Answers to Tutorial 17A Photoelectric Effect

Self-Review Questions:

S1 1, 3, 2 **S3** B **S4** A **S7** C **S8** D

S9 **A** **S10** D

Discussion Questions:

D1 D

D2(b)(i) 3.01 x 10^{12} s^{-1}, **(ii)** 3.125 x 10^{10} s^{-1}, **(d)** 1.60 x 10^{-19} J, **(e)** 2.41 eV

D3 λ_2 is less than λ_1, ϕ_E is greater than ϕ_F

D4(a)(i) 6.4 x 10^{14} Hz, **(ii)** 7.69 x 10^{-19} J, **(iii)** 6.67 x 10^{-34} J s, **(c)(i)** h/e ; $-\phi/e$, **(ii)** hc; $-\phi$, **(iii)** hc/e ; $-\phi/e$

D5(a) Maximum KE = 5.92 x 10^{-21} J, max. speed = **1.14 x 10^{5} m s^{-1}**

D6(a) 2.49 eV, **(b)** 1.2 x 10^{14} Hz, **(c)** 1.01 x 10^{9} s^{-1}

D7 8.43 x 10^{-12} C

17.2 Atomic Structure and Line Spectra

17.2.1 Bohr's Model of Atomic Quantization

Rutherford's Planetary Model of Atom and its Flaws

Throughout mankind's history, the quest to understand the composition of matter and the structure of the atom has evolved through various theories[3]. By the 20th century, Ernest Rutherford's nuclear model was the most successful of all the various proposals.

In Rutherford's model of atom, the atom was made up of a massive, positive central nucleus, with electrons orbiting around this nucleus. However, although this model provided crucial insights into the atomic structure of the atom, it also faced significant flaws.

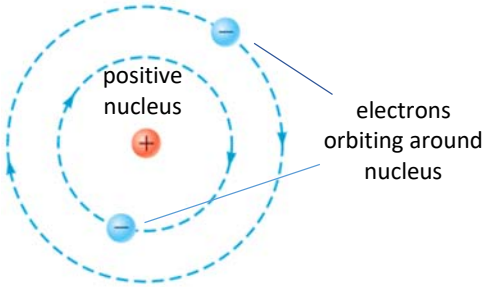


Fig. 9.1. Rutherford's planetary model of the atom

Fig. 9.2. Classical theory predicts that the orbiting electrons should spiral into the nucleus and radiate electromagnetic with a continuous spectrum.

According to Maxwell's classical theory of electromagnetism, any accelerating electrical charge radiates electromagnetic waves. An electron orbiting in an atom would always have a centripetal acceleration and so should always be emitting electromagnetic radiation. With the electron continuously losing energy, the radius of the orbit should decrease continuously, and the electron should spiral into the nucleus with a fraction of a second. (Fig. 9.2)

Furthermore, in the 1800s, Michael Faraday had conducted a series of experiments and observed that atoms emit (and absorbs) only certain frequencies of electromagnetic radiation that were characteristics of the element – producing discrete line spectra. (Fig. 10)

According to classical theory, the *frequency* of the electromagnetic waves emitted by the orbing electrons is equal to the frequency of revolution. As the electron spiral towards the nucleus, its frequency of revolution continuously increases and hence should emit a *continuous* spectrum (a continuous variation of all frequencies). Classical theory, hence, could not explain the line spectrum emitted or absorbed by atoms.

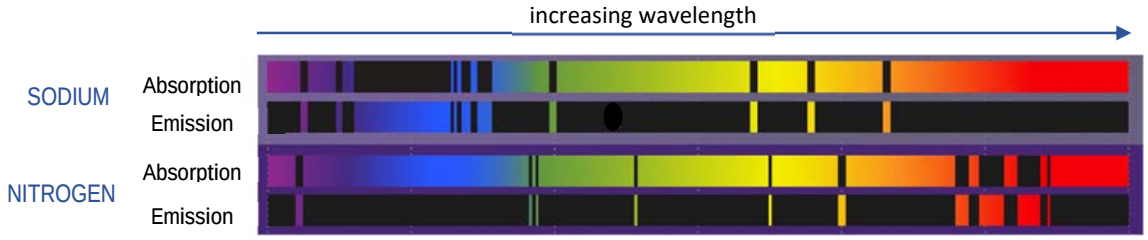


Fig. 10. The absorption and emission line spectral of sodium and nitrogen.

[3] Around 500 BC, Democritus proposed the concept of atoms, suggesting that matter consists of indivisible particles. Later, Joseph John Thomson introduced the plum-pudding model, envisioning electrons embedded within a positively charged mass. This model was challenged by the gold-foil experiment conducted by Geiger and Marsden, leading Ernest Rutherford to propose a new "planetary model" featuring a dense, positively charged nucleus surrounded by orbiting electrons, akin to a miniature solar system. We will learn more about JJ Thomson's plumb pudding model and Rutherford's planetary model of atoms in our study of Nuclear Physics.

Bohr's model of the atom

In 1913, Niels Bohr presented a new model of the atom to address this problem. Bohr applied Planck's idea of quantized energy levels in blackbody radiation, Einstein's concept of the photon, Rutherford's planetary model of the atom, and Newtonian mechanics to arrive at a semiclassical model[4] based on some revolutionary ideas. **The postulates of the Bohr theory as it applies to the hydrogen atom are as follows**:

1. The electron moves in circular orbits around the nucleus under the influence of the electric force of attraction. However, **only certain electron orbits are stable. These states are called the stationary states**. When in one of these stationary states**,** the electron does not emit energy in the form of radiation, even though it is accelerating. Hence, the total energy of the atom remains constant. Furthermore, because the electron does not continuously emit radiation, or lose energy, it will not eventually spiral into the nucleus.
2. **The atom emits radiation, in a single photon, when the electron makes a transition from a more energetic initial stationary state**, at a higher orbit, **to a lower-energy stationary state** at a lower orbit. The frequency *f* of the photon emitted in the transition is related to the change in the atom's energy and is not equal to the frequency of the electron's orbital motion.

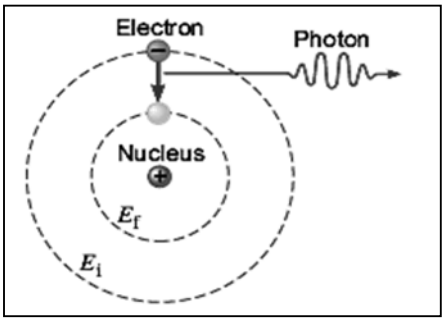


Fig. 11

De-excitation and Excitation

When an electron in an initial orbit with a larger energy E_i changes to a final orbit with a smaller energy E_f, the emitted photon has an energy of E_i - E_f, consistent with the law of conservation of energy,

$$hf = E_i - E_f$$

where *f* is the frequency of the emitted light. This process is known as **de-excitation**.

Electrons in the atom can also jump from a lower-energy orbit to a higher-energy orbit by acquiring energy. This process is known as **excitation**.

Excitation can occur

- through **collision of the atom** with another particle, where the colliding particle will impart energy to the atom equal to the difference between two energy levels in the atom, while any remaining energy is retained by the colliding particle as its kinetic energy;
- through **absorbing a photon** of energy **exactly** equal to the energy difference between two levels of the atom. Upon absorption, the photon disappears and the atom makes a transition to the higher energy state.

The Bohr model is sometimes known as the semi-classical model of the atom, as it adds some primitive quantization conditions to what is otherwise a classical-mechanics treatment. The Bohr model is certainly not a full quantum-mechanical description of the atom. The full quantum-mechanical model, or electron-cloud model, describes the electron as a three-dimensional shape ("cloud") where there is at least a minimal probability of finding the electron.

[4] Although Bohr's model is now considered obsolete and has been completely replaced by a probabilistic quantum-mechanical theory, it is historically important as the Bohr model developed the notions of energy quantization and angular momentum quantization as applied to atomic-sized systems.

Application of Bohr's Model to the Hydrogen Atom

The simplest Bohr model would be that of the hydrogen atom, which consists of just one electron and one proton. Bohr was able to derive the energy levels E_n for the hydrogen atom[5]:

$$E_n = -\frac{13.6}{n^2} \text{ eV}$$

where n is known as the principal quantum number and is a positive integer (i.e., 1, 2, 3, …).

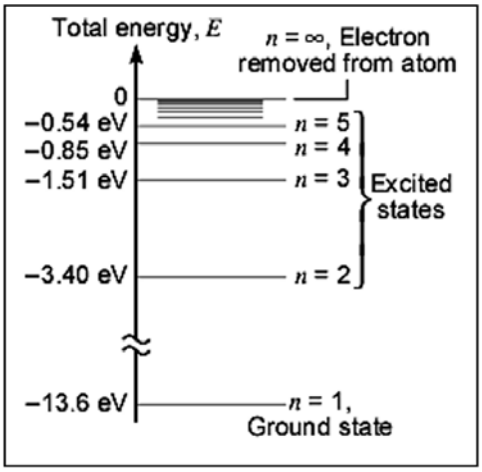


Fig. 12

The electron normally occupies the energy level corresponding to n = 1, also known as the **ground state**. At the ground state, the electron has the lowest energy, -13.6 eV, and orbits nearest to the proton (hydrogen nucleus).

The higher energy levels are called **excited states**. Fig. 12 shows how the energies of the excited states come closer and closer together as n increases.

The electron in a hydrogen atom at room temperature spends most of its time in the ground state. To raise the electron from the ground state (n = 1) to the highest possible excited state (n = ∞), 13.6 eV of energy must be supplied. Supplying this amount of energy removes the electron from the atom, producing the positive hydrogen ion H^+. 13.6 eV is known as the **ionisation energy** of the hydrogen atom.

The following video is about the energy levels in the hydrogen atom:
Electron Energy Levels – A Level Physics https://www.youtube.com/watch?v=DzK0MjMfb-w

In general, for a multi-electron atom,

> The **ionisation energy** is the minimum energy needed to remove the outermost electron from the atom.

When the electron is given an amount of energy equal to the ionisation energy, the electron is just able to become a "free" electron. If the electron is given an amount of energy greater than the ionisation energy, the excess energy becomes the kinetic energy of the free electron outside the atom. **The free electron outside an atom can have a continuous range of energies. Inside the atom, however, an electron can only have an energy that is characteristic to that atom.**

Evidence that the energy states of the atoms are discrete and quantised is provided by the existence of atomic line spectra, the subject of the next two sections.

[5] You are not required to memorise this formula or know how Bohr arrived at it. In fact, you are not required to know the energy levels or spectra for any particular element.

17.2.2 Emission Line Spectra

When gases such as hydrogen, neon or sodium are placed inside a narrow discharge tube at low pressure (so that the gas atoms are effectively isolated from one another), a high voltage (e.g., 1 kV) applied across its ends causes the gas atoms to ionise, forming ions and electrons. The ions and electrons accelerate towards opposite ends of the tube, under the applied electric field. **Collisions between the electrons and atoms result in transfer of energy from the electrons to the atoms and the atoms are excited from their ground state.** Excited states are not stable, and **an excited atom will quickly de-excite to a lower-energy state emitting a photon of energy corresponding to the difference between the initial higher energy level and final lower energy level.** The excited atom can, but may not de-excite directly to ground state immediately. It may make a series of jumps through intermediate energy states to ground state, releasing more than one photon in the process.

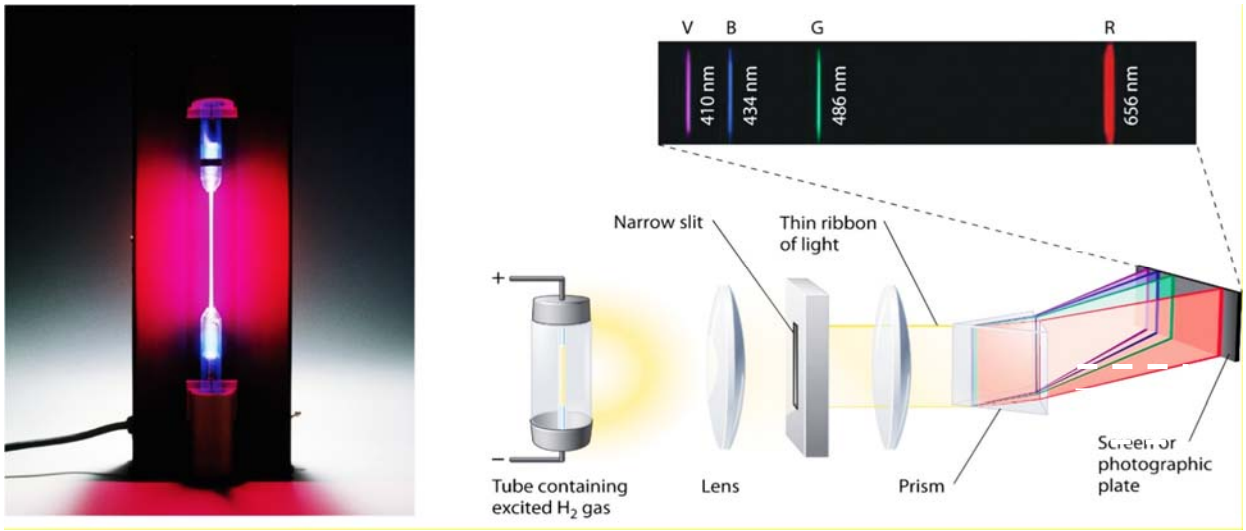


Fig. 13

The photons emitted by the gas discharge tube are passed through a collimator slit, to produce a narrow beam of photons. Next, the photons are passed through a diffraction grating to produce an **emission line spectrum** on a dark screen behind the grating (Fig. 13).

The frequency f of an emitted photon depends on the energy difference between the initial and final energy states, as given by $hf = E_i - E_f$. The frequency or wavelength of the photons emitted can be determined by studying the spectrum seen on the screen.

The emission line spectrum consists of a series of **distinct coloured lines on a dark background**. The different colours or frequencies are unique to the gas, i.e., no two different gases emit the same set of discrete frequencies of light, and atoms of the same element always emit the same set of discrete frequencies of light.

Using his model for the energy levels of the hydrogen atom, Bohr was able to account for the spectral pattern of hydrogen gas. For hydrogen gas, several distinct groups of lines can be detected, though only one group is in the visible range (Fig. 14(a)).

The Lyman (ultraviolet) series of lines corresponds to electron transitions down to the level $n = 1$, the Balmer (visible) series is produced by transitions down to $n = 2$, and the Paschen (infrared) series are produced by transitions down to $n = 3$; each series is named after its discoverer. The Balmer series is shown in Fig. 14(b).

You may find a YouTube video about the Atomic Line Spectrum of Hydrogen here:
Atomic Line Spectrum H https://www.youtube.com/watch?v=5aCWCn9Dhqs

The following video talks about different spectra:
Emission and Absorption Line Spectra https://www.youtube.com/watch?v=xhaDuAkpF8A

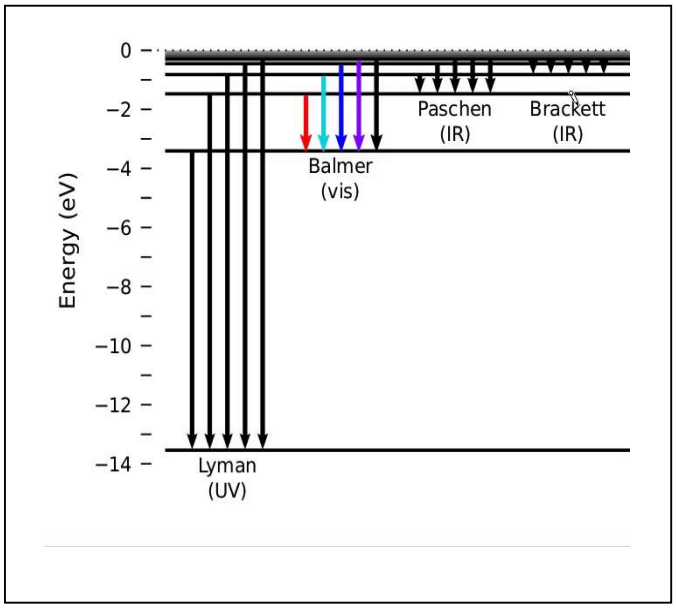


Fig. 14(a)

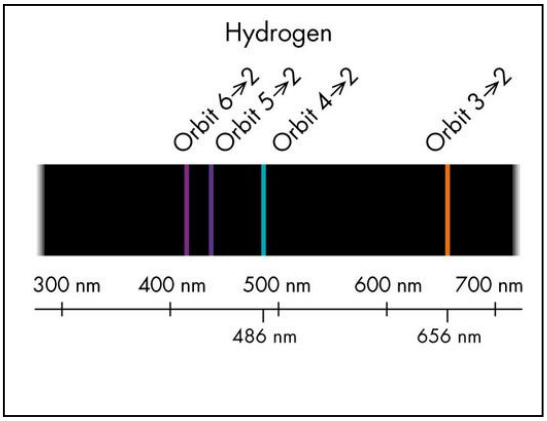


Fig. 14(b)

Typically, gas at low pressure is used for spectral analysis, as the atoms are sufficiently far apart and have minimal interaction with one another. When the atoms are close together, the discrete energy levels will become several continuous energy *bands* due to interactions between the atoms, and the distinctive spectral lines will no longer be observable.

You may read on “band theory” on your own if you are interested. It forms the foundation of the understanding of all solid-state devices (transistors, solar cells, etc.).

17.2.3 Absorption Line Spectra

An absorption spectrum is formed when the continuous spectrum emitted from an incandescent lamp (or white light) passes through a cool gas. The electrons in **the gas absorb the photons with frequencies corresponding to the energies they need to jump from a lower energy state to a higher energy state.** These reduction of the photons of these frequencies show up as "black lines" in the otherwise continuous spectrum of colours. An absorption spectrum, in a sense, is the inverse of an emission spectrum of the same series.

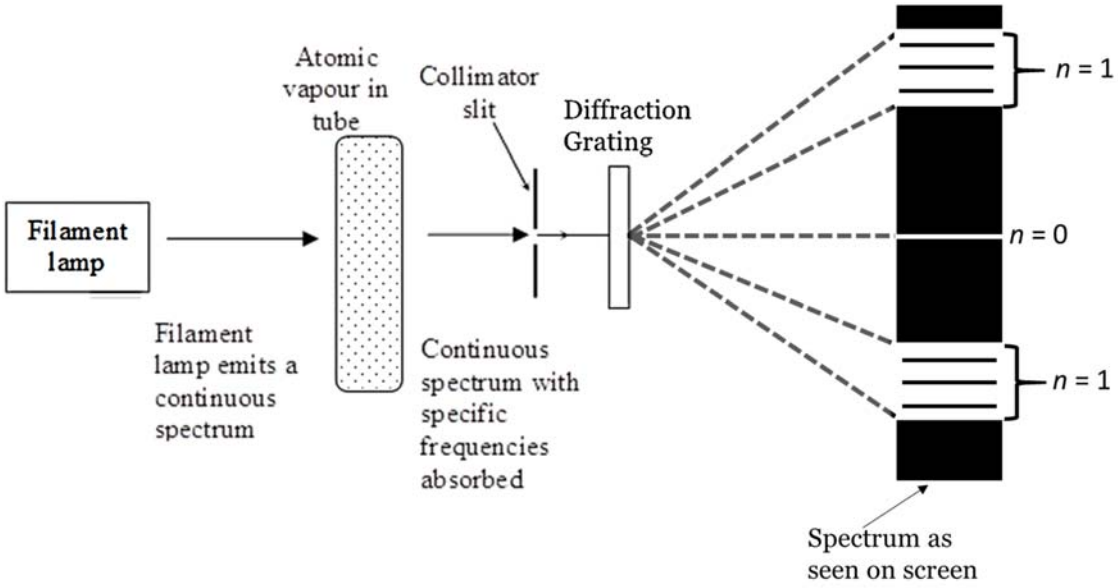


Fig. 15(a)

In the set-up shown in Fig. 15(a), a filament lamp emits a continuous spectrum of light, which is incident on a tube containing *cool* atomic vapour at low pressure ("cool" meaning atomic vapour in the ground state). As a photon is incident on an atom, if its energy corresponds to the energy difference between the ground state and an excited state, it may be absorbed by the atom. As such, the atomic vapour absorbs photons of specific energies/frequencies.

What is observed on the screen is the original **continuous spectrum emitted by the hot filament lamp with darker lines corresponding to the absorbed frequencies**. Since excited states are not stable, the atoms eventually return to the ground state. However, **the photons are emitted in all directions** and not just in the forward direction. Hence, the parts of the spectrum corresponding to these photon frequencies appear darker in comparison to the other frequencies (see Fig. 15(b)).

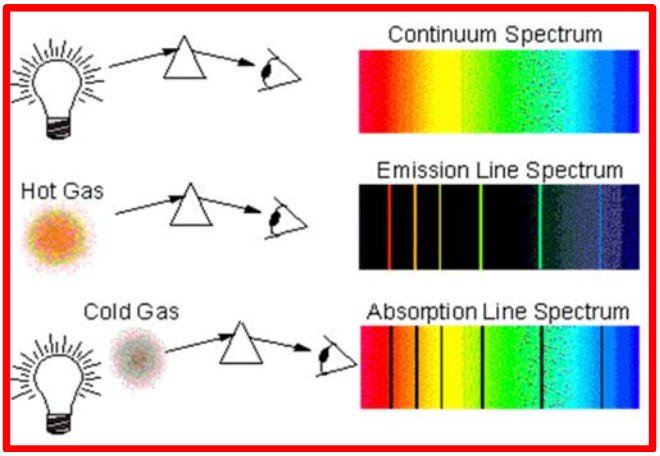
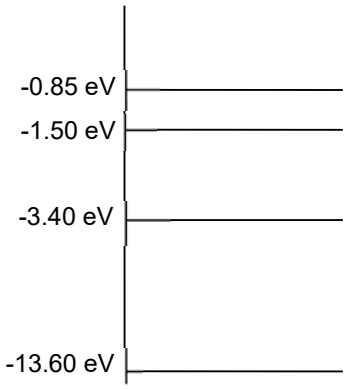


Fig. 15(b)

Example 8

The diagram represents the energy levels of the four lowest states of a hydrogen atom. The energies are in units of electron volts.



(a) Calculate the longest wavelength which might be emitted by a spectral transition between any pair of these four levels.

(b) Calculate the excitation energy for a jump from the ground state to the next higher level. Express your answer in J.

(c) Determine the total number of spectral lines which might be detected in the emission spectrum of atomic hydrogen, due to transitions between these four states.

(d) Suggest why the energies of the different levels are negative.

17.3 X-ray Spectra

17.3.1 X-ray Production

X-rays were discovered by Wilhelm K. Röntgen in 1895. He found that a gas-discharge tube which he was using with a high voltage had caused some wrapped photographic plates to become fogged. He discovered that the tube emitted a radiation that could penetrate paper, wood, glass, rubber and even thick aluminium. Röntgen could not find out whether the radiation was a stream of particles or a train of waves and he named it **X-rays**.

X-rays are produced when high-speed electrons, accelerated through a large potential difference, collide with a metal target and are slowed down. The target is contained within an evacuated glass tube known as an X-ray tube (see Fig. 16).

The production of X-rays can be seen as the **inverse process** of the photoelectric effect. In X-ray production, an incoming electron causes a photon to be emitted. In the photoelectric effect, an incoming photon causes an electron to be emitted. However, the energies involved in the two processes are considerably different.

Photoelectric emission	***X-ray production***
Transformation of the *energy of a photon* into the *kinetic energy of an electron.*	Transformation of the *kinetic energy of an electron* into the *energy of a photon*.
Typical energy of photon: a few electron volts (eV).	Typical energy of electron: in the kilo electron volts (keV) range.

A typical plot of X-ray intensity versus wavelength is shown in Fig. 17. It consists of sharp peaks or lines superimposed on a broad continuous spectrum. The sharp peaks are called **characteristic lines** or **characteristic X-rays** because they are **characteristic of the target metal used**. The broad continuous spectrum is referred to as **bremsstrahlung** (German for "braking radiation") and is emitted when the bombarding electrons decelerate upon hitting the target.

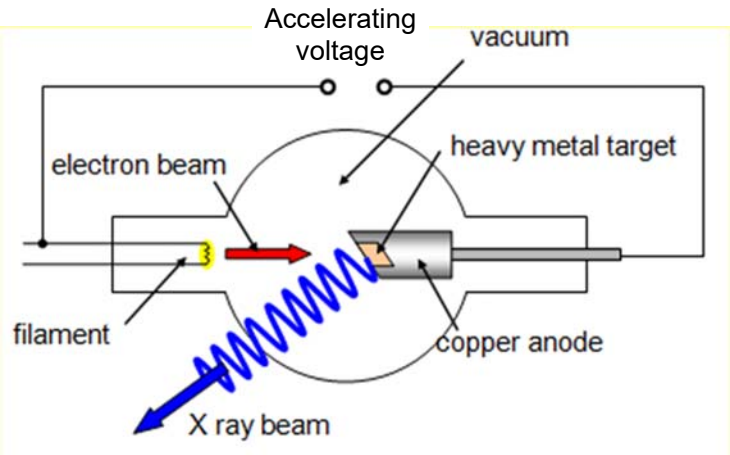


Fig. 16 X-ray tube

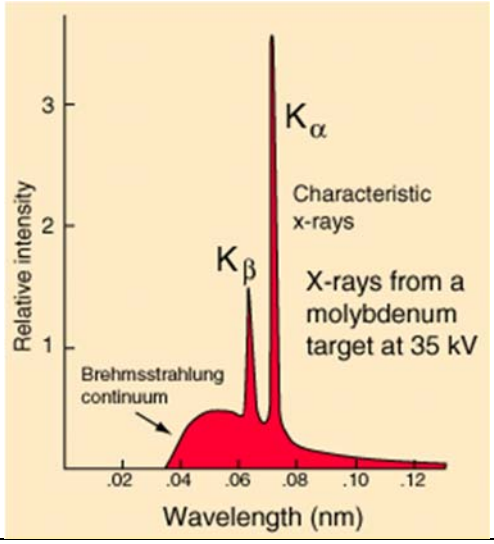


Fig. 17 X-ray spectrum

17.3.2 Explanation of X-ray Emission Spectra

Two distinct processes are involved in X-ray emission (corresponding to the two distinct features of the X-ray spectrum in Fig. 17). The first process accounts for the bremsstrahlung, while the second accounts for the characteristic lines.

Bremsstrahlung

In the first process, some of the electrons pass through the electric field of the atomic nuclei in the metal target. As a result of the interaction with the atomic nuclei, part or all of the kinetic energy of these electrons is converted directly to a **continuous spectrum** of photons, including X-rays. The electrons are slowed down or stopped by the target as a result. Because the electrons are slowed down in the process, this radiation is known as *bremsstrahlung* (German word for 'braking radiation'.)

The intensity of the continuous background radiation decreases gradually for increasing wavelength, but is cut off sharply at the other end of the spectrum with a minimum wavelength λ_{min}.

Cutoff Wavelength, λ_{min}

Using the quantum theory of light, one can explain the existence of λ_{min} (which is often known as the *cutoff wavelength).*

When the bombarding electrons interact with the metal atoms in the target, they can lose any fraction of their initial energy, up to the maximum in a single interaction. If the electron loses all its energy in a single interaction, based on quantum theory, a single quantum of electromagnetic radiation (photon) is produced with maximum energy, and hence **minimum wavelength** (and **highest frequency**).

If the accelerating voltage of the X-ray tube is V, and an electron has charge $-e$, the electron gains kinetic energy eV. Since, the most energetic photons are produced when this entire electron's kinetic energy is converted into a single photon, then, by conservation of energy,

$$\text{Energy of photon emitted} = \text{Kinetic energy of the electron} \Rightarrow h\frac{c}{\lambda_{min}} = eV$$

Hence, the **cutoff wavelength λ_{min}:**

$$\boxed{\lambda_{min} = \frac{hc}{eV}}$$

(for an electron completely stopped in its first interaction with the metal target)

From, the derivation, we can also see that the minimum wavelength λ_{min} in the bremsstrahlung process is ***independent*** of the target material. It only depends on the accelerating voltage.

The Characteristic Lines

The second process accounts for the peaks in the X-ray spectrum at certain frequencies (i.e., the characteristic X-rays in Fig. 17).

Electrons in atoms are grouped in various energy states called "shells." The shell nearest to the nucleus is known as the **K** shell. This is followed by the **L** shell, the **M** shell, the **N** shell and so on. (Fig 18 (a),(b))

In the X-ray tube, highly energetic electrons which have been accelerated by tens to hundreds of kilovolts of potential difference bombard the metal target. These highly energetic bombarding electrons can eject electrons from the *inner shells* (i.e., closer to the nucleus) of the atoms of the metal target. The vacancies created will be quickly filled by electrons dropping down from higher levels, emitting X-rays with sharply defined frequencies associated with the difference between the atomic energy levels of the target atoms.

If E is the difference in energy between the two shells where the electron transition occurs, the frequency f of the radiation emitted is related to E by

$$E = hf.$$

As E is very high for electronic transitions between the K, L and M shells, the frequency f of the emitted radiation is also very high and the wavelength correspondingly short, typically of the order of 10^{-9} m or less (i.e., in the X-ray region of the electromagnetic spectrum).

The K series of X-ray lines, K_α, K_β, K_γ, etc. is due to the transition of electrons from the L, M, N... shells to the K shell (Fig. 18(a)). Similarly, L lines can be produced from transitions to the L shell (Fig. 18(b)).

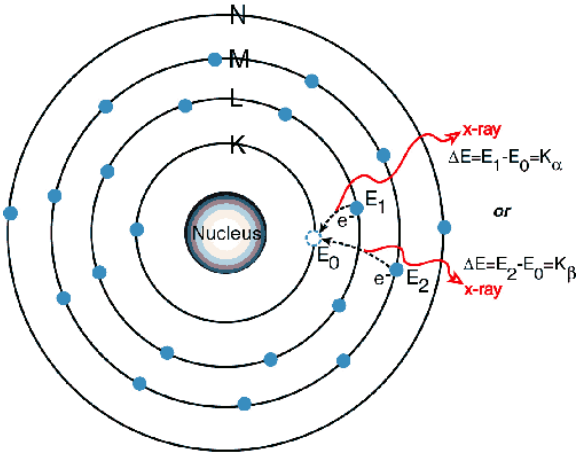


Fig. 18(a)

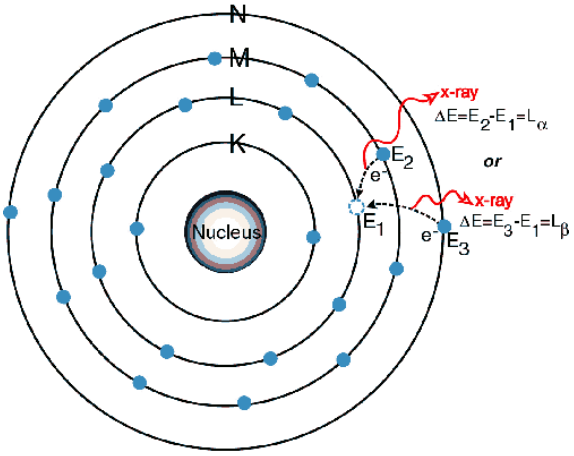


Fig. 18(b)

Part of the shell structure of Tungsten (74 protons) is shown in Fig. 18(c). If a vacancy in the K shell is filled by an L-shell electron, a photon is emitted that contributes to the characteristic K_α line; if it is filled by an M-shell electron, a K_β photon is emitted, etc.

Thus, **energy changes (de-excitation) in electrons close to the nucleus of metals** result in the characteristic X-rays. In contrast, the emission and absorption spectra we studied in section 17.2 are due to energy changes within the outermost shell of the atom, and the energy changes producing optical spectra are about 1000 times smaller than those producing X-rays. Hence, the frequencies of the optical lines are about 1000 times smaller, that is, their wavelengths are about 1000 times longer than those of X-rays.

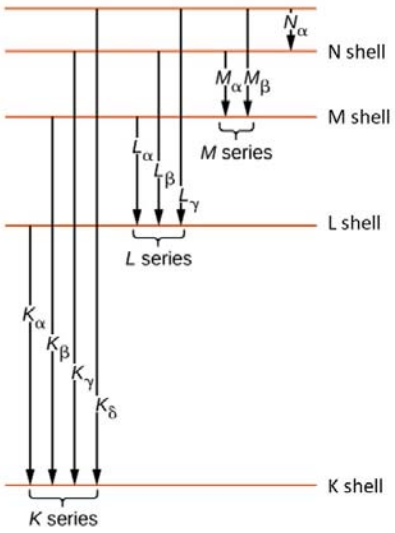


Fig. 18(c)

Unlike the minimum wavelength of an X-ray spectrum, **the wavelengths of the characteristic lines are dependent on the target materia**l. In fact, that is why they are called the characteristic lines: they are characteristic of the atomic structure of the target material.

Fig. 19 illustrates the bremsstrahlung and characteristic line radiation production.

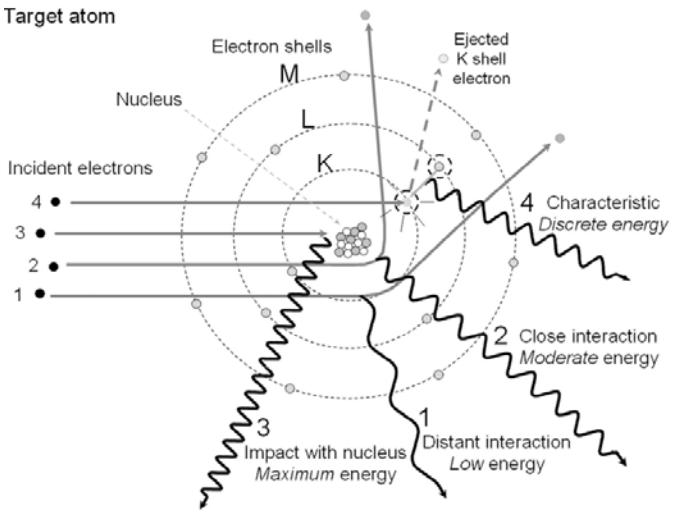


Fig. 19(a)
Bremsstrahlung (1,2,3) and K_α characteristic radiation (4) mechanisms

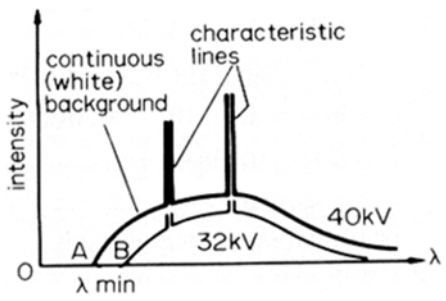


Fig. 19(b).
X-ray characteristic lines and background radiation (bremsstrahlung)

Fig. 19(b) shows the characteristic lines K_α and K_β and the continuous background radiation (bremsstrahlung) for two values of the accelerating potential, 40 kV and 32 kV, for a certain X-ray tube, i.e., the target material is the same for both spectra.

Two observations can be made from Fig. 19(b):

1) The wavelengths of the characteristic lines are independent of the applied potential difference – they are characteristic of the metal.
2) λ_{min} for the 40 kV spectrum is smaller than that for the 32 kV spectrum – this is to be expected as photons produced by electrons accelerated through a potential difference of 40 kV would have a greater maximum energy than those produced by electrons accelerated through a potential difference of 32 kV.

How does an X-ray Tube Work (Radiation Protection)

https://www.youtube.com/watch?v=Bc0eOjWkxpU

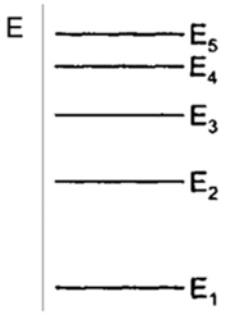
How does an X-ray Tube Work

Tutorial 17B: Line Spectra & X-ray Spectra

Line Spectra Self-Review Questions

S1 An atom emits a spectral line of wavelength λ when an electron makes a transition between levels of energy E_1 and E_2. Give the expression which relates λ, E_1 and E_2.

S2 The diagram shows the first five energy levels of an atom.



Photons of electromagnetic radiation are emitted when an electron falls from one energy level to a lower level.

Which transition corresponds to a photon with the greatest wavelength?

A E_2 to E_1

B E_5 to E_1

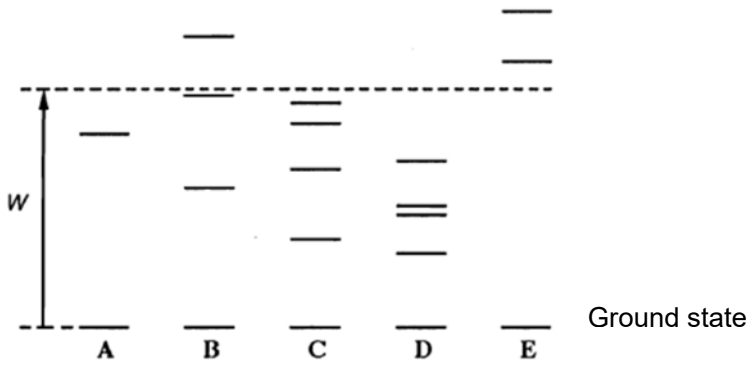
C E_4 to E_3

D E_5 to E_4

UCLES N2005/I/29

S3 The diagram shows the electron energy levels, referred to the ground state (the lowest possible energy) as zero, for five different isolated atoms.

Which atom can produce radiation of the shortest wavelength when atoms in the ground state are bombarded with electrons of energy W?



Question : If the atom was bombarded with photons with energy W, will your answer be different? Explain.

CIE J89/I/29 (modified)

S4 When a parallel beam of white light passes through a metal vapour, dark lines appear in the spectrum of the emergent light. This is principally because energy is absorbed and

A is not re-radiated at all.

B is re-radiated as infra-red.

C is re-radiated as ultra-violet.

D is re-radiated gradually over a long period of time.

E is re-radiated uniformly in all directions.

CIE J82/I/34

X-ray Spectra Self-Review Questions

S5 Calculate the shortest wavelength X-ray that can be produced when a target is struck by an electron that has been accelerated through a potential difference of

(a) 15.0 kV,

(b) 100 kV.

S6 **(a)** Explain what X-rays are and how they are formed.

(b) State one similarity and one difference between visible-light emission spectra and X-ray emission spectra.

(c) Discuss the differences between bremsstrahlung X-rays and characteristic X-rays.

(d) Electrons are accelerated from rest through a potential difference of 10 kV in an X-ray tube. Calculate the minimum wavelength of X-rays generated.

Line Spectra Discussion Questions

D1 **(a)** Explain what is meant by a photon. [2]

(b) **(i)** Describe the appearance of a visible line emission spectrum. [2]

(ii) State how line spectra, together with the concept of a photon, provide evidence for discrete energy levels in isolated atoms. [2]

(c) Some electron energy levels of the hydrogen atom are illustrated in Fig. 8.1.

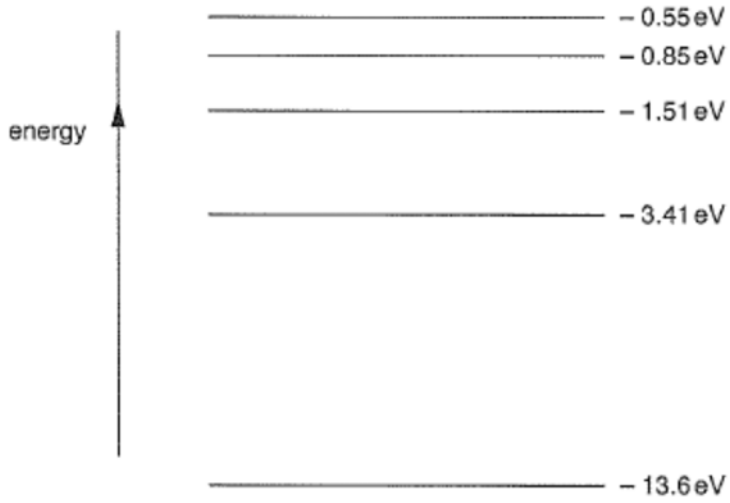


Fig. 8.1 (not to scale)

(i) The wavelength of violet light is approximately 400 nm.

By calculating the energy, in eV, of a photon of violet light, explain why electron transitions to the energy level – 13.6 eV do not result in emissions in the visible spectrum. [4]

(ii) The minimum energy of a photon of visible light is 1.6 eV.

Use Fig. 8.1 to determine, for the electron energy levels shown,

1. the number of lines that lie in the visible spectrum, [1]
2. the shortest wavelength of photons in this visible spectrum [2]

(d) Considering transition between only these levels, how many spectral emission lines might be produced by transitions among these levels?

(e) The radiation emitted from transitions between levels shown in Fig. 8.1 is incident on the surface of a sheet of platinum. Platinum has a work function energy of 5.6 eV.

(i) On Fig. 8.1, mark with letter P the transition that will give rise to radiation with the longest wavelength that will cause emission of electrons from the platinum. [1]

(ii) Calculate the maximum energy (in eV) of an electron emitted from the platinum by the radiation in (i). [2]

(f) When the spectrum of gamma radiation from a nucleus is examined, it is found that it too has a line emission spectrum. Suggest what can be deduced from this observation. [2]

UCLES N2016/9646/P3/Q8 part; N84/III/6; J93/III/6 (part)

D2 Light in a beam has a continuous spectrum that lies within the visible region. The photons of light have energies ranging from 1.60 eV to 2.60 eV.

The beam passes through some hydrogen gas. It then passes through a diffraction grating and an absorption spectrum is observed. The dark lines in the absorption spectrum are the result of electron transitions between energy levels.

Some of the energy levels of an electron in a hydrogen atom are illustrated in Fig. 10.1.

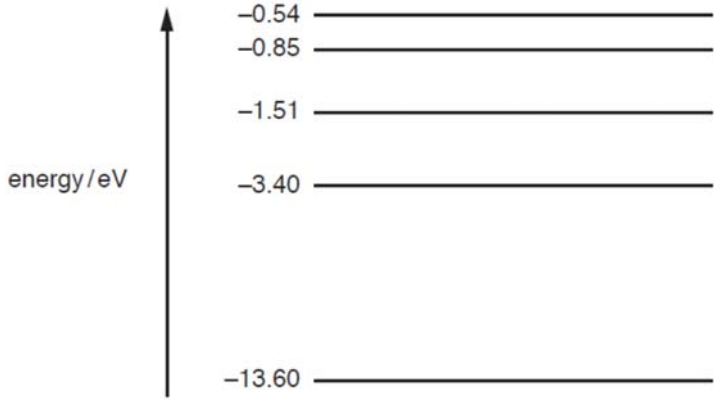


Fig. 10.1 (not to scale)

(a) **(i)** Explain how the figure above can be used to account for absorption spectra. [2]

(ii) All of the light absorbed by the hydrogen is re-emitted. Explain why dark lines are still observed in the absorption spectrum. [1]

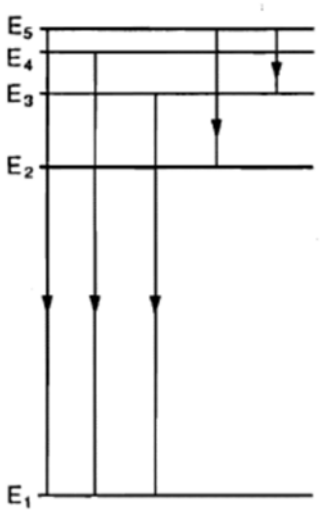
(b) On Fig. 10.1, draw arrows to show the initial electron transitions between energy levels that could give rise to dark lines in the absorption spectrum. [2]

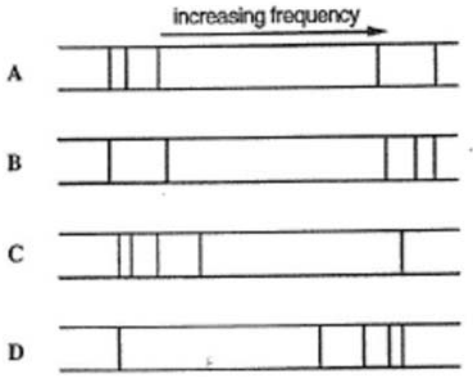
(c) Calculate the shortest wavelength of the light in the beam that is absorbed. [2]

CIE9702 March 2017 P42 Q10

D3 The diagram on the right shows five energy levels of an atom. Five possible transitions between the levels are indicated. Each transition produces a photon of definite energy and frequency.

Which spectrum corresponds most closely to the transitions shown?





CIE N97/I/28; N85/I/29; J83/II/35

D4 In an experiment, sodium vapour with its atoms in the ground state is bombarded by a beam of electrons, each of energy 3.65 eV.

Fig. 3.1 depicts an electron that is scattered by a sodium atom which is originally in the ground state (n = 1). The energy levels of atomic sodium are shown in Fig. 3.2.

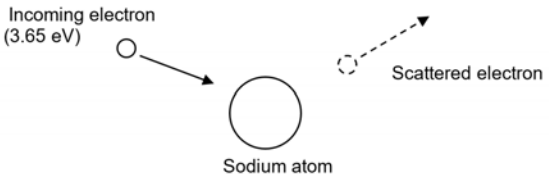


Fig. 3.1

Fig. 3.2

(a) What is the ionization energy of a sodium atom? Give your answer in both eV and joules.

(b) If a photon with an energy of 0.42 eV is produced by one particular collision of a sodium atom with an electron, what is the energy of the scattered electron? Explain your answer clearly.

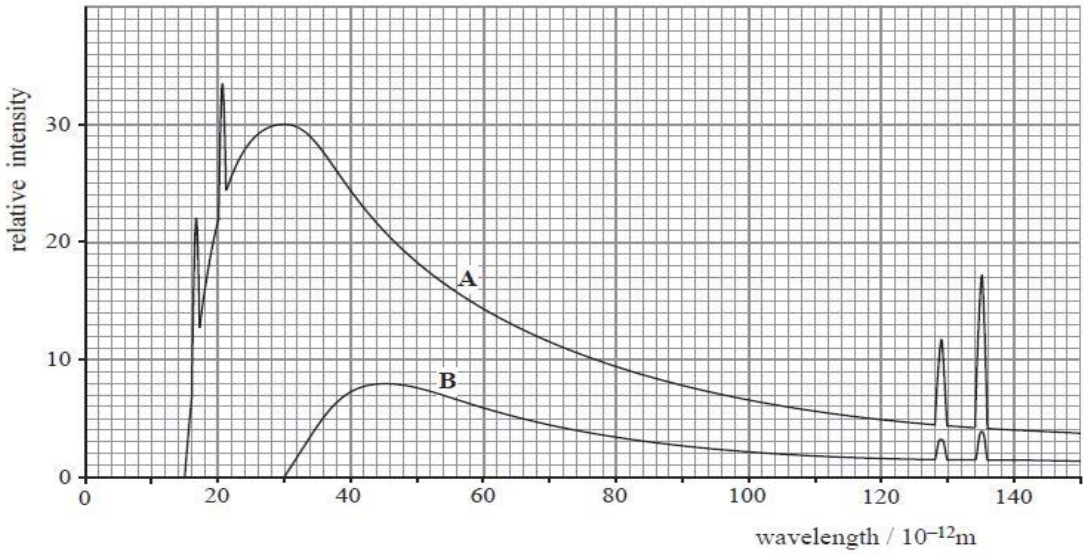
(c) If a sodium atom is excited by a collision with an electron, will it be possible for a scattered electron to have energy of 2.41 eV? Explain your answer.

(d) What would happen if photons of energy 3.15 eV each were incident on sodium gas?

2003 HCI Prelim, P3, Qn 6 (part)

X-ray Spectra Discussion Questions

D5 The diagram below shows two X-ray spectra produced by X-ray tubes.



(a) Label Graph A with the following:

(i) bremsstrahlung,

(ii) characteristic lines,

(iii) cutoff wavelength, and

(iv) K_α, K_β, L_α and L_β. [2]

(b) The material of the target is the same in both cases. Explain how the graphs support this statement. [2]

(c) Explain why there is a minimum wavelength for each of the emitted X-rays spectrum. [3]

(d) Spectrum A was produced using an accelerating voltage of 80 kV.

Showing your reasoning, calculate a value for the Planck constant. [2]

(e)* Suggest and explain how spectrum B is produced.

(f) Calculate the accelerating voltage used to produce spectrum B. [2]

(g)* With the aid of Fig. 5.2,

(i) map a partial energy level diagram for the atom of the target material. Label the energy levels in your sketch in eV. Include the ionization level in your diagram. Also clearly indicate also in your diagram the energy levels corresponding to K, L, M and N-shell of the target material. You may assume that energy of K-shell is 0 eV .

(ii) Draw labelled arrows representing the transitions for K_α, K_β, L_α and L_β in your sketch

WJEC 2006 Jun. PH542/01. Q5 (c). Extended*.

D6 The *K* shell ionisation energy of copper is 8 979 eV. The *L* shell ionisation energy is 951 eV. Determine the wavelength of the K_α emission line of copper.

What must be the minimum voltage on an X-ray tube with a copper target in order to see the K_α line?

Answers to Tutorial 17B: Line Spectra & X-ray Spectra

Self-Review Questions

S2 D **S3** B **S4** E **S5 (a)** 8.29 x 10^{-11} m, **(b)** 1.24 x 10^{-11} m

S6 (d) 1.24 x 10^{-10} m

Discussion Questions

D1(c)(ii)1. 3, **2.** 4.35 x 10^{-7} m, **(d)** 10, **(e)(ii)** 4.59 eV **D2(c)** 4.88 x 10^{-7} m

D3 B **D4(a)** 5.13 eV, 8.21 x 10^{-19} J, **(b)** 0.04 eV, **(c)** No.

D5(d) 6.4 x 10^{-34} J s, **(f)** 41.4 kV, **D6** 1.55 x 10^{-10} m, 8979 V

17.4 Wave-Particle Duality and the Uncertainty Principle

What is the Wave/Particle Duality? Part 1 https://www.youtube.com/watch?v=Q_h4IoPJXZw
The Wave/Particle Duality – Part 2 https://www.youtube.com/watch?v=_riIY-v2Ym8

Minute Physics version

17.4.1 Introduction

We have seen evidence that light can behave like a particle. But can Einstein's particulate theory of light explain the interference and diffraction of light? Unfortunately, it cannot. Such behaviour can only be explained using the wave theory of light. Hence, light also exhibits wave-like properties like diffraction, interference and polarization. (Experiments like Young's double-slit experiment clearly demonstrate this wave behavior.)

Therefore:

Light behaves **both** as particles and as a wave.

In fact, the equation *E = hf* contains this contradictory nature of light. The energy of a photon – a *particle* of light – depends on the *frequency* of the light. Light can have a frequency only if it is a wave.

17.4.2 De Broglie's Hypothesis – Matter Waves

As a graduate student in 1923, Louis de Broglie made an astounding suggestion. Since light can exhibit particle-like behaviour, then the converse should be true. Particles (like electrons) should also exhibit wave-like behaviour.

De-Broglie Relationship:

De Broglie boldly proposed that a relationship between momentum (of the particle) and wavelength (of the **associated matter wave**) exists for **all particles**:

$\lambda = \frac{h}{p}$ where *p* is the momentum of the particle

λ is known as the ***de Broglie wavelength*** of the particle.

Example 9 (J77/II/33)
Calculate the de Broglie wavelength of a rifle bullet of mass 0.02 kg moving at a speed of 300 m s^{-1}.

17.4.3 Evidence of the Wave-like Nature of Particles

To exhibit diffraction with light, we had to make very narrow slits that are **approximately equal to the wavelength of light (10^{-7} m for visible light).** For his experiment, Young used a razor blade to scratch slits on a piece of glass that had been blackened with soot.]

In order to validate de Broglie's hypothesis, we must be able to observe wave-like behaviour of particles (e.g., electrons).

The **de Broglie wavelength of electrons** is determined to be even shorter than the wavelength of visible light. In fact, it **is of the same order of magnitude as the size of a typical atom (10^{-10} m).** This holds the key to showing diffraction of electrons. Diffraction experiments were successfully conducted using a metal crystal. The **regularly arranged atoms of the crystal acted as a diffraction grating for the electron "waves".** A number of important experiments conducted in the 1930s was based on this principle and clearly demonstrated the wave-like nature of moving electrons.

Clinton Davisson and Lester Germer directed beams of electrons onto nickel. The surface of a crystal of nickel acted like a diffraction grating and it was observed that the electrons exhibited diffraction behaviour.

Further experiments to demonstrate the wave nature of electrons were conducted by G.P. Thomson. He targeted mono-energetic beams of electrons at a thin metal foil (Fig. 20). A photographic plate placed behind the foil revealed a bright central maximum surrounded by diffraction rings (Fig. 21).

G.P. Thomson is the son of J.J. Thomson, who discovered the electron. The father was awarded a Nobel Prize for demonstrating that the electron was a particle. The son was awarded one for showing that electrons can behave like waves!

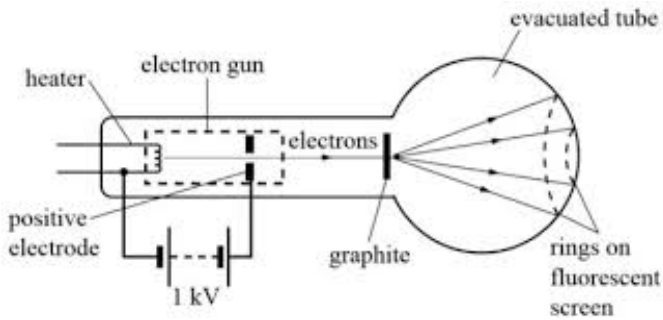


Fig. 20 Schematics of a basic electron diffraction experiment

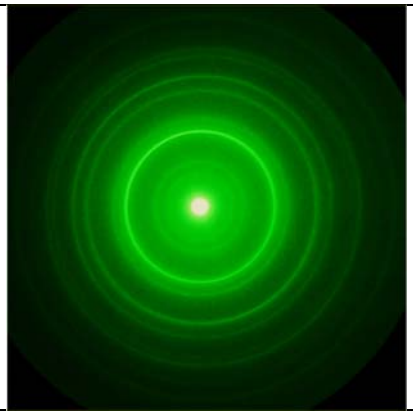


Fig. 21 Image on the fluorescent screen (the portions hit by electrons will light up)

The following video tells us how to use an electron diffraction tube
Using an electron diffraction tube https://www.youtube.com/watch?v=MiIAnxIZKQo

These **electron diffraction experiments clearly show that electrons can behave like waves**, and detailed calculations show that de Broglie's equation is valid. The wave nature of electrons is something that we now take for granted, and has been put to practical use in devices like the electron microscope.

We conclude **that matter or radiation cannot be completely described by the wave or particle model alone. The wave and particle descriptions are complementary**, even though in explaining any particular phenomenon we just need to invoke one model.

Example 10
What is the de Broglie wavelength of an electron that has been accelerated from rest through a potential difference of 54 V?

17.4.4 The Heisenberg Uncertainty Principle

A video which tries to answer the question: "What is the Uncertainty Principle?"

https://www.youtube.com/watch?v=7vc-Uvp3vwg

When we take measurements in an experiment, there will always be experimental uncertainties or errors involved, no matter how good the measuring apparatus are. However, according to classical mechanics, there is no fundamental barrier imposed by nature to an ultimate refinement of the apparatus or experimental procedures. That is, the uncertainty in a measurement can be made indefinitely small by using more precise instruments.

Quantum theory predicts, however, that such a barrier *does* exist and there is a **definite limit** to the precision of any measurement. This limit is **not** due to the limitation of the measuring instrument; rather, it is **inherent in nature** as a result of the **wave-particle nature of matter**.

Imagine generating a wave in a string by periodically shaking it up and down. One would not be able to precisely point out where the wave is along the string, but once could certainly give a value for the wavelength, and hence, using de Broglie's equation, the momentum.

On the other hand, if one were to give the string a sudden jerk, one would get a relatively narrow bump travelling down the string. In this case, the location of the wave along the string is more defined, while the wavelength could not be stated (it is not even periodic). For intermediate cases, one could say that the wave has a fairly well-defined wavelength (and therefore momentum) and a fairly well-defined position, but there is an inescapable trade-off between the momentum and the position.

In 1927, Werner Heisenberg formally introduced this notion, which is now known as the **Heisenberg Uncertainty Principle**:

$$\Delta x \Delta p_x \gtrsim h$$

If a measurement of position of a particle is made with uncertainty Δx and a simultaneous measurement of linear momentum is made with uncertainty Δp_x, then the product of the two uncertainties can never be smaller than h, the Planck constant.

In general terms, the uncertainty principle states that it is impossible to know the exact values of two **related** physical quantities, such as the position and the momentum of a particle, **at the same time**.

Example 11

The speed of an electron is 5.00×10^3 m s^{-1} to an uncertainty of 0.00300%.

Find the inherent uncertainty in determining the position of this electron.

Tutorial 17C Wave-Particle Duality and the Uncertainty Principle

Wave-Particle Duality Self-Review Questions

S1 The wave nature of electrons is suggested by experiments on

A line spectra of atoms.

B the production of X-rays.

C the photoelectric effect.

D electron diffraction by a crystalline material.

UCLES N79/II/33

S2 Smaller objects may be distinguished in electron microscopes than in optical microscopes because

A electrons are smaller than visible quanta.

B the electrons travel faster than light.

C the electron wavelength is much shorter than that of visible light.

D the electrons are not diffracted.

UCLES N79/II/39

S3 Determine the momentum of light photons of frequency 5.0×10^{14} Hz.

CIE J76/II/13 modified

S4 **(a)** If the wavelength of an electron is equal to 5.00×10^{-7} m, how fast is it moving?

(b) If an electron has a speed of 1.00×10^{7} m s^{-1}, what is its wavelength?

S5 **Doorway Diffraction.** If your wavelength were 1.0 m, you would undergo considerable diffraction in moving through a doorway.

(a) What must be your speed be for you to have this wavelength? (Assume your mass is 60.0 kg.)

(b) At the speed calculated in part (a), how many years would if take you to move 0.80 m (one step)?

(c) Will you notice diffraction effects as you walk through doorways? Explain.

Young and Freedman (14E). P39.75

Uncertainty Principle Self-Review Questions

S6 A light source is used to determine the location of an electron in an atom to a precision of 0.05 nm. What is the uncertainty of the speed of the electron?

S7 A golf ball has a mass of 50 g and travels horizontally with a velocity of 80 m s^{-1}. What limit of uncertainty is placed on a measurement of its position by the Heisenberg uncertainty principle, if the uncertainty in the velocity is 0.01 m s^{-1}? Assume that the uncertainty in the mass is negligible.

Wave-Particle Duality Discussion Questions

D1 Electrons are accelerated in a vacuum before passing through a graphite film, as illustrated in Fig. 6.1. The electrons are then incident on a fluorescent screen.

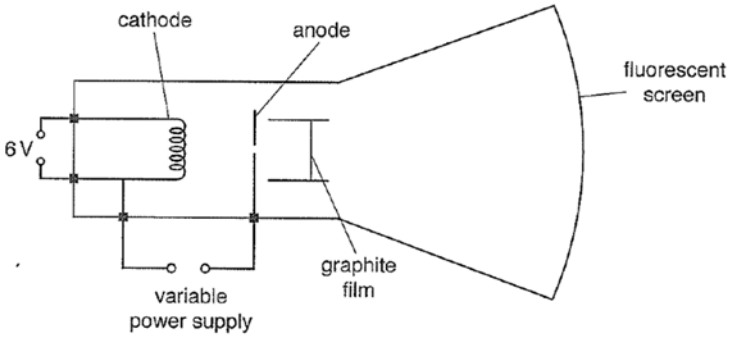


Fig. 6.1

Concentric rings of light are observed on the screen.

(a) Explain why the electrons produce concentric rings after passing through the graphite film. [1]

(b) The potential difference (p.d.) between the anode and the cathode is increased. This increase the brightness of the rings.

State and explain another effect of this change in p.d. on the pattern observed on the fluorescent screen. *Sketch diagrams to illustrate what will be seen on the fluorescent screen before and after the increase of the p.d to illustrate the change. [3]

(c) The electrons are accelerated from rest through a potential difference of 1.2 kV.

Calculate the de Broglie wavelength of the accelerated electrons. [4]

(d)* Comment on the wavelength you have calculated in relation to the size of an atom and what you can infer about the graphite film.

UCLES H2 N2017 P2 Q6 * Extended

D2 A narrow parallel beam of laser light is incident on a surface at an angle of incidence of 52°, as shown in Fig. 6.1.

The beam has a circular cross-section of diameter 1.2 mm.

The laser light has wavelength 680 nm and intensity 3.1 x 10^3 W m^{-2}.

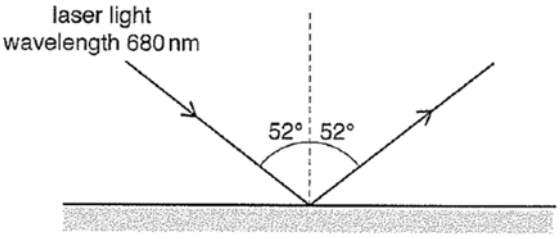


Fig. 6.1

(a) Determine the energy of a photon of the laser light. [2]

(b) Show that the number of photons incident per unit time on the surface is 1.2 x 10^{16} s^{-1}. [2]

(c) For the laser light incident on the surface in Fig. 6.1, only 55% of the photons are reflected and the remainder are absorbed.

Calculate the force *F* normal to the surface that is exerted by the laser light on the surface. [6]

UCLES H2 N2018 P3 Q6 (part)

D3 A stationary nickel nucleus of mass 9.95 X 10^{-26} kg emits a γ-ray photon of energy 1.17 MeV.

(a) For the γ-ray photon,

(i) show that its wavelength is approximately 1.06 X 10^{-12} m. [3]

(ii) calculate its momentum. [2]

(b) The direction in which the γ-ray photon is emitted is shown in Fig. 2.1.

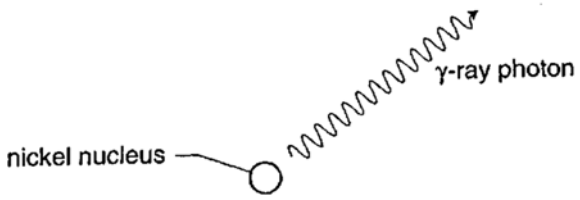


Fig. 2.1

(i) On Fig. 2.1, draw an arrow to show the direction of motion of the nickel nucleus after the emission of the photon. [1]

(ii) Calculate the speed of the nickel nucleus after the emission of the photon. [2]

(c) A second nickel nucleus that is moving emits a γ- ray photon.

Suggest, with a reason, whether the angle between the final direction of motion of the nucleus and that of the emitted photon will be the same as that in Fig. 2.1. [2]

UCLES H2 N2005 P2 Q2

Uncertainty Principle Discussion Questions

D4 A pulse of a radio wave lasts for 1.0 x 10^{-5} s. A photon of the radio wave may be considered to be at a point anywhere within this pulse, although the location of the point is not known.

(a) Calculate the length of the pulse, [1]

(b) By estimating the uncertainty in the position of the photon as your answer in (a), estimate the uncertainty in the momentum of the photon. [3]

UCLES H2 2007 Paper 3 Q7 (part, modified)

D5 A common, but erroneous description of the absolute zero of temperature is "that temperature at which all molecular motion ceases." How can the uncertainty principle be used to argue against this description?

D6 A beam of electrons is directed along a normal towards a barrier, as shown in Fig. 10.1.

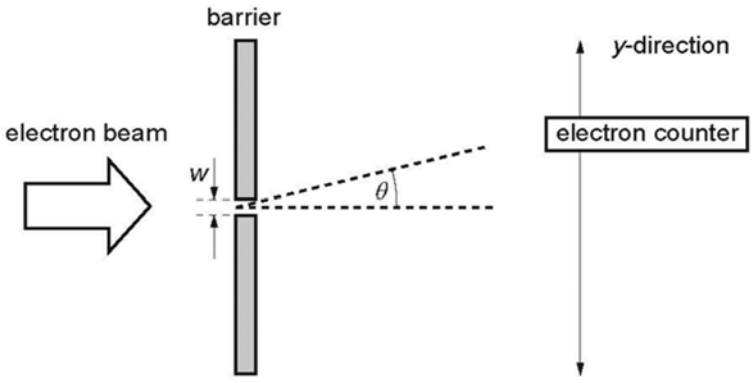


Fig. 10.1

The barrier contains a single slit of width *w*. Beyond the slit there is a detector that counts electrons. This can be moved in the *y*-direction to compare the rate of arrival of electrons at different values of the angle θ from the original direction of the beam.

(a) **(i)** Louis de Broglie suggested that some aspects of the behaviour of electrons can be explained using a wave model. Describe the main features of de Broglie's model of the electron. [3]

(ii) Use the wave model to explain how the electron beam spreads out beyond the slit. [2]

(b) State and explain one aspect of this experiment that cannot be explained using the wave model. [2]

(c) Werner Heisenberg used a different approach involving what is now known as the uncertainty principle. This can also be used to explain why the electron beam spreads out after passing through the slit. One version of this involves the equation

$$\Delta p \Delta y \gtrsim h$$

Explain how the terms below apply to electrons **as they pass through** the slit.

(i) Δy [2]

(ii) Δp [2]

(d) Hence use the uncertainty principle to explain why

(i) the beam spreads out, [2]

(ii) the beam is spread out more when the slit is narrower (smaller *w*), [3]

(iii) the beam is spread out less when the incoming electrons have smaller de Broglie wavelength. (You might find it helpful to include a vector diagram). [4]

CIE 9792 Jun 2010 Paper 3 Q10 (20 marks)

Answers to Tutorial 17C: Wave Particle Duality and Uncertainty Principle

Self-Review Questions

S1 D **S2** C **S3** 1.1 x 10^{-27} N s **S4(a)** 1460 m s^{-1}, **(b)** 7.28 x 10^{-11} m

S5(a) 1.11 x 10^{-35} m s^{-1}, **(b)** 2.29 x 10^{27} years **S6** 1 x 10^{7} m s^{-1} **S7** 1 x 10^{-30} m

Discussion Questions

D1(c) 3.54 x 10^{-11} m **D2(a)** 2.93 x 10^{-19} J, **(b)** 1.20 x 10^{16} s^{-1}, **(c)** 1.12 x 10^{-11} N

D3(a)(ii) 6.25 x 10^{-22} N s, **(b)(ii)** 6.28 x 10^{3} m s^{-1} **D4(a)** 3000 m, **(b)** 2.21 x 10^{-37} N s

Chapter 18

Nuclear Physics

Playlist for Lecture Examples: https://for.edu.sg/k82kxa

Content

- The nucleus
- Isotopes
- Nuclear processes
- Mass defect and nuclear binding energy
- Radioactive decay
- Biological effects of radiation

Learning Outcomes

Candidates should be able to:

(a) infer from the results of the Rutherford α-particle scattering experiment the existence and small size of the nucleus

(b) distinguish between nucleon number (mass number) and proton number (atomic number)

(c) show an understanding that an element can exist in various isotopic forms each with a different number of neutrons in the nucleus.

(d) use the usual notation for the representation of nuclides and represent simple nuclear reactions by nuclear equations of the form ${}^{14}_{7}\mathrm{N} + {}^{4}_{2}\mathrm{He} \rightarrow {}^{17}_{8}\mathrm{O} + {}^{1}_{1}\mathrm{H}$

(e) state and apply to problem solving the concept that nucleon number, charge and mass-energy are all conserved in nuclear processes

(f) show an understanding of the concept of mass defect

(g) recall and apply the equivalence between energy and mass as represented by $E = mc^2$ to solve problems

(h) show an understanding of the concept of nuclear binding energy and its relation to mass defect

(i) sketch the variation of binding energy per nucleon with nucleon number

(j) explain the relevance of binding energy per nucleon to nuclear fusion and to nuclear fission

(k) show an understanding of the spontaneous and random nature of nuclear decay

(l) infer the random nature of radioactive decay from the fluctuations in count rate

(m) show an understanding of the origin and significance of background radiation

(n) show an understanding of the nature of α, β and γ radiations (knowledge of positron emission is not required)

(o) show an understanding of how the conservation laws for energy and momentum in β decay were used to predict the existence of the neutrino (knowledge of antineutrino and antiparticles is not required)

(p) define the terms activity and decay constant and recall and solve problems using $A = \lambda N$

(q) infer and sketch the exponential nature of radioactive decay and solve problems using the relationship $x = x_0 \exp(-\lambda t)$ where x could represent activity, number of undecayed particles or received count rate

(r) define and use half-life as the time taken for a quantity x to reduce to half its initial value

(s) solve problems using the relation $\lambda = \frac{\ln 2}{t_{1/2}}$

(t) discuss qualitatively the effects, both direct and indirect, of ionising radiation on living tissues and cells.

Nine mini lectures, discussing several of the concepts of the chapter, were created and made available in SLS and on YouTube. You can find them all in the playlist "Nuclear Physics:" https://for.edu.sg/nbd29f.

18.1 Introduction

Everything around us consists of molecules, which are made up of atoms, which in turn contain protons, neutrons and electrons. One might wonder whether these can be broken down into even smaller constituents. The question of what the smallest component is, from which our physical world might be constructed, has always been a driving motivation for natural philosophers and scientists in their quest to uncover the constituent fabric of nature.

- **The Beliefs of the Early Greeks**
  - Early Greeks believed that all matter was composed of only four elements, fire, earth, water and air. Ancient cultures in Egypt, Babylonia, Japan, Tibet and India had similar lists.
  - The Greek philosopher Leucippus and his student Democritus were the first to propose that matter is made up of indivisible particles that they called *atomos,* a word derived from *a*-(not) and *tomos* (cut or derived). This idea was, however, not readily accepted.

- **Modern Atomic Theory - John Dalton (1766 - 1844)**
  - In the early 19th century, John Dalton used *scientific methods* to discover that many chemical phenomena could be explained if elements were made up of atoms, which he proposed to be basic, indivisible building blocks of matter. In Dalton's theory, all atoms of an element are identical, while different elements have atoms of differing size and mass. As it turned out, not all atoms of an element are identical, but they may have a varying mass (see section 18.3.3 on *isotopes* below).

- **The Electron**
  - By 1909, Millikan had completed his first measure of the elementary charge. In 1910, J.J. Thomson discovered the electron.
  - These and other experiments (e.g., Rutherford's alpha-particle scattering experiment) showed that all the mass of an atom had to be associated with the *positive charge* and not with the negatively charged electrons. The neutron had not yet been discovered at that time.
  - It was also known that the overall size of atoms is of the order of 10^{-10} m and all atoms except hydrogen contain more than one electron, unless they are ionised.

- **Thomson's Plum Pudding Model**

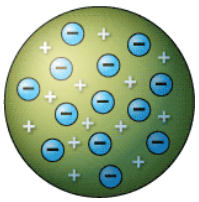


Figure 1.
Thomson's plum pudding model

Figure 2.
A plum pudding

  - At that time, what was *not* known was how the mass and charge were distributed within the atom. In 1904, Thomson proposed the "*plum pudding model.*" He proposed:
    - *the atom consists of a sphere of positive charge* (like the pudding), of the order of 10^{-10} m in diameter
    - electrons are embedded in the positively charged sphere (like the bits of plum in a more or less spherical muffin).

18.2 Rutherford's Alpha-Particle Scattering Experiment

Mini Lecture 1.1 – Rutherford's Alpha-Particle Scattering Experiment, https://for.edu.sg/bvvocd.

Hans Geiger and *Ernest Marsden* carried out a series of experiments under the direction of *Ernest Rutherford* in 1908-1913 at the *University of Manchester* in England, where they shot helium nuclei at a thin layer of gold foil. The results of the experiments disproved Thomson's plum pudding model.

18.2.1 Details of Rutherford's Experiment

- **Aim:** To probe the internal structure of the atom.

- **Description**

  In Rutherford's alpha(α)-particle scattering experiment, α-particles (helium-4 nuclei) from a radioactive source were directed towards a sheet of gold foil in a vacuum chamber.

  - The *vacuum* is necessary because alpha particles only have a short range of about 5 cm in air.
  - The gold foil, being very ductile, is rolled into a very thin sheet of 0.00004 cm (about 100 atoms) thick, so that the high-energy alpha particles can pass through it, instead of being absorbed by it.

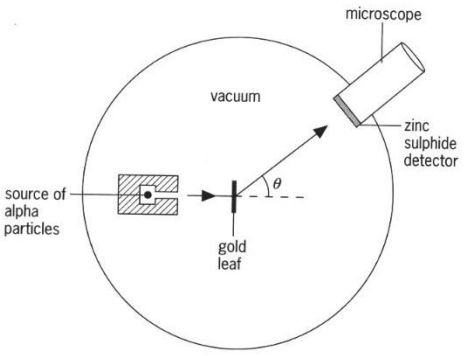


Figure 3.
Set up of Rutherford's Scattering Experiment.

  - The angular deflection of alpha particles is measured using a microscope that has a glass screen coated with zinc sulphide (Fig. 3). A charged particle impacting with the screen will produce a tiny flash of light, called a **scintillation**.
  - To determine how the alpha particles are scattered, the number of scintillations *N* is obtained for each angular deflection θ. Fig. 4 shows the results of the alpha-particle scattering experiment. Note that the vertical axis is marked in powers of 10. If a linear scale were used, no deflections over angles larger than 10 or 20 degrees would be visible, and the relevant fact that about 1 in 8000 alpha particles are backscattered would be lost.

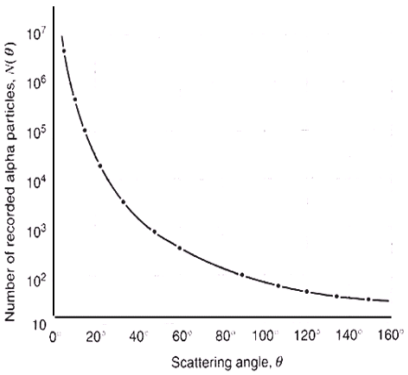


Figure 4*. Graph showing the results of the α-particle scattering experiment. Note that the vertical axis is marked in powers of 10.*

- **Predicted Alpha Scattering Based on Thomson's Model (Hypothesis)**

  - The positive charge and the negative electrons are distributed throughout the whole atom. Hence, the electric field inside the atoms should be quite small.
  - Alpha particles are doubly charged helium ions. As the electrical force on an α-particle that enters the atoms should be quite weak, the maximum deflection θ observed should at most be only a few degrees.

- **Actual Observation**

  - *More than 99% of α-particles passed through with little or no angular deflection.*
  - *A few α-particles (~1 in 8000) backscattered, i.e., suffered deflections of more than 90°. A very small number were even deflected backwards.*

    These results were so unexpected that Rutherford later wrote:

    "*It was quite the most incredible event that ever happened to me in my life. It was almost as incredible as if you had fired a 15-inch shell at a piece of tissue paper and it came back and hit you.*"

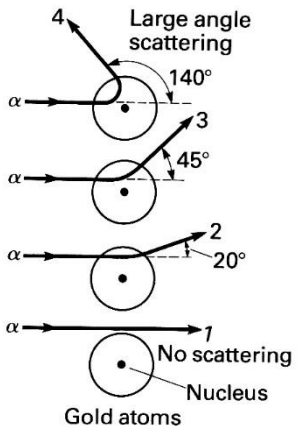


Figure 5. Scattering of alpha particles by the nucleus.

- **Interpretation of Results**

Rutherford concluded from the α-particle scattering experiment that:

i. all the positive charge and mass of the gold atom were concentrated in a very small volume within the atom

ii. the rest of the atom, other than the electrons, is empty space.

From the data, Rutherford also calculated that the radius of the nucleus should be less than 10^{-14} m, at least 10^4 times smaller than the radius of the atom. (Today, it is deduced that the typical size of a nucleus is about 10^{-15} m, i.e., 10^5 times smaller than the size of the atom.) He worked out a detailed formula for the scattering, which matched the experimental data to high precision, and later proposed a new structure for the atom.

Fig. 5 shows how alpha particles are scattered, depending on how close they pass the nucleus of a gold atom. **1**: Alpha particles that pass outside the gold atom altogether will experience it as an uncharged sphere and are not deflected at all. **2**: Alpha particles that pass through a gold atom's electron cloud experience a net positive charge and are deflected by the repulsive Coulomb force. **3**: The closer the alpha particle passes the nucleus, the greater the angle of deflection. **4**: About 1 in 8000 alpha particles hit a gold nucleus almost face-on, with the result that they are backscattered, i.e., scattered over angles larger than 90 degrees. About 1 in a million even get reflected in almost the same direction as where they came from.

18.2.2 Rutherford's Nuclear Model of an Atom[1]

Based on his experiment, Rutherford put forward his nuclear model of the atom, which had the following properties (Fig. 6).

(i) There is a positively charged centre in an atom called the nucleus. Nearly all the mass of the atom resides in the nucleus.

(ii) The electrons revolve around the nucleus in well-defined orbits.

(iii) The size of the nucleus is very small compared to the size of the atom.

https://for.edu.sg/rutherfordgoldfoilexperiment

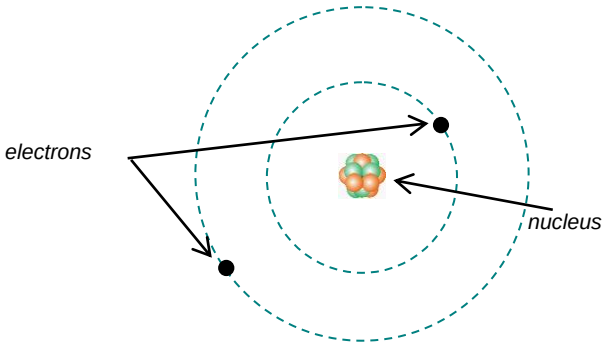


Figure 6. *Atomic structure proposed by Rutherford.*

[1] **Drawbacks of Rutherford's model of the atom:**

(i) It does not obey the Maxwell theory of electrodynamics which dictates that "an accelerated charged particle continuously emits energy." If an electron were to do this, it would continuously lose energy and spiral inwards, ultimately falling into the nucleus. However, atoms are observed to be quite stable.

(ii) It could not explain the discontinuous nature of the line spectrum of the hydrogen atom.

These problems were overcome by Niels Bohr's model of the atom, as outlined in the chapter on Quantum Physics.

Example 1: Rutherford Experiment
Young and Freedman 11th Edition - Page 1461 Example 38.5

An α-particle is aimed directly at a gold nucleus. An α-particle has two protons and a charge of +2*e*, while a gold nucleus has 79 protons and a charge of +79*e*.
What minimum kinetic energy must the alpha particle have to approach to within 5.0×10^{-14} m of the centre of the gold nucleus?
Assume that the gold nucleus, which has about 50 times the mass of an alpha particle, remains at rest.

Example 2: Size of a Nucleus
Akrill, Bennet and Millar (3rd Edition) - Page 159 Prob. 14.57

a) Estimate the size of the full stop at the end of this sentence.
b) If this dot represents the nucleus of a gold atom, how large, on this scale, would the atom be?
c) What is in the space around the nucleus?

18.3 Some Properties of the Nucleus

Mini Lecture 1.2 – Some Properties of the Nucleus: https://for.edu.sg/of9qdl.

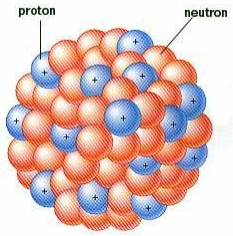


Figure 7*. The nucleus is approximately spherical and contains protons & neutrons.*

18.3.1 Basic Building Blocks of the Nucleus

The basic building blocks of the nucleus are the **proton** and the **neutron.**

- Experiments show that we can model a nucleus as a sphere with a radius *R* that depends on the total number of ***nucleons*** (neutrons and protons) in the nucleus (Fig. 7).
- In a **neutral** atom, the **number of protons in the nucleus is equal to the number of electrons orbiting the nucleus**.
- The masses and charges of these particles are given below:

	Proton, $^{1}_{1}$p	**Neutron, $^{1}_{0}$n**	**Electron, $^{0}_{-1}$e**
Mass	m_p 1.007276 u (1.672622×10^{-27} kg)	m_n 1.008665 u (1.674927×10^{-27} kg)	m_e 0.000548580 u (9.10938×10^{-31} kg)
Charge	+ *e*	0 (neutral)	- *e*

Elementary charge *e* = 1.60×10^{-19} C

Unified atomic mass u = 1.66×10^{-27} kg (see next page)

- From the table above, we can see that the nucleus contains over 99.9% of the mass of an atom.

18.3.2 Nuclide Notation

- A **nuclide** is a species of atom characterized by the specific constitution of its nucleus, i.e., its *number of protons* and its *number of neutrons.*
- The ***proton number*, *Z***, is the number of protons in the nucleus. *Z* is also called the *charge number* or the ***atomic number***.
- The ***nucleon number*, *A***, is the total number of **nucleons**. A nucleon is either a neutron or proton in the nucleus. *A* is also known as the ***mass number***.
- Therefore, the number of neutrons in the nucleus is $N = A - Z$
- The **notation** for a nuclide *X* with proton number *Z* and nucleon number *A* is $^{A}_{Z}X$.

18.3.3 Isotopes

- **Isotopes** are atoms that have the same number of <u>protons</u> but different numbers of <u>neutrons</u>.
- Examples of isotopes of hydrogen: $^{1}_{1}\mathrm{H}$ (hydrogen); $^{2}_{1}\mathrm{H}$ (deuterium); $^{3}_{1}\mathrm{H}$ (tritium).
  Examples of isotopes of carbon:

	Abundance in Nature	***Remarks***
$^{11}_{6}\mathrm{C}$	Trace	Radioactive: half-life of 20.4 minutes
$^{12}_{6}\mathrm{C}$	98.90 %	Stable
$^{13}_{6}\mathrm{C}$	1.10 %	Stable
$^{14}_{6}\mathrm{C}$	Trace	Radioactive: half-life of 5730 years

- Isotopes have nearly identical chemical properties (same proton number), but different physical properties (different nucleon number).

18.3.4 Common Units used in Nuclear Physics

- **Atomic Mass:** atomic masses are usually expressed in terms of the unified atomic mass unit, u.
- **Atomic Mass Unit (u):** one unified atomic mass unit is one-twelfth the mass of the carbon-12 atom.

Hence, $1\ u = \frac{1}{12}\left(\frac{\text{mass of 1 mole of carbon-12}}{\text{Avogadro's number}}\right) = \frac{1}{12}\left(\frac{0.012\ \text{kg}}{6.02\times10^{23}}\right) = 1.66\times10^{-27}\ \text{kg}$.

Notes:
- ✓ The mass number *A* should not be confused with the atomic mass. Mass number is an integer, specific to a nuclide, and has no unit – it is simply a count of the number of nucleons.

- **The Electron-Volt:** in nuclear physics, the unit of energy called an electron-volt (eV) is often used:

$1\ \text{eV} = (1.60 \times 10^{-19}\ \text{C})(1\ \text{J/C}) = 1.60 \times 10^{-19}\ \text{J}$

Usually, slightly larger units are more convenient, such as:

$1\ \text{keV} = 1.0 \times 10^{3}\ \text{eV}$ $(= 1.60 \times 10^{-16}\ \text{J})$ and $1\ \text{MeV} = 1.0 \times 10^{6}\ \text{eV}$ $(= 1.60 \times 10^{-13}\ \text{J})$

18.4 Nuclear Reactions

The nuclei of atoms can change, or be made to change, in various ways. Changes to the nuclei of atoms are called nuclear reactions.

Examples

- **Radioactive decay**
  Radioactive decay is a process where an unstable nucleus *spontaneously* and *randomly* emits radiation, changing into a more stable nucleus.

- **Nuclear fission**
  Nuclear fission occurs when a massive nucleus splits into two less massive nuclei of approximately the same nucleon number.

- **Nuclear fusion**

Mini Lecture 2.2 – Nuclear Fusion: https://for.edu.sg/b86k00.

Nuclear fusion occurs when two less massive nuclei combine (fuse) to form a single more massive nucleus.

This can only occur when the nuclei have very high energy (e.g., at very high temperatures). The reason for this is that, as the nuclei approach each other, kinetic energy is converted into electrical potential energy. If the nuclei lose all their kinetic energy before they bind, fusion cannot occur (see Example 1).

18.4.1 Nuclear Reactions and Conservation Laws

Mini Lecture 1.3 – Nuclear Reactions: https://for.edu.sg/3uli87.

Nuclear reactions may be shown in a form similar to chemical equations, for example

$$ {}^{1}_{0}\mathrm{n} + {}^{14}_{7}\mathrm{N} \rightarrow {}^{14}_{6}\mathrm{C} + {}^{1}_{1}\mathrm{p} $$

The transformations of particles must follow certain conservation laws. In all nuclear reactions, **the following quantities** must be equal on each side of the equation:

- Sum of nucleon numbers
- Sum of proton numbers
- Net electric charge
- Total momentum
- Total energy

  - It is observed that in a nuclear reaction, the total mass of reactants and products is different. A decrease in mass is always accompanied by a release of energy, whereas an increase in mass is always accompanied by an absorption of energy.
  - The energy *E* associated with a change of mass *m* is given by the mass-energy equivalence relation $\boldsymbol{E = mc^2}$, where *c* is the speed of light.

Example 3: Nuclear Reaction Equations

Milner SAQ 1.5 and SAQ 1.6

a) The nuclear equation for a particular fission process is given by: $ {}^{235}_{92}\mathrm{U} + {}^{1}_{0}\mathrm{n} \rightarrow {}^{141}_{55}\mathrm{Cs} + {}^{?}_{?}\mathrm{Rb} + 2\ {}^{1}_{0}\mathrm{n} $

Determine i) the nucleon number, and ii) the proton number of the *rubidium* atom

b) Two deuterium nuclei can fuse to produce a tritium nucleus and a proton. Write a nuclear equation for this reaction.

Answer:

Example 4: Energy-Mass Equivalence

What is the energy equivalent, in MeV, of one unified atomic mass, 1 u?

Answer:

Example 5: Conservation of Mass Energy:
Serway College Physics p958 – main text (modified)

Consider the following nuclear reaction: $^{2}_{1}\mathrm{H} + {}^{14}_{7}\mathrm{N} \rightarrow {}^{12}_{6}\mathrm{C} + {}^{4}_{2}\mathrm{He}$

The masses of the nuclei in the above equation are given by:

Mass of $^{2}_{1}\mathrm{H}$ = 2.014 102 u Mass of $^{12}_{6}\mathrm{C}$ = 12.000 000 u

Mass of $^{14}_{7}\mathrm{N}$ = 14.003 074 u Mass of $^{4}_{2}\mathrm{He}$ = 4.002 602 u

a) What is the mass "lost" by the system in the nuclear reaction?
b) What is therefore the equivalent energy "lost" by the system, in MeV?
c) What is this “lost” energy manifested as?
d) Suggest why the combined energies of the products are usually higher than the value calculated in (b).

Example 6: Conservation of Mass-Energy:
(N82/I/11 – modified)

A nitrogen nucleus $^{14}_{7}\mathrm{N}$, bombarded by an alpha particle with a certain kinetic energy, transmutes to an oxygen nucleus $^{17}_{8}\mathrm{O}$ and a proton.

a) Write an equation for this nuclear reaction, showing the mass numbers and atomic numbers of the particles involved.
b) Find the minimum initial kinetic energy the alpha particle must have in order for this reaction to occur.
c) By also considering the linear momentum of the nuclei, suggest why it is unlikely for this reaction to occur, even if the alpha particle were to possess this minimum energy.

Given:

Mass of $^{4}_{2}\mathrm{He}$ = 4.002 602 u Mass of $^{17}_{8}\mathrm{O}$ = 16.999 133 u

Mass of $^{14}_{7}\mathrm{N}$ = 14.003 074 u Mass of $^{1}_{1}\mathrm{H}$ = 1.007 825 u

18.5 Nuclear Force & Nuclear Binding Energy

Mini Lecture 1.4 – Nuclear Force & Nuclear Binding Energy: https://for.edu.sg/g3fa5m.

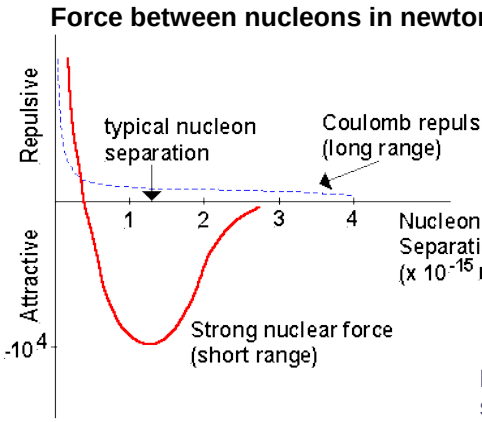
An atomic nucleus consists of positively charged protons and neutral neutrons, all concentrated in a tiny volume in space. Electrostatic repulsion between the positively charged protons should have caused the disintegration of the nuclei spontaneously. However, we know that most nuclei are very stable, except for radioactive nuclei.

It dawned on physicists that there must be another force, other than the electrostatic force, which holds the nucleons together. This is now known as the strong nuclear force.

18.5.1 Strong Nuclear Force & the Stability of Nuclei

- **Nature and Range of the Strong Nuclear Force**

  - At typical nucleon separations of 1.3 fm (1.3 x 10^{-15} m), it is a very strong attractive force: ~10^4 N.
  - At much smaller separations between nucleons, the force becomes repulsive.
  - Beyond about 1.3 fm separation, the force quickly drops to zero.
  - Thus, the nuclear force is a **short-range force**.
  - The much smaller Coulomb force between protons has a much longer range and becomes the only significant force between protons when their separation exceeds about 2.5 x 10^{-15} m.



https://for.edu.sg/strongnuclearforce

Figure 8. *Graph of forces between two n*

- **The nuclear force is independent of charge**

  The nuclear forces associated with proton-proton, proton-neutron and neutron-neutron interactions are the same.

18.5.2 Stability of a Nucleus

Apart from the nucleus of the $^{1}_{1}\mathrm{H}$ atom, *there is a very delicate balance between the strong nuclear forces that act between adjacent nucleons and the electrostatic forces that act between the protons*. This determines how **stable** a particular nucleus is.

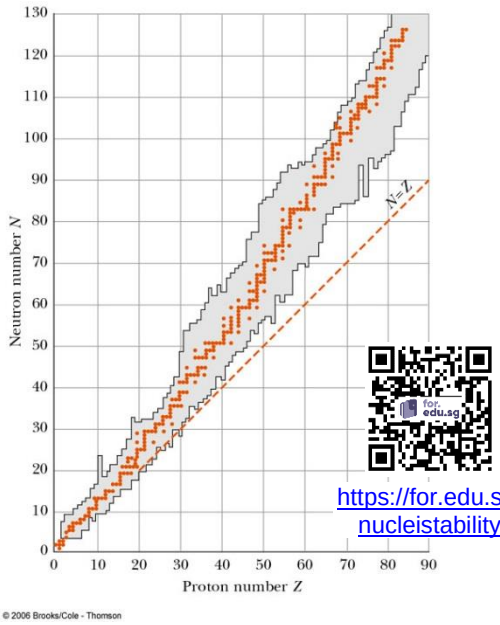
Out of the 339 naturally occurring nuclides, 254 are stable and 85 are unstable. Beyond the 339 naturally occurring nuclides, more than 3000 radionuclides have been artificially produced.

A plot of *N* vs. *Z* for a number of stable nuclides is given in Fig. 9.

- Note that for **light nuclides**, they are *most stable if they contain almost equal numbers of neutrons and protons* (i.e. *N* = *Z*).
- **Heavier nuclides** are *more stable* if ***N* > *Z***.

  This difference can be partially understood by recognizing that *as the number of protons increases, the electrostatic repulsive force increases*. Hence, *more neutrons are needed to keep the nucleus stable* as neutrons are affected only by the attractive strong nuclear force.
- **Beyond *Z* = 83**, there are no stable nuclides. This is because the repulsive forces between the protons can no longer be compensated by the addition of neutrons.

  Nuclides with more than 83 protons are unstable and tend to decay or disintegrate into other nuclides.



https://for.edu.sg/nucleistability

Figure 9.

A plot of the neutron number N versus the proton number Z for the stable nuclides (solid points). The dashed line corresponds to N = Z. The plots are centred on the so-called line of stability. The shaded area shows radioactive (unstable) nuclides.

18.5.3 Nuclear Binding Energy

- Like the ionisation energy (also known as electron binding energy), that has to be supplied to an atom in the ground state (lowest energy level) to completely liberate an orbital electron, work has to be done against the strong nuclear force to disassemble a nucleus into individual unbound protons and neutrons. This *nuclear binding energy* is typically about a million times larger than a typical electron binding energy.

> The ***binding energy*** *of a nucleus* is the energy required to separate the nucleus into its constituent free neutrons and protons (Fig. 10).

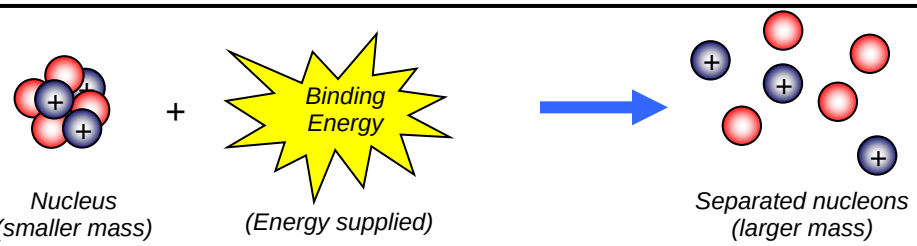


Figure 10. *Energy must be supplied to break up a nucleus into its constituent neutrons and protons.*

- This binding energy is also the *energy that is released when free protons and neutrons are bound together in a nucleus* (Fig. 11).

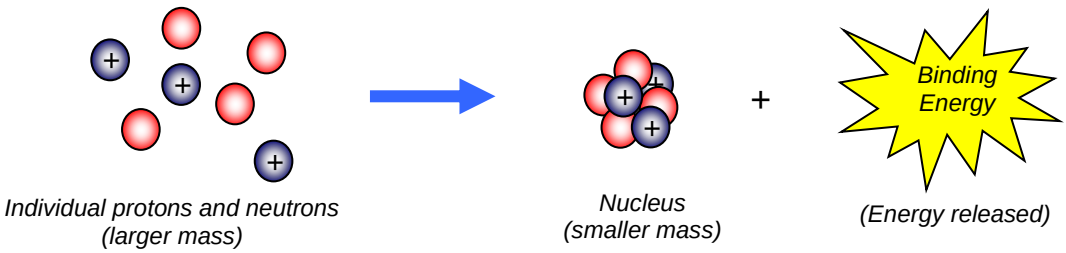


Figure 11. *If the nucleons bind together to form a nucleus, energy will be released.*

18.5.4 Mass Defect

- Recall that *mass* and *energy* are equivalent and are related by Einstein's equation $\boldsymbol{E = mc^2}$.

- Since energy is released when the individual protons and neutrons bind to form a stable or semi-stable nucleus, **the nucleus will have a smaller mass than the total mass of the separate neutrons and protons.**

  This difference in mass Δm is known as the ***mass defect***.

- To calculate the mass defect Δm for a ***nucleus*** that has Z protons and N neutrons:

$$\Delta m = (\textit{Sum of mass of constituent nucleons}) - (\textit{Mass of nucleus})$$

$$\Delta m = (Zm_p + Nm_n) - (M_{nucleus})$$

  where m_p is the mass of a proton, m_n is the mass of a neutron and $M_{nucleus}$ is the mass of the nucleus.

- It therefore follows from Einstein's mass-energy-equivalence relationship that:

  $\textit{Binding Energy } E = (\Delta m)\, c^2$ where Δm is the mass defect.

https://for.edu.sg/nuclearbindingenergy

Example 7: Binding Energy of Deuterium
Serway College Physics p943

The nucleus of a deuterium atom, called a deuteron, consists of one proton and one neutron. Calculate the deuteron's binding energy in MeV, given that its nuclear mass is 2.013 553 u.
Mass of proton m_p = 1.007 276 u
Mass of neutron m_n = 1.008 665 u

18.5.5 Binding Energy per Nucleon (B.E./*A*)

- Although the larger an atomic nucleus, the greater its total binding energy, it is the **binding energy per nucleon that determines the *stability* of the nucleus.**

> The **binding energy per nucleon** represents *the average energy required to remove a nucleon from the nucleus*.

- Therefore, the **greater the binding energy per nucleon, the more stable the nucleus**.

Element	Mass of protons & neutrons/ u	Nuclear mass/ u	Binding energy / MeV	(B.E./*A*) / MeV
Deuterium	2.01594	2.01355	2.23	1.12
Helium-4	4.03188	4.00151	28.29	7.07
Lithium-7	7.05649	7.01336	40.15	5.74
Beryllium-9	9.07243	9.00999	58.13	6.46
Iron-56	56.44913	55.92069	492.24	8.7906
Silver-107	107.86187	106.87934	915.23	8.55
Iodine-127	128.02684	126.87544	1072.53	8.45
Lead-206	207.67109	205.92952	1622.27	7.88
Polonium-210	211.70297	209.93683	1645.16	7.83
Uranium-235	236.90849	234.99351	1783.80	7.59
Uranium-238	239.93448	238.00037	1801.63	7.57

Figure 12a. *Binding energies per nucleon for some common elements*

- A *graph of binding energy per nucleon against the number of nucleons (Fig. 12b)* shows several interesting features:

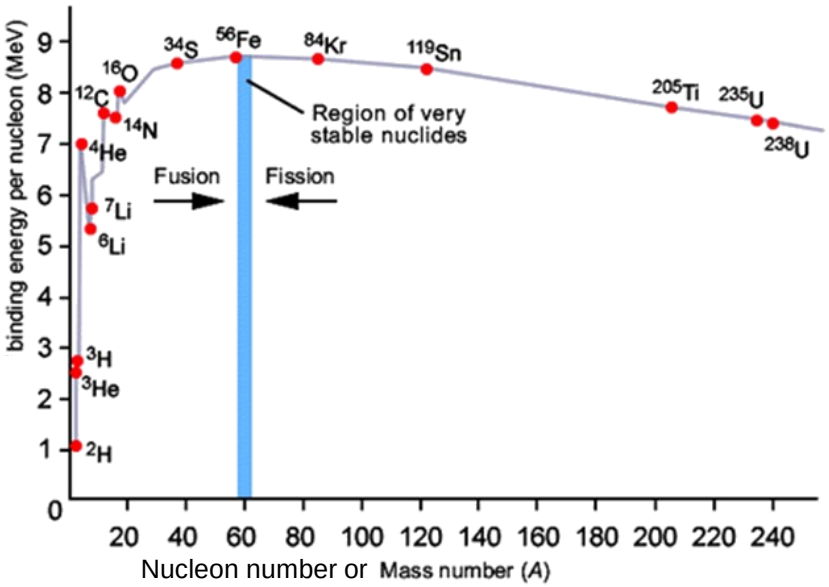


Figure 12b. *Graph of binding energy per nucleon vs. nucleon number*

1) Very small nuclei and very large nuclei are both less stable than nuclei of intermediate sizes.
2) The average binding energy per nucleon is about 8 MeV.
3) ^{56}Fe is one of the most stable and abundant nuclei. It resides close to the top of the curve, with a binding energy per nucleon of 8.8 MeV. The most stable nucleus of all is that of nickel-62.
4) Very small nuclei tend to form more stable heavier nuclei through **fusion**.
5) Similarly, very large nuclei can be split into smaller, more stable nuclei through **fission.**

Example 8: Binding Energy per Nucleon

TYS N85/II/11 (part)

The conventional notation for a certain nuclide is $^{81}_{35}\mathrm{Br}$.

a) Write down (i) the number of protons, (ii) the number of neutrons, in its nucleus.

b) The mass of its nucleus is 80.8971 u. Taking the proton and neutron masses to be 1.0073 u and 1.0087 u respectively, deduce the binding energy per nucleon.

c) Sketch a labelled graph showing how the binding energy per nucleon varies with mass number, indicating the approximate position of $^{81}_{35}\mathrm{Br}$ on the curve.

d) Explain the implications for the stability of a given nuclide of a low value of binding energy per nucleon.

Example 9: Binding Energy per Nucleon

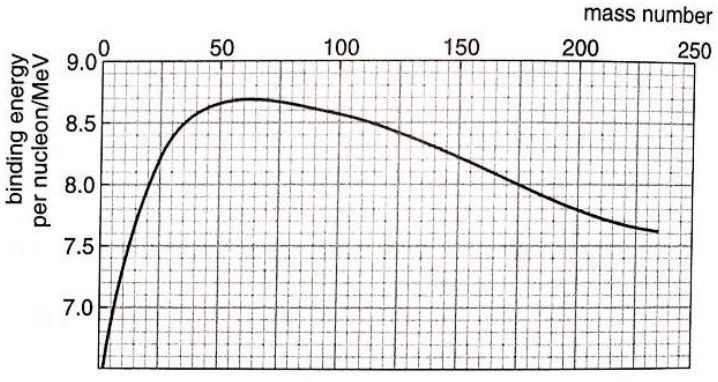
TYS N96/III/6 (part, modified – graph from Akrill, Benett and Millar)

The binding energy per nucleon varies with nucleon number in the way shown in the figure. During one particular fission process, a uranium-236 nucleus gives, among its fission products, a strontium-90 nucleus and a xenon-143 nucleus.

a) Use the values on the figure to calculate the energy released during this fission process.

b) What other fission particles are produced by this process? Why do these particles not have to be taken into account in the calculation in (a)?

c) Why does a release of energy occur when there is an increase in the binding energy?

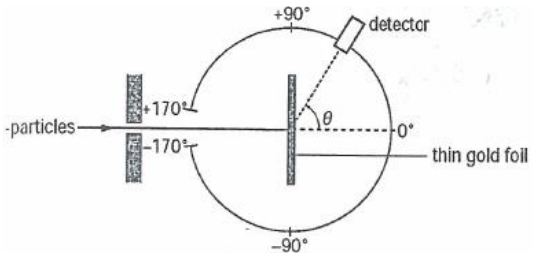


Tutorial 18A: Basic Concepts of Nuclear Physics

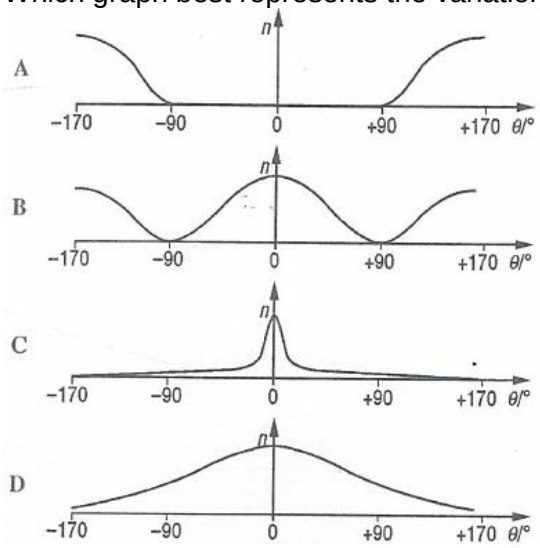
Self-Review Questions

S1 N02/I/30

In an *α*-particle scattering experiment, a student determined the number *n* of *α*-particles incident per unit time on a detector held at various angular position *θ*.

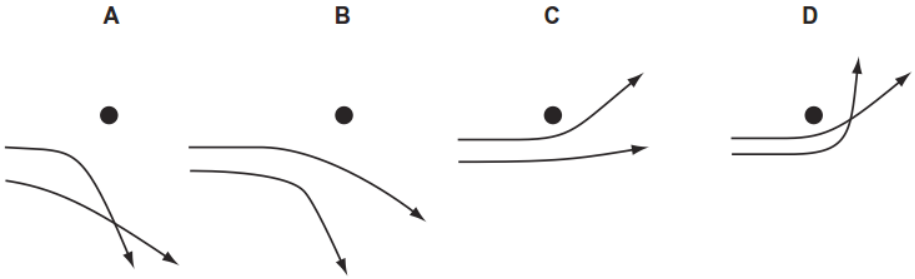


Which graph best represents the variation of *n* with *θ*?



S2 (N13 CIE9702/13/1/40)

Two *α*-particles with equal energies are deflected by a large nucleus.
Which diagram best represents their paths?



S3 N10/I/39

Which pair of nuclides has nuclei containing the same number of neutrons?

A $^{107}_{47}Ag$ and $^{104}_{45}Rh$
B $^{109}_{47}Ag$ and $^{109}_{46}Pd$
C $^{108}_{46}Pd$ and $^{109}_{47}Ag$
D $^{105}_{45}Rh$ and $^{106}_{47}Ag$

S4 N07/I/40

Nuclei of atoms can exist in excited states. When an excited nucleus returns to its state of lowest energy (the ground state), a *γ*-ray photon may be emitted.
The mass of a nucleus in its ground state is 59.9308 u. The energy of the photon emitted when this nucleus returns from an excited state to the ground state is 2.13×10^{-13} J.
What is the mass of the nucleus in the excited state?

A 59.9280 u **B** 59.9294 u **C** 59.9322 u **D** 59.9337 u

S5 N90/III/6 (part)

An atom of magnesium consists of:

12 electrons, each of mass 0.00055 u;

12 protons, each of mass 1.00728 u;

13 neutrons, each of mass 1.00866 u.

(i) Which particles are in the nucleus and what is the nucleon number (mass number) of this nucleus?

(ii) Which number determines that the atom is an atom of magnesium rather than of any other element?

(iii) The mass of this magnesium atom is 24.98584 u. Calculate the total mass of the constituent particles. Explain the difference.

(iv) Calculate the binding energy of the atom in joules.

S6 N09/I/39

The nucleus Z has the notation ${}^{y}_{x}Z$. The mass defect of this nucleus is Δm.

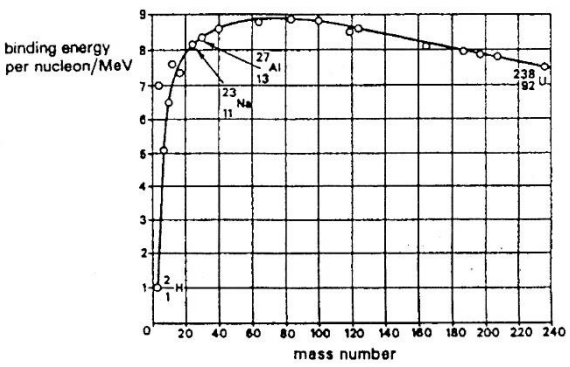
What is the binding energy per nucleon of the nucleus?

A $\dfrac{\Delta m}{x}$ **B** $\dfrac{\Delta m}{y}$

C $\dfrac{c^2\Delta m}{x}$ **D** $\dfrac{c^2\Delta m}{y}$

S7 J89/I/27

The diagram shows a graph of the binding energy per nucleon for a number of naturally occurring nuclides, plotted against their mass numbers.



Which of the following statements is a correct deduction from the graph?

A Of the nuclides plotted, ${}^{2}_{1}\mathrm{H}$ is the most stable.

B Energy will be released if a nucleus with a mass number greater than about 80 undergoes fusion with another nucleus.

C Energy will be released if a nucleus with a mass number less than about 80 undergoes fission as a result of particle bombardment.

D ${}^{27}_{13}\mathrm{Al}$ will not spontaneously emit an alpha particle to become ${}^{23}_{11}\mathrm{Na}$.

E ${}^{238}_{92}\mathrm{U}$ is the stable endpoint of a number of radioactive series.

Discussion Questions

D1 N18/III/1 & N21/II/7 (Adapted)

(a) An *α*-particle is approaching a stationary gold nucleus.
The *α*-particle is deviated through an angle of approximately 45° as it passes the gold nucleus.

On Fig. 1.3, sketch the path of the *α*-particle as it passes the gold nucleus.

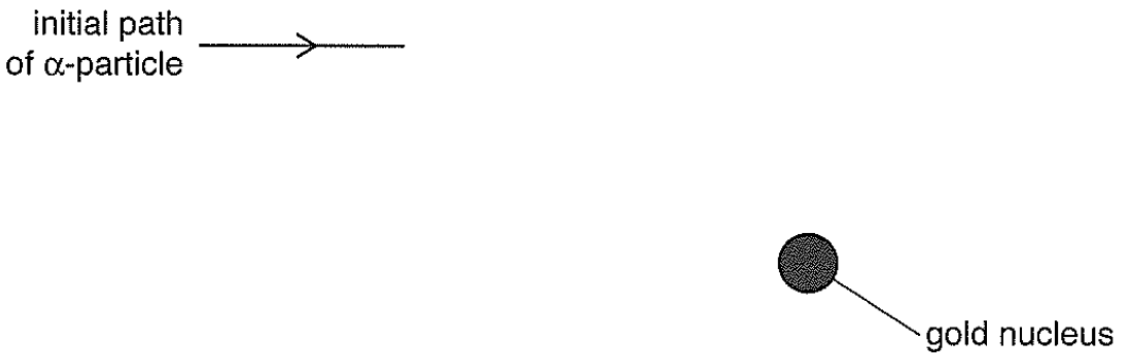


Fig. 1.3

(b) State and explain the evidence from Rutherford's alpha scattering experiment that changed the model of the atom in terms of its charge distribution and its mass distribution.

D2 N86/II/12 (part) – modified.

In an *α*-particle scattering experiment, an *α*-particle and a gold nucleus (in a piece of thin gold foil) collide head on and the alpha particle rebounds.
[for the α-particle, $A = 4$, $Z = 2$; for the gold atom, $A = 197$, $Z = 79$]

(a) Using the data given, write down a numerical expression for the electrostatic force of repulsion F acting on the alpha particle at the instant of collision when the distance between the particle and the gold nucleus is x.

(b) Also write down an expression for the electric potential energy of the system of the *α*-particle and the gold nucleus when the distance between them is x.

(c) Hence or otherwise, sketch the variation of

(i) the electrostatic potential energy V,

(ii) the kinetic energy T

of the alpha particle with x in one graph. Label the two curves appropriately.

(iii) State the relationship between the two graphs drawn in (i) and (ii).

(d) The distance of closest approach of the *α*-particle to the gold nucleus is s.

(i) Calculate a value for s, given that the initial kinetic energy of the *α*-particle is 1.8 MeV. Assume that the gold nucleus remains at rest throughout the collision.

(ii) What indication does the value of s give about the size of the particles involved?

D3 N93/III/6 (part)

When beryllium is bombarded with α-particles of energy 8.0 x 10^{-13} J, carbon atoms are produced, together with very penetrating radiation. The nuclear reaction might be

EITHER

(i) $^{9}_{4}Be + {}^{4}_{2}He \rightarrow {}^{13}_{6}C + \gamma$

OR

(ii) $^{9}_{4}Be + {}^{4}_{2}He \rightarrow {}^{12}_{6}C + {}^{1}_{0}n$

(a) Explain what is meant by $^{13}_{6}\mathrm{C}$.

(b) The energy of the penetrating radiation is found to be at least 8.8 x 10^{-12} J for each γ or $^{1}_{0}n$ produced. Show that equation (i) cannot be valid. Explain your reasoning carefully.

Nuclide	Mass/u
$^{9}_{4}\mathrm{Be}$	9.0150
$^{4}_{2}\mathrm{He}$	4.0040
$^{13}_{6}\mathrm{C}$	13.0075

[7]

D4 TYS N97/III/6 (part)

When uranium-235 nuclei are fissioned by slow-moving neutrons, two possible reactions are

Reaction 1:

$^{235}_{92}\mathrm{U} + {}^{1}_{0}\mathrm{n} \rightarrow {}^{139}_{54}\mathrm{Xe} + {}^{95}_{38}\mathrm{Sr} + 2\,{}^{1}_{0}\mathrm{n} + \text{energy}$

Reaction 2:

$^{235}_{92}\mathrm{U} + {}^{1}_{0}\mathrm{n} \rightarrow 2\,{}^{116}_{46}\mathrm{Pd} + x\mathrm{c} + \text{energy}$

https://for.edu.sg/nuclearfission

(i) For reaction 2, identify the particle c and state the number *x* of such particles produced in the reaction. [2]

(ii) The binding energy per nucleon *E* for a number of nuclides is given in the table below.

nuclide	***E* / MeV**
$^{95}_{38}\mathrm{Sr}$	8.74
$^{139}_{54}\mathrm{Xe}$	8.39
$^{235}_{92}\mathrm{U}$	7.60

1. State what is meant by the *binding energy per nucleon* of a nucleus.
2. Show that the energy released in reaction 1 is 210 MeV.
3. The energy released in reaction 2 is 163 MeV. Suggest, with a reason, which one of the two reactions is more likely to occur. [7]

D5 N03/II/8. Data Analysis (part)

The large amount of energy released in a nuclear fission reaction, together with the emission of more than one neutron, has made it possible for neutron-induced fission to be used as a source of useful energy.

When a neutron is captured by a uranium-235 nucleus, it causes the nucleus to fission. On average, 2.5 neutrons are emitted in these fission reactions. This is illustrated in Fig. 8.1.

When the conditions are suitable, a chain reaction can occur. If this chain reaction is not controlled, an explosion is likely.

However, if the chain reaction is controlled, as in a nuclear reactor, a source of continuous power may be created.

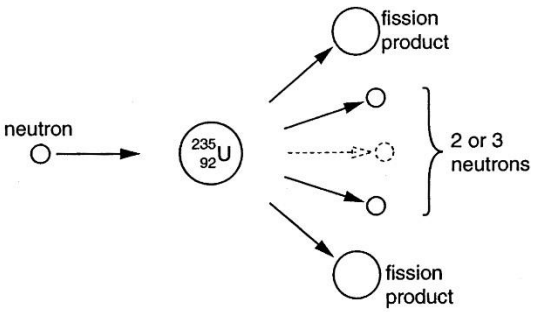


Fig. 8.1

(a) **(i)** Explain, with the aid of a diagram, what is meant by a *chain reaction*.

(ii) Suggest why, in an uncontrolled chain reaction where all neutrons are captured by uranium-235 nuclei, the majority of the energy is released during the final stages of the fission of a sample of the uranium. [3]

(b) The induced fission reaction of uranium-235 may be represented by a nuclear equation of the form

$$^{235}_{92}\mathrm{U} + {}^{1}_{0}n \rightarrow {}^{a}_{b}\mathrm{P} + {}^{c}_{d}\mathrm{Q} + (2 \text{ or } 3)\ {}^{1}_{0}n + 198\ \mathrm{MeV}$$

The fission products P and Q have approximately equal masses. However, when any two nuclei are fissioned, the fission products may not be the same. If a sample of uranium-235 fissions, many different fission products will be produced. The percentage amount of each fission product in the fissioned material is referred to as the *percentage yield*. The variation with nucleon number of the percentage yield of different fission products is referred to as a 'fission yield curve' and is illustrated in Fig. 8.2.

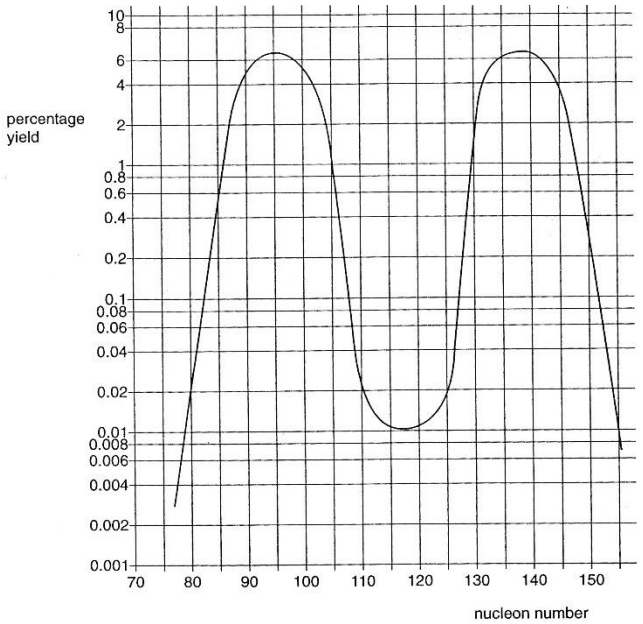


Fig. 8.2

(i) Suggest why the percentage yield is shown on a logarithm scale. [1]

(ii) Use Fig. 8.2 to determine the nucleon number of those fission products that have the same percentage yield as the nuclide with a nucleon number of 82. [2]

(iii) By reference to the nuclear equation and your answer in **(ii)**, suggest the nucleon number of the nuclide that would be produced in the same fission reaction as the nuclide with nucleon number 82. [1]

(iv) Use Fig. 8.2 to determine the percentage yield of fission products having nucleon numbers of 95 and 139. [1]

(v) Hence, show that the fission products in **(iv)** are about 600 times more likely to be produced than those having masses equal to each other. [1]

(c) The energy released in the reaction in **(b)** occurs partly as kinetic energy of the fission products (167 MeV) and of the neutrons (5 MeV).

(i) Suggest **one** other mechanism by which energy is released in the fission reaction. [1]

(ii) In a nuclear power station, 25% of the energy of the fission products is converted into electrical energy. Calculate, for the fission of a mass of 1.0 kg of Uranium-235,

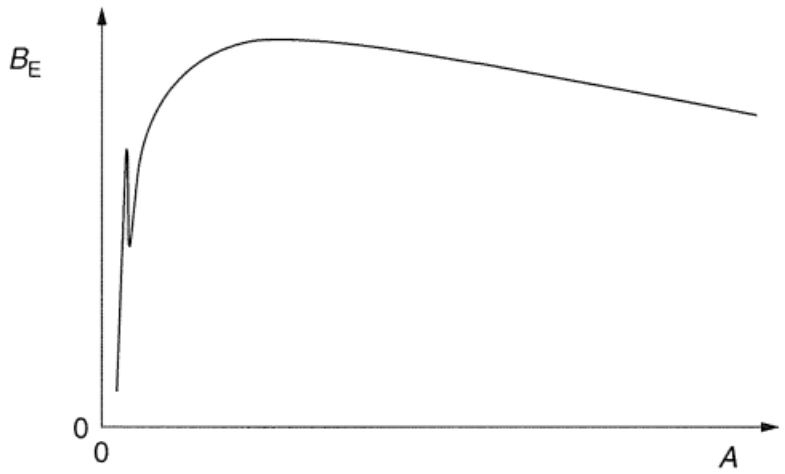
1. the number of nuclei in 1.0 kg of uranium-235. [2]
2. the electrical energy generated, [3]
3. the average power output, in megawatts, of the power station if the uranium is fissioned in a time of 24 hours. [2]

D6 (N2011/3/5)

An induced nuclear fission reaction may be represented by the equation

$$^{235}_{92}U + ^{1}_{0}n \rightarrow ^{141}_{56}Ba + ^{92}_{36}Kr + 3\,^{1}_{0}n + \text{energy}$$

The variation with nucleon number A of the binding energy per nucleon B_E is illustrated as follows.



(a) State an approximate value, in MeV, for the maximum binding energy per nucleon.

(b) On the above diagram, mark approximate positions for the nuclei of

(i) uranium-235 (label the position U),

(ii) barium-141 (label the position Ba),

(iii) krypton-92 (label the position Kr).

(c) By reference to binding energy per nucleon, explain why energy is released in this fission reaction.

Selected Answers to Tutorial 18A

S1	C	**S2**	A	**S3**	C	**S4**	C
S5	**(iii)** 25.20654 u	**(iv)**	3.30×10^{-11} J				
S6	D	**S7**	D				

D2	**(d)**	**(i)**	1.26×10^{-13} m	
D5	**(b)**	**(ii)**	108, 127, 152	
	(b)	**(iii)**	152	
	(c)	**(ii)**	**1.**	2.56×10^{24}
	(c)	**(ii)**	**2.**	1.71×10^{13} J
	(c)	**(ii)**	**3.**	198 MW

18.6 Radioactivity

As we have seen, the stability of a nucleus is determined by its binding energy per nucleon. One of the possible routes via which unstable nuclei can be transmuted into more stable nuclei is via *radioactive decay*.

https://for.edu.sg/whatisradioactivity

- Radioactive decay is a **spontaneous** and **random** process where an unstable nucleus changes into a more stable nucleus, emitting radiation as it does so.

  - **Spontaneous Process**:
    When we say that radioactive decay is a ***spontaneous*** process, it means that radioactive decay is not triggered by external factors or influences and that the rate of decay is not affected by environmental conditions.

  - **Random Process:**
    When we say that radioactive decay is a ***random*** process, it means that it is not possible to predict which nucleus will decay or when a particular nucleus will decay, while the *probability of decay per unit time* is constant and the same for all nuclei in the sample.

- Hence, the decay of a radioactive atom *is not affected by any chemical reaction*. It c*an neither be sped up nor slowed down* and is *independent of physical conditions* such as temperature, pressure and the decay of other atoms.

- **Evidence of Randomness of Radioactivity**

  - The random nature of radioactive decay can be demonstrated by observing the fluctuations in *count rate*. The *activity* (see 18.6.2 below) of a source is the rate at which a source of unstable nuclei decays. If you use a detector to measure the number of emissions, you will only record a fraction of the total emission. This is because normal detectors do not usually surround the radioactive source. However, the detected *count rate* will generally be *directly proportional* to the *activity* of the source.
  - When a Geiger-Müller tube (see appendix A) is placed near a source, the fluctuations in the count rate recorded by the tube show the randomness of radioactive decay (Fig. 13).

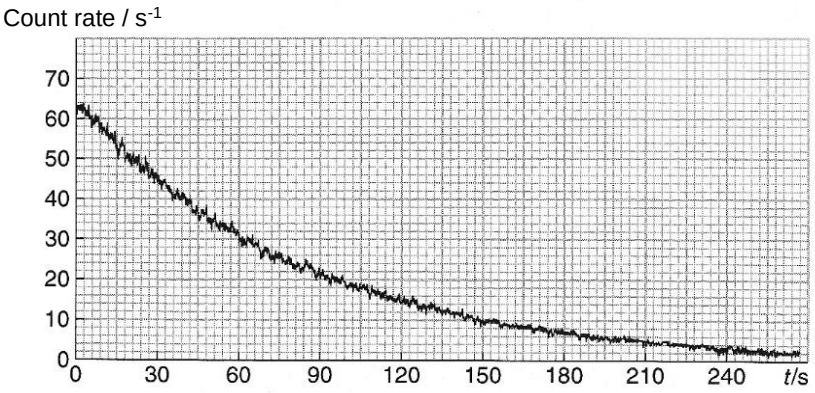


Figure 13. *The fluctuations in the count rate of a radioactive decay demonstrate its random nature.*

18.6.1 Types of Radioactive Decay

Mini Lecture 2.1 – Types of Radioactive Decay: https://for.edu.sg/j36xk2.

Alpha Decay

- An *α*-particle is the nucleus of a helium-4 ($^4_2\mathrm{He}$) atom, with a charge of +2*e*. It may be emitted from an unstable nucleus.
- Alpha decay occurs when an unstable heavy nucleus (the "parent") ejects an alpha particle, transforming itself into a different nuclide (the "daughter") after losing two protons and two neutrons (Fig. 14).
- Alpha decay can be written symbolically as: $^A_Z\mathrm{X} \rightarrow {}^{A-4}_{Z-2}\mathrm{Y} + {}^4_2\mathrm{He}$

  where $^A_Z\mathrm{X}$ is known as the **parent nucleus** and $^{A-4}_{Z-2}\mathrm{Y}$ is known as the **daughter nucleus**, e.g. ^{238}U is an alpha emitter and decays according to the scheme $^{238}_{92}\mathrm{U} \rightarrow {}^{234}_{90}\mathrm{Th} + {}^4_2\mathrm{He}$ (Fig. 14).

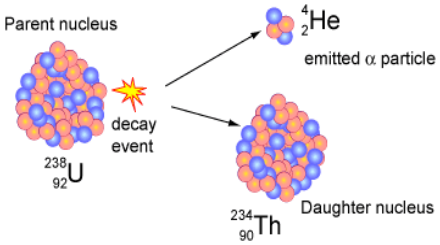


***Figure 14**. Alpha decay of U-238.*

- Alpha-emitting nuclei tend to be very heavy nuclei (with a very high proton number).
- **Energy and Momentum in Alpha Decay**

  Energy is released during an alpha decay. By the mass-energy equivalence principle, the combined mass of the daughter nucleus and the alpha particle is smaller than the mass of the parent nucleus. The energy released shows up in the ***kinetic energy* of the daughter nucleus and the *α*-particle.**

  Since the total linear momentum is conserved, if the parent nucleus was initially at rest, **the daughter nucleus and the *α*-particle must have momenta that are equal in magnitude and opposite in direction.** Since the *α*-particle has a much smaller mass, it carries away most of the kinetic energy.

  (Useful alternative expression of the kinetic energy of a particle: $\mathrm{KE} = \frac{p^2}{2m}$.

Example 10: Alpha Decay
TYS N83/II/13 (part)

A stationary radon atom may decay spontaneously into a polonium atom and a helium atom: $^{222}_{86}\mathrm{Rn} \rightarrow {}^{218}_{84}\mathrm{Po} + {}^4_2\mathrm{He}$

The masses of these atoms are: ^{222}Rn: 222.0176 u; ^{218}Po: 218.0090 u; ^{4}He: 4.0026 u.

a) Explain carefully how energy is conserved in this decay. Assuming that no γ ray is emitted, calculate the total kinetic energy of the decay products.

b) Explain carefully how momentum is conserved.

Beta Decay

- **Beta decay** is a type of radioactive decay in which a β-particle is emitted.
- A β-particle can be
  - a high-speed electron ejected from an unstable nucleus, referred to as a β^- particle, or
  - a high-speed positron[2] ejected from an unstable nucleus, referred to as a β^+ particle.

  (For the purpose of the A-level, a β-particle is an electron.)
- Beta decay can be:
  - **Beta-minus (β^-) decay**, where an **unstable neutron <u>in the nucleus</u> *transforms into a proton and an electron*** **and ejects the electron (Fig. 15a).**

    This decay process can be written as:

    $$^{A}_{Z}\mathrm{X} \rightarrow {}^{A}_{Z+1}\mathrm{Y} + {}^{0}_{-1}e$$

  - **Beta-plus (β^+) decay**, where an **unstable proton <u>in the nucleus</u> *transforms into a neutron and a positron*** **and ejects the positron (Fig. 15b).**

    The decay process can be written as:

    $$^{A}_{Z}\mathrm{X} \rightarrow {}^{A}_{Z-1}\mathrm{Y} + {}^{0}_{1}e$$

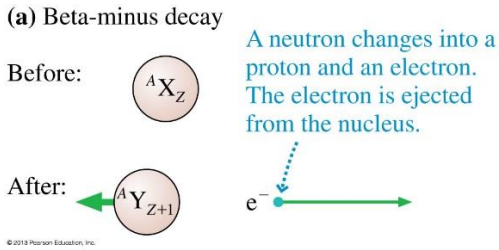


Figure 15(a)

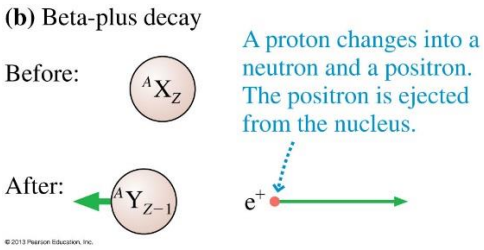


Figure 15(b)

- In both cases:
  1. As the proton number has changed, the parent nucleus **transmutes** into another nucleus.
  2. Although the respective numbers of protons and neutrons in the nucleus change during beta decay, the nucleon number remains the same.
  3. Although β^- decay results in an electron being ejected from the nucleus, the electron ***does not*** exist within the nucleus prior to the β-decay. It is created during the radioactive decay process. (This process is ***not*** ionisation, where an orbital electron is freed from an atom.)

[2] The positron is the antiparticle of the electron. For most kinds of particles, there is an associated antiparticle with the same mass and opposite charge. Thus, the positron has a mass m_e and a charge $+e$.

More on Beta Decay – the Existence of Neutrinos

Consider the beta-minus or beta-plus decay represented above (Fig. 15). If you apply the principles of *conservation of linear momentum* and *conservation of mass-energy* to each decay, you will see that for a radioactive isotope ${}^{A}_{Z}X$, the β radiation emitted should have only a single momentum and energy (Fig. 16).

However, experimentally, it has been found that beta particles from a single type of nucleus are emitted with a range of energies and momenta, apparently violating the principles of conservation of energy and momentum.

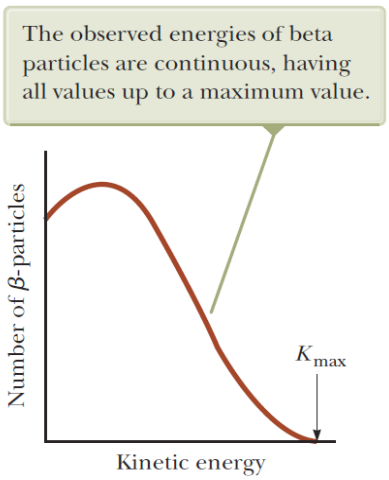


Figure 16

In 1930, Pauli proposed that a third particle must be present in the decay products. Enrico Fermi named this particle the *neutrino* (symbol "ν"), meaning the little neutral one.

- **The beta processes in their most correct and complete form** are therefore:
  - **Beta-minus (β^-) decay**: ${}^{A}_{Z}\mathrm{X} \rightarrow {}^{A}_{Z+1}\mathrm{Y} + {}^{0}_{-1}e + \bar{\nu}$ (complete expression)
  - **Beta-plus (β^+) decay:** ${}^{A}_{Z}\mathrm{X} \rightarrow {}^{A}_{Z-1}\mathrm{Y} + {}^{0}_{1}e + \nu$ (complete expression)

  where $\bar{\nu}$ represents an antineutrino (the antiparticle of the neutrino) and
  ν represents a neutrino.

- The energy emitted in each reaction is shared between the β-particle and the neutrino (or antineutrino in the case of beta-minus decay). This allows the β particles to be emitted from the nucleus with a range of energies and momenta, ensuring that the principles of conservation of mass-energy and of linear momentum are obeyed.

- **Properties of the Neutrino:**
  - From the conservation of charge, we can see from the equations above that the neutrino and antineutrino *neutrally charged*.
  - The neutrino also has *negligible mass*.
  - The neutrino *interacts* only *very weakly* with other matter and hence eluded detection for many years.

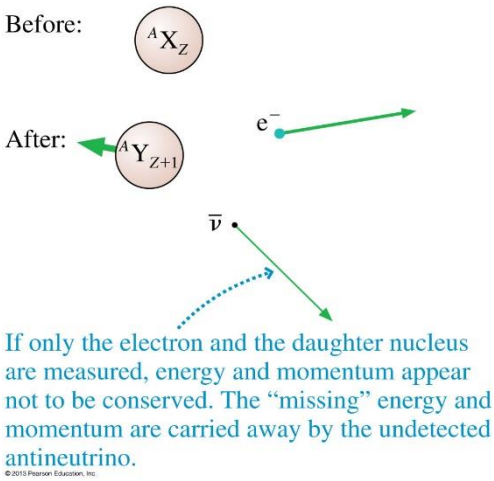


Figure 17

Gamma Decay

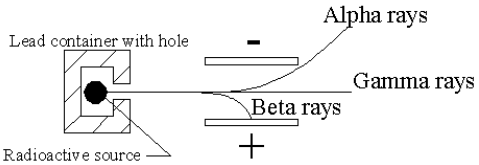
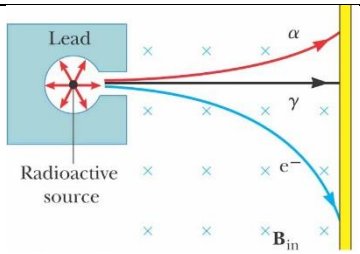
- Very often, after an alpha or beta decay, the nucleus is left in an *excited* state.

  A typical nucleus has a set of allowed energy levels, including a ground state and several excited states. A nucleus in an excited state is unstable and will undergo decay to a lower energy state by emitting one or more high-energy gamma photons:

  ${}^{A}_{Z}\mathrm{X}^{*} \rightarrow {}^{A}_{Z}\mathrm{X} + \gamma$ (The excited nucleus is denoted by an asterisk *.)

Properties of Radioactive Emissions

The table below shows the properties of the three types of radiations.

	α-particle	β-particle	γ-ray
Nature	$^{4}_{2}\mathrm{He}$ Helium nucleus	$^{0}_{-1}e$ Electron (β^- particle)	Photon
Charge	+ 2e	- e	0
Ionisation	Very strong (producing 10^3 to 10^4 ions per mm of path)	Strong	Weak
Penetrating power	Easily stopped by paper, skin, or a few cm of air	Stopped by ~1 m air or 3 mm of aluminium	Generally requires at least several cm of lead to be stopped
Deflection by an electric field	 ***Figure 18.***		
Deflection by a magnetic field	 ***Figure 19.***		

https://for.edu.sg/radiationpenetratingpower

18.6.2 Laws of Radioactive Decay

Mini Lecture 2.3 – Laws of Radioactive Decay: https://for.edu.sg/47s8yx.

Activity (*A*) of a Radioactive Decay

The *rate of decay* or ***activity*** of a radioactive sample is the number of disintegrations per unit time.

$$A = -\frac{dN}{dt}$$

where *N* is the number of radioactive nuclei present.

S.I. unit of activity: s^{-1} or Bq (becquerel)

Notes:

- $\frac{dN}{dt}$ is a negative number, since *N*, the number of undecayed nuclei, decreases with time. The activity is reported as a positive number, since it represents the number of nuclei that decay per unit time.
- 1 Bq = 1 disintegration per second.
- Another commonly used unit of activity is **curie** (Ci), defined as 1 Ci = 3.7 x 10^{10} Bq.

  The curie is a very large unit. Most of the activity of radioactive substances is measured in millicurie (mCi) or microcurie (μCi).

The Geiger–Müller Tube

(Refer to appendix A for more information about Geiger–Müller Tube)

- A Geiger-Müller tube, **when connected to a counter**, measures the **count rate *C*** instead of the activity. This is because the detector only captures *a portion* of the radioactive emissions from a sample of radioactive nuclides.
- **The count rate is a fraction of the activity, so $C \propto A$**

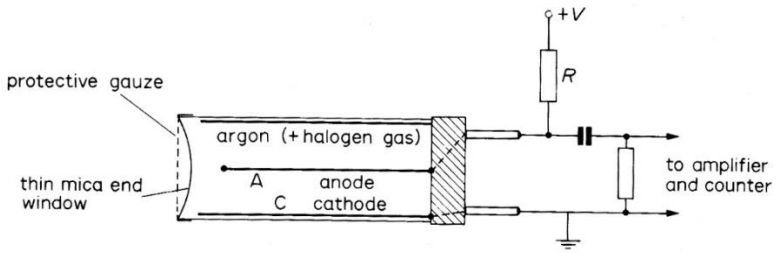


Figure 20. *Geiger-Müller Tube*

- When ionising radiation passes through the tube, each ionising particle produces a pulse voltage in the external circuit of a Geiger-Müller tube detector.
- **In order to measure the number of pulses from the detectors, a counter or a ratemeter must be connected to the Geiger-Müller tube.**
- A scaler counts the actual number of pulses while a ratemeter gives directly the average number of pulses per second or *count rate*.

Background Radiation

- In the absence of a specific radioactive source, a radiation detector still picks up a non-zero count rate (typically 20 to 50 counts per minute) from the environment. This measurement is known as the **background count**.
- Background count is due to **ionising radiation** that is constantly present in the environment. This radiation is known as **background radiation** and its sources can be natural or man-made.
- Naturally occurring sources account for about 80% of our exposure, and is contributed by things such as rocks, radon (a colourless and odourless gas that is easily trapped in buildings) and cosmic radiation. Man-made sources include those used in nuclear medicine, nuclear power plants and in consumer products (such as tobacco, which contains radioactive Pb-210). For more information on background radiation, refer to appendix B.
- **In experiments to study radioactivity, the background count is usually subtracted from the measurement, to eliminate the systematic error in the count rate.**

Decay Constant

Although radioactive decay is a random process, if the *number of radioactive nuclei is large*, we can associate a probability of decay to the nuclei.

- **The rate at which a particular decay process occurs in a sample is proportional to the number of radioactive nuclei present:**

$$\frac{dN}{dt} = -\lambda N \quad \text{..... (1)}$$

where λ is the proportionality constant known as the **decay constant**.

Note:

1. Since Activity $A = -\frac{dN}{dt}$, we can also write equation (1) as

$$\boxed{A = \lambda N} \quad \text{..... (2)}$$

2. Equation (1) can also be written as: $\lambda = \frac{(-\frac{dN}{dt})}{N}$

The ***decay constant*** is the probability per unit time that a nucleus will decay.

3. Rewriting equation (1), we can also have:

$$\frac{dN}{N} = -\lambda dt$$

Integrating, $\int_{N_o}^{N} \frac{1}{N'} dN' = -\lambda \int_0^t dt'$

$$\Rightarrow \quad \ln(\frac{N}{N_o}) = -\lambda t$$

$$\Rightarrow \quad \boxed{N = N_o e^{-\lambda t}} \quad \text{..... (3)}$$

where N is the number of radioactive nuclei present at time t and N_o represents the number of undecayed radioactive nuclei at t = 0.

Equation (3) shows that *the number of undecayed radioactive nuclei in a sample decreases exponentially with time* (Fig. 21).

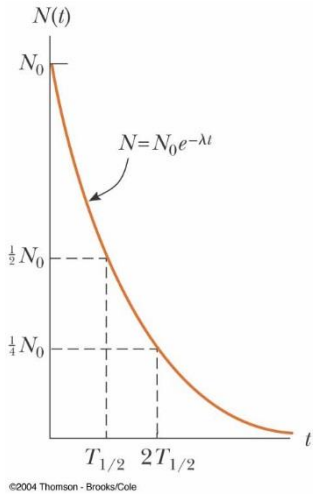


Figure 21. *Plot of the exponential decay law for radioactive nuclei. The curve is jagged (see Fig. 13). The smooth curve is fitted to represent the trend of the decay. The time $T_{1/2}$ is the half-life of the sample, the average time it takes for half of the nuclei to decay.*

4. **Variation of Activity with Time**

Recall that the activity *A* of a sample of radioactive nuclei is proportional to the number of radioactive nuclei present *N*, therefore

$A \propto N$ and $A_o \propto N_o$

Equation (3) can also be written as

$$A = A_0 e^{-\lambda t} \quad \text{..... (4)}$$

where *A* is the activity of the sample of radioactive nuclei at time *t* and A_o is the initial activity of the sample of radioactive nuclei at time $t = 0$.

5. **Variation of Count Rate with Time**

Furthermore, since the count rate *C* of a sample of radioactive nuclei is a proportion of the activity *A*, we also have:

$C \propto A$ and $C_o \propto A_o$

Therefore, equation (4) is equivalent to: $C = C_o e^{-\lambda t}$

Half-life $t_{1/2}$

Another parameter that is useful in characterising nuclear decay is the **half-life $t_{1/2}$**.

> The half-life of a radioactive nuclide is the average time taken for half of the original number of nuclei in a sample of the radioactive nuclide to decay.

or

> The half-life of a radioactive nuclide is the average time taken for the activity of a sample of the radioactive nuclide to halve.

The **above definition assumes a statistically large number of particles** and the half-life represents an expectation over a long time.

- As an example, consider a radioisotope with a $t_{1/2}$ = 7 years.

Time, t	***No of radioactive nuclei present, N***
0	$N_0 = \frac{N_o}{2^0}$
7 years (1 $t_{1/2}$)	$\frac{N_0}{2} = \frac{N_0}{2^1}$
14 years (2 $t_{1/2}$)	$\frac{N_0}{4} = \frac{N_o}{2^2}$
21 years (3 $t_{1/2}$)	$\frac{N_0}{8} = \frac{N_0}{2^3}$

https://for.edu.sg/halflife

Therefore, we see that after *n* half-lives, the time elapsed: $t = n \cdot t_{1/2}$,

and the number of radioactive nuclei present in the sample: $N = \frac{N_o}{2^n}$.

- **Relationship between decay constant and half-life**

  From the definition of half-life, we can also see that:

  At $t = t_{1/2} \Rightarrow N = \frac{N_o}{2}$ and $N = N_o e^{-\lambda t} \Rightarrow \frac{N_o}{2} = N_o e^{-\lambda t_{1/2}} \Rightarrow$ $\boxed{t_{1/2} = \frac{\ln 2}{\lambda}}$

Example 11: Radon Gas

Serway College Physics. Page 948 Example 29.4

Radon $^{222}_{86}\mathrm{Rn}$ is a radioactive gas that can be trapped in the basements of homes and its presence in high concentrations is a known health hazard. Radon has a half-life of 3.83 days. A gas sample initially contains 4.00×10^8 radon atoms.

a) How many atoms will remain after 14.0 days have passed if no more radon leaks in?

b) What is the activity of the radon sample after 14.0 days?

c) How long will it be before 99% of the sample has decayed?

18.6.3 Uses of Radioactive Nuclides

Mini Lecture 3.2 – Nuclear Boons and Banes: https://for.edu.sg/knylzi.

Radioactive nuclides are used in a myriad of applications from medicine to industry. Two interesting applications – one in medicine and the other in archaeology – are described below. Appendix C contains more applications of radioactive nuclides.

Medicine – Assessment of Thyroid Function

Iodine, which is absorbed into the blood stream, is collected by the thyroid gland for use in making thyroid hormones. Both overactive and underactive thyroids can lead to medical problems. To investigate the behaviour of the thyroid, tests are carried out to measure the accumulation of radioactive ^{131}I by the gland. The patient is given a tracer dose of a diluted sodium-iodide solution and the count rate from the thyroid is taken over a period of time and compared with that of a normal thyroid (Fig. 22).

More recently, ^{99}Tc^{m}, an isotope of technetium, has been used instead of iodine. The patient is given sodium pertechnetate, $NaTcO_4$, which is taken up in a similar way as iodine but is more easily released from the body. In addition, the half-life of ^{99}Tc^{m} is much shorter: six hours, compared to eight days for ^{131}I. The person thus remains radioactive for a much shorter period.

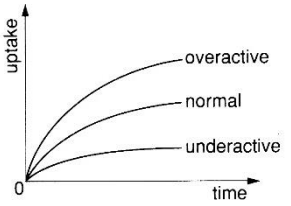


Figure 22 *Assessment of thyroid function*

Archaeology – Carbon Dating

Mini Lecture 3.1 – Carbon Dating: https://for.edu.sg/pzzjtr.

- Carbon-14 dating is a way of determining the age of certain archaeological artefacts of biological origin, up to about 50,000 years old.
- Carbon isotopes that occur naturally are

  $^{12}_{6}\mathrm{C}$: stable, 99% abundance

  $^{13}_{6}\mathrm{C}$: stable, 1% abundance

  $^{14}_{6}\mathrm{C}$: radioactive, $t_{1/2}$ = 5730 years, abundance of ~10^{-10} %

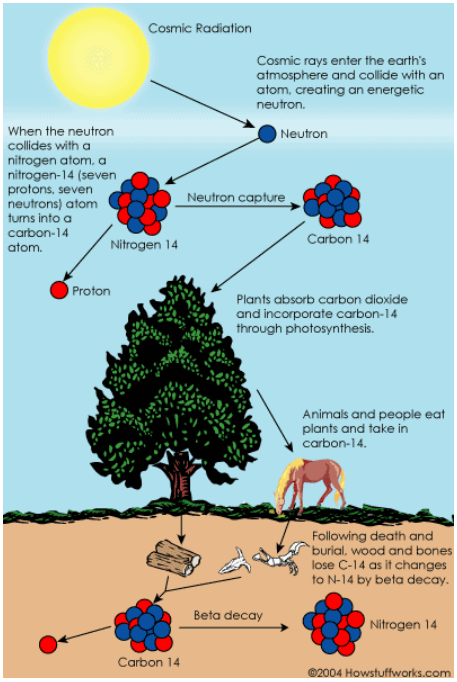


Figure 23. *Carbon-14 Dating.*

- When cosmic rays from outer space collide with molecules in the upper atmosphere, one of the nuclear reactions that occur is the conversion of nitrogen-14 to carbon-14 by neutron capture:

$$^{14}_{7}\mathrm{N} + {}^{1}_{0}\mathrm{n} \rightarrow {}^{14}_{6}\mathrm{C} + {}^{1}_{1}\mathrm{p}$$

  The carbon-14 atoms then undergo beta-decay:

$$^{14}_{6}\mathrm{C} \rightarrow {}^{14}_{7}\mathrm{N} + {}^{0}_{-1}e$$

- ^{14}C can occur in carbon dioxide CO_2. This radioactive CO_2 is absorbed by plants via photosynthesis. All living organisms that eat plant matter will also ingest this radioactive CO_2.
- In the atmosphere, *the rate of production of ^{14}C is approximately balanced by the rate of removal of ^{14}C and hence the ratio of ^{14}C to ^{12}C in the atmosphere is constant at approximately 1.3 x 10^{-12}.*

  The carbon atoms in all living organisms will correspondingly have this same $^{14}C/^{12}C$ ratio, since these organisms continuously exchange carbon dioxide with their surroundings.

https://for.edu.sg/carbondating

- When an organism dies, the carbon-14 atoms that decay are not replenished, while the amount of carbon-12 remains constant in the organism. Thus, the $^{14}C/^{12}C$ ratio begins to decline.
- It is therefore possible to determine the time elapsed since death by measuring the $^{14}C/^{12}C$ ratio of a sample.
- Using this technique, scientists have been able to identify samples of wood, charcoal, bone and shell as having lived from 1000 to 25000 years ago. This knowledge has helped us reconstruct the history of living organisms, including humans, during this time span.

Example 12: Carbon Dating
TYS N83/II/6 (part)

Living material contains a fairly constant small fraction of the radioactive isotope ^{14}C. After death, the activity from the ^{14}C present in the material decays with a half-life of 5.7 x 10^3 years. A certain quantity of carbon from recently dead organic matter is converted to carbon dioxide and placed in a Geiger-Müller tube, giving 12 counts per minute above the background. If the same quantity of carbon were taken from a linen shroud and similarly treated, what count rate would you expect if the shroud were

a) a relic dating from the early Christian era, about A.D. 33,
b) a medieval forgery dating from about A.D. 1100?

18.6.4 Radiation Hazard

Radioactive emissions can cause damage to the cells in our bodies. Most of this damage is due to **ionisation** when the radiation passes. If levels of radiation are high enough, there can also be damage due to heating effects as the body absorbs the energy from the radiation.

○ Ionising Radiation

Ionising radiation encompasses energetic subatomic particles, ions and atoms moving at high speeds and electromagnetic waves at the high-energy end of the electromagnetic spectrum. During an interaction with an atom, the ionising radiation may remove electrons from the atom, causing the atom to be electrically charged or ionised (Fig. 24).

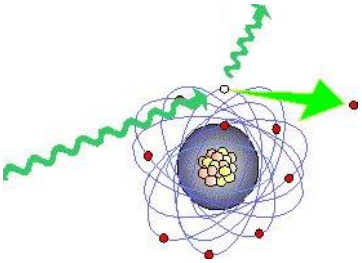


Figure 24. *Incoming radiation interacting with a neutral atom, causing ionisation*

○ Damage to Cells

The nucleus of a biological cell has been shown to be the part of the cell most vulnerable to radiation damage. This is because the nucleus contains the DNA molecules which carry the organism's genetic code.

○ Types of damage to cells

Damage may be **direct** or **indirect**.

1. Ionisation may cause **direct damage** to the DNA within the nucleus, when the whole DNA molecule may either be broken into fragments or have sections of it removed.

2. **Indirect damage** occurs when the radiation interacts with other molecules in the cell, producing ions and radicals (H^+, OH^-, H•, OH•), which can then "attack" other cell structures and DNA. The radicals can also combine to form toxic substances like H_2O_2 which could initiate harmful reactions within the cells.

○ Effects of Damage to DNA

- **Cell death**: If the damage within the cell is great, then **cell death** occurs immediately or within a very short time.
- **Cell is unable to replicate**: The DNA in the cell nucleus may have been damaged so badly that the **cell is unable to replicate itself by cell division**. This may have serious long-term consequences for the organism.
- **Cancer, genetic mutation and sterility:** At low doses, cellular reproduction may continue for many cell generations, giving rise to delayed effects such as **enhanced risk of developing cancer**. Rapidly multiplying cells are far more susceptible to damage, so special attention is paid in the use of ionising radiation on unborn children and young persons. The reproductive system may also be affected, giving rise to an **increased risk of genetic mutations**. If the damage is severe enough, it can cause **sterility**.

https://for.edu.sg/radiationhazard

Appendix A: Detecting Radiation

Nuclear radiation may come from the following:

1. Naturally occurring radioactive materials, such as compounds of uranium, thorium and potassium.
2. Purified or man-made sources, which are stronger than the natural sources. The ones you find in the school laboratory are sealed and are called *closed* sources. They are stored in lead-lined boxes, as most radiation is reduced considerably by a layer of lead.
3. Space: High-energy particles, such as muons and pions, reach the earth's surface from space. Such exotic, high-energy particles are also produced artificially in particle accelerators. All these particles decay into other ionising particles very quickly. It is unlikely that you will be able to detect the original particles in the laboratory, but the effects of the daughter particles contribute to the background radiation.

Wilson Cloud Chamber

A cloud chamber (Fig. 25) is a device that makes the paths of ionising radiation entering it visible. Water is introduced into chamber Y. If sufficient time is given, the air in the chamber becomes saturated with water vapour.

A series of adiabatic expansions is carried out to cool the air in chamber Y. As the air cools, the water condenses on the dust particles in the air and will fall off as water droplets. After a series of controlled adiabatic expansions, the air can be made dust-free and supersaturated, i.e. the vapour pressure is greater than the saturated vapour pressure at the reduced temperature reached, but no water vapour condenses.

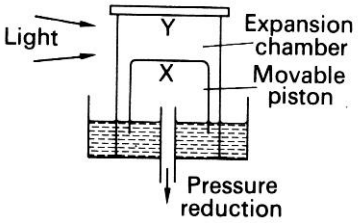


Figure 25.
Wilson Cloud Chamber

An ionising particle that enters chamber Y will collide with the air and water molecules to produce many ions. Each of these ions may act as a nucleating centre on which water can condense. Hence, a passing ionising particle will leave a track of water droplets, showing the route the particle took. Light that is scattered from the tracks makes them visible.

Glaser's Bubble Chamber

In the same way that air can be super-saturated with water vapour, a liquid under pressure can be heated to a temperature higher than that at which boiling normally takes place; it is then said to be superheated. If the pressure is suddenly reduced, bubbles may not form in the liquid for a certain time interval. During this quiet period, if the liquid is exposed to ionising radiation, nucleating centres are obtained for bubble formation and the liquid quickly evaporates into bubbles which grow rapidly. Hence, the bubble track reveals the path of the ionising radiation.

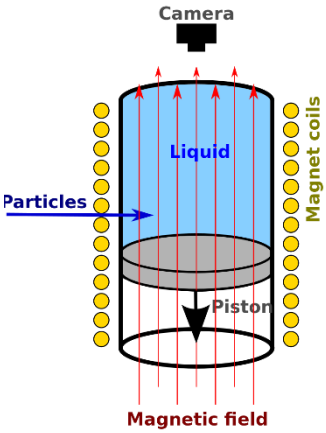


Figure 26.
A bubble chamber

Fig. 26 shows a schematic diagram of a bubble chamber. The magnetic field acts on charged particles and causes them to travel in helical paths, whose radii are determined by the particles' charge-to-mass ratios and their velocities. Because the magnitude of the charge of all known, charged, long-lived subatomic particles is the same as that of an electron, their radius of curvature must be proportional to their momentum. Thus, by measuring a particle's radius of curvature, its momentum can be determined.

The Geiger–Müller Tube

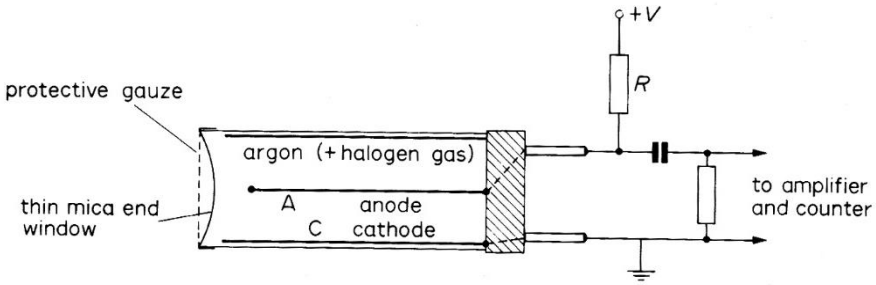


Figure 27. *Geiger-Müller Tube*

A Geiger-Müller (GM) tube consists of a tube filled with a low-pressure inert gas, such as helium, neon or argon, and an organic vapour or a halogen gas (see Fig. 27). The tube contains electrodes, between which there is a potential difference of several hundred volts, but no current flowing. The walls of the tube are either entirely metal or have their inside surface coated with a conductor to form the cathode, while the anode is a wire passing up the centre of the tube.

When ionising radiation passes through the tube, some of the gas molecules are ionised, creating positively charged ions and electrons. The strong electric field created by the tube's electrodes accelerates the ions towards the cathode and the electrons towards the anode. The ion pairs gain sufficient energy to ionise more gas molecules through collisions on the way, creating an avalanche of charged particles. This results in a short, intense pulse of current which passes from the negative electrode to the positive electrode and is measured or counted.

The usual form of a GM tube is an *end-window tube*. This type is so named, because the tube has a window at one end through which ionising radiation can easily penetrate. The other end normally has the electrical connectors. There are two types of end-window tubes: the *glass-mantle* type and the *mica-window* type. The glass-window type will not detect alpha radiation, since alpha-particles are unable to penetrate the glass, but is usually cheaper and will usually detect beta radiation and X-rays. The mica-window type will detect alpha radiation but is more fragile.

The ideal GM tube should produce a single current pulse on entry of every ionising particle. It must not give any spurious pulses and must recover quickly to the passive state. Unfortunately for these requirements, when positive argon ions reach the cathode and become neutral argon atoms again by obtaining electrons from it, the atoms can acquire their electrons in enhanced energy levels. These atoms then return to their ground state by emitting photons, which can in turn produce further ionisations and cause spurious secondary pulse discharges. Some form of quenching of the ionisation is therefore essential. The disadvantage of quenching is that for a short time after a discharge pulse has occurred (the so-called *dead time*, which is typically a few microseconds), the tube is rendered insensitive and is temporarily unable to detect the arrival of any new ionising particles. This effectively causes a loss of counts at sufficiently high count rates.

External quenching uses control electronics to temporarily remove the high voltage between the electrodes. *Self-quenching* or *internal-quenching* tubes stop the discharge without external assistance, by means of the addition of a small amount of an organic vapour or a halogen such as bromine or chlorine. In the latter case, the choice of quencher molecules is important. If a poor gas quencher is introduced in the tube, the positive argon ions have multiple collisions with the quencher gas and transfer their charge and some energy to them. Neutral argon atoms are then produced and the quencher gas ions reach the cathode, get excited and decay by photon emission, thereby producing spurious tube discharges as before. However, effective quencher molecules, when excited do not lose their energy by photon emission, but by dissociation into neutral quencher atoms. No spurious output pulses are then produced.

- ***Counters – Scalers and Ratemeters***

Each ionising particle or radiation produces a pulse voltage in the external circuit of a GM detector. In order to measure the number of pulses from the detectors, some form of counter must be used (Fig. 27). A scaler counts the actual number of pulses while a ratemeter gives directly the average number of pulses per second or ***count rate***.

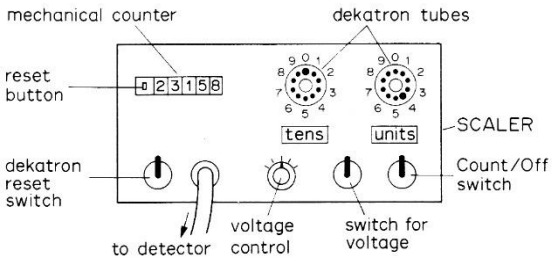


Figure 27. *(a) A scaler*

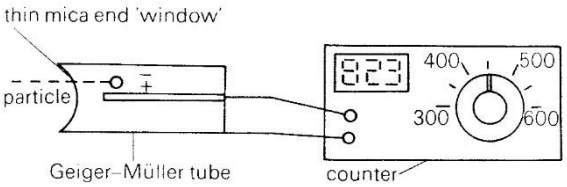


Figure 27. *(b) Connection of a scaler to a G-M tube for detection of nuclear particles.*

Scintillator

Gas-filled detectors like the GM tube are inefficient as gamma detectors as the probability of an interaction is low. A scintillator is a “solid" detector which addresses this shortfall. The incoming ionising radiation interacts with the scintillator to produce many electronic transitions. The decay of these transitions produces many photons of low energy, i.e., the scintillator material *fluoresces*.

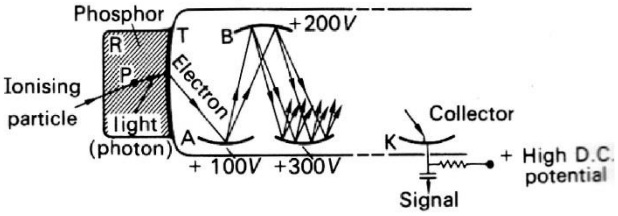


Figure 28. *Scintillator with photomultiplier*

Each emitted photon travels through the transparent scintillator material and ejects an electron from a photosensitive surface. Each of these electrons goes on to emit more secondary electrons: the numbers of electrons are multiplied by a photomultiplier and turn into a voltage pulse. The pulse amplitude is a measure of the energy of the incident ionising radiation.

Appendix B: Background Radiation

Background radiation is the ionising radiation that is constantly present in the environment. Sources of such radiation can be natural or man-made (Fig. 29).

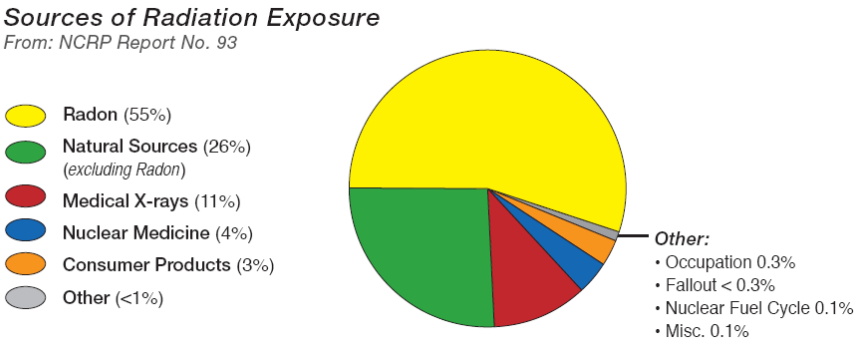


Figure 29. *Sources of radiation exposure*

Natural Sources of Radiation

Naturally occurring radiation accounts for about 80% of our exposure. Most of the exposure is from indoor radon, followed by radiation from outer space and from the Earth's crust.

- **Radon**
  - Approximately 55% of our exposure to radiation comes from radon, which is a *colourless*, *odourless* and *relatively inert*, naturally occurring radioactive gas. Radon-222 is one of the elements in the long radioactive decay chain from uranium-238 (Fig. 30) and the less common isotope radon-220 is part of the decay series from thorium-232 (Fig. 31), both of which can be found in rocks and soil.
  - **Dangers of Radon Gas**

    Radon is the second leading cause of lung cancer in the United States, with only smoking causing more lung-cancer deaths. Key dangers associated with radon gas are not so much the radon itself, but its radioactive daughters, which are solid particles that easily attach themselves to dust particles, which may be inhaled. When inhaled, these particles stick to the surfaces of the bronchial tubes and the *α*-particle emitted in the decay of the radon daughters damage the cells they are attached to, causing cancer.
  - **How Does Radon Get Into Homes?**

    Radon usually migrates into homes and other buildings from the ground through openings in the floor or wall. Buildings trap radon inside, where it accumulates and may become a health hazard. Any home or building may have a radon problem, be it a new or an old home, a well-sealed or a draughty one, as well as homes with and without basements.

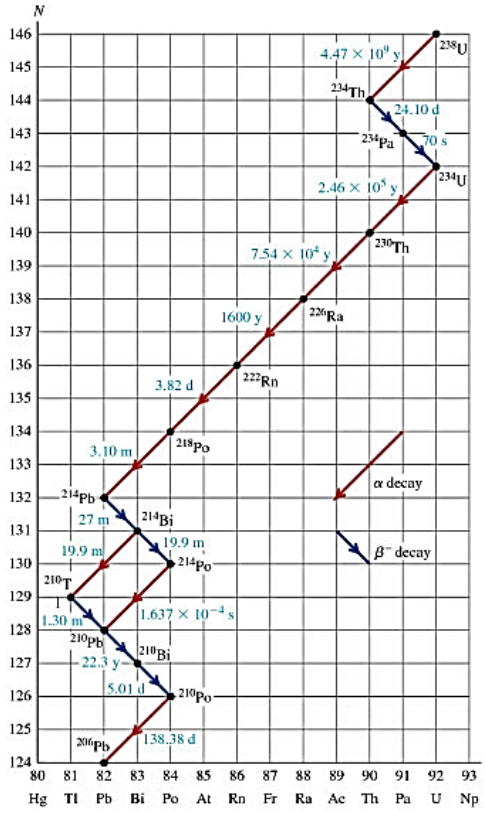


Figure 30. *The ^{238}U decay series, terminating with stable nuclide ^{206}Pb*

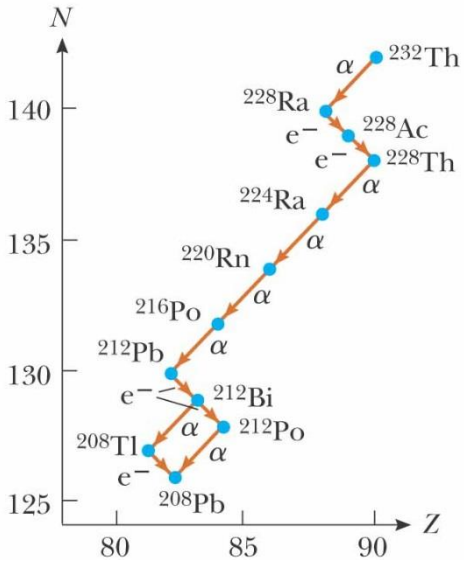


Figure 31. *The ^{232}Th series*

- **Soil and Buildings**
  - Natural radioactive material in rocks and soil account for about another 8% of the radiation dose a person typically receives in a year. The earth's crust contains small amounts of uranium, thorium and radium, as well as radioactive isotopes of several elements, including potassium. The radiation dose also comes from some building materials (such as bricks and concrete).

- **Cosmic Radiation**

  Yet another 8% of our exposure comes from cosmic radiation

  - Cosmic radiation mainly originates from the sun and outer space. Radioactive nuclides (mainly carbon-14) are produced through interactions of the cosmic rays with atoms in the atmosphere.
  - Cosmic rays are extremely energetic particles, primarily protons, which originate in the sun, other stars and from violent cataclysms in the far reaches of space. Cosmic-ray particles interact with the upper atmosphere of the earth and produce showers of lower-energy particles. Many of these lower-energy particles are absorbed by the earth's atmosphere, but a considerable fraction reaches the surface of the earth.
  - The cosmic-radiation exposure varies with altitude. The exposure of an individual to cosmic rays is greater at higher elevations than at sea level. For example, a passenger in a jetliner travelling at 37,000 feet would receive about 60 times as much cosmic radiation as a person standing at sea level for the same length of time.

- **Natural Radioactivity in the Body**
  - Small traces of many naturally occurring radioactive materials are present in the human body. These come mainly from naturally occurring radioactive nuclides present in the food we eat and in the air we breathe. These isotopes include tritium (^{3}H), carbon-14, and potassium-40.

Man-Made Sources

Since the discovery of radiation, people have benefited from the use of radiation in medicine and industry. Man-made sources of radiation account for about 20% of our total exposure to radiation.

- **Medicine**
  - The radiation used in medicine is the largest source of man-made radiation to which people in most countries are exposed. Most of the exposure is from diagnostic X-rays. Physicians use X-rays in more than half of all medical diagnoses to determine the extent of a disease or physical injury. Radiation is also used in cancer treatment, where precisely targeted radiation destroys diseased cells without killing nearby healthy cells. Furthermore, radiopharmaceuticals are used to locate tumours in a patient's body and to treat cancer. One-third of all successful cancer treatments involve radiation.

- **Consumer Products**
  - Small amounts of man-made background radiation can be attributed to consumer products like televisions, smoke detectors, gas-lantern mantles, natural-gas heating and mining and agriculture products.

- **Nuclear-Power Plants**
  - Nuclear-power reactors, which use radioactive uranium, reduce the reliance on fossil fuels. Nuclear-power plant operations account for less than a hundredth of a percent of the average American's total radiation exposure.

- **Nuclear Fallouts**
  - Nuclear weapons derive their explosive power from the uncontrolled radioactive breakup of plutonium and uranium. This yields many radioactive daughter products that are blown high into the atmosphere and are carried around the earth. These radioactive elements gradually fall back to earth (fallout) over a period of many years.
  - During the 1950s and early 1960s, many test explosions were carried out in the atmosphere. In 1963, an Atmospheric Test Ban Treaty was signed, and most subsequent tests have been conducted underground.
  - The peak year for radioactive fallout was 1963 and levels have been declining since then. Levels of fallout are now less than ten per cent of what they were at their peak.

- **Tobacco**
  - The tobacco in cigarettes contains Pb-210. Pb-210 is a naturally occurring radionuclide that precipitates out of the atmosphere and deposits on the leaves of tobacco.
  - When the tobacco is inhaled, the smoker receives a dose from the inhaled Pb-210 as well as Po-210, the decay product of Pb-210. Pb-210 is deposited on the surfaces of bones and Po-210 is deposited in the liver, kidney, and spleen.

Appendix C: Uses of Radioactive Nuclides

Radioactive nuclides are used in some industries and in nuclear medicine for their chemical properties and as sources of radiation. Some examples are listed below.

https://for.edu.sg/nuclearmedicine

- **In Medicine:**
  - As a tracer. Any radioactive nuclide which is to be used as a tracer for diagnostic purposes must be incorporated into a pharmaceutical product which can be administered to the patient.

    The pharmaceutical product and the radioactive nuclide which it incorporates should have the following properties:
    - They should concentrate in the organ or system which is to be imaged or assessed.
    - They should not change the functioning of the organ, i.e., the pharmaceutical behaves in the same way as the substance it is replacing.
    - The half-life of the radioactive nuclide should be as short as possible to reduce radiation damage, but sufficiently long to enable measurements to be made.

    Some examples of the use of radioactive tracers in medicine are listed below:
    - **Thrombosis (blood clots).** A radioactive tracer, iodinated fibrinogen, is introduced into the blood. If there is a blood clot forming in some part of the body, a concentration of the radioactive fibrinogen will occur at that point. The concentration can be located using a radiation detector.
    - **Haemorrhage inside the body.** Like thrombosis, the radioactive tracers are used to locate the precise site of haemorrhage in an emergency. Radioactive chromium is injected into the body and is taken up by the red blood cells and carried uniformly throughout the body. However, there will be more blood accumulating at the haemorrhage site and the greater activity there can be located using a radiation detector.
  - Cancer treatment. Immature cells and cells that are growing or dividing most rapidly are most sensitive to radiation. This is used in radiation treatment of cancer.

    Often, cancer cells are growing rapidly. Hence, they are much more likely to be killed by a high dose of gamma radiation from a cobalt-60 source than normal cells that divide less frequently.

- **In Industry:**
  - As a tracer. Tracers are used to detect leaks in underground pipes by doping the water in the pipes and using a detector to check where the water flows. Tracers are also used to check the wear of a machine part. For example, we can check the wear of a piston if it has been doped with a radioactive material during manufacture by checking the activity in the oil that is used with the piston.
  - To check for cracks and imperfections
    - Gamma rays can be used in place of X-rays to photograph solid objects. Gamma rays are directed at the object and the image is captured on a photographic film placed at the other side of the object. The film can reveal any imperfections such as cracks. The technique has an advantage over X-rays as no high-voltage electrical sources are needed.
    - Radiation can also be used to check and control the thickness of paper or plastic during the manufacturing process. The thickness of rolled sheets of plastic can be controlled by placing a beta source on top of the sheet and a detector beneath it. If the sheet is too thin, the count rate will rise. This rise in count rate can then be directly coupled to the rollers, so that their separation is increased.

- **In Agriculture:**
  - Suppose a plant research centre wishes to investigate the ability of a particular plant to take up magnesium through its roots. This is possible by watering the plant with a fertiliser containing a known, extremely low concentration of radioactive magnesium-28. If a detector is placed in contact with the leaves of the plant, an increased count rate soon after watering in the fertiliser indicates that the magnesium has travelled through the plant to the leaves. This can happen within a time as short as 20 minutes.

- **As a sterilisation tool:**
  - Radiation can also be used as a means of preserving food, because exposure to high levels of radiation can destroy or incapacitate bacteria and mould spores. Techniques include exposing foods to gamma rays, high-energy electron beams and X-rays. Food preserved this way can be placed in a sealed container (to keep out new spoiling agents) and stored for long periods of time.

Appendix D: Calculating Radiation Doses

- **Radiation dosimetry** is the quantitative description of the effect of radiation on living tissue.
- **Absorbed dose:**
  - The **absorbed dose** of radiation is defined as the energy delivered to tissue per unit mass.
  - The SI unit of absorbed dose is the *gray* (Gy):
    - 1 Gy = 1 J/kg
  - Another commonly used unit is the *rad*.
    - 1 rad = 0.01 J/kg = 0.01 Gy
- **Relative Biological Effectiveness (RBE):**
  - The absorbed dose by itself is not an adequate measure of biological effect, because equal energies of different radiation may cause different extents of biological effect. This variation is described by a numerical factor called the **Relative Biological Effectiveness (RBE)** or the **quality factor (QF)** of each specific radiation.
  - RBE is defined as the ratio of a dose of standard radiation to the dose of test radiation to produce the same biological effects. As such, RBE is a dimensionless quantity.
  - The table below shows some approximate values of RBE for several types of radiation. These values are obtained from experimental observation. All these values depend on the kind of tissue in which the radiation is absorbed and on the energy of the radiation.

Radiation	RBE
X-rays, γ-rays, β-particles, muons	1
Protons, charged pions	2
Slow neutrons	3 - 5
Fast neutrons	10
α-particles, nuclear fission products, heavy nuclei	20

- **Biologically Equivalent Dose:**
  - The biological effect is described by the quantity **biological equivalent dose**, or simply the **equivalent dose**. The equivalent dose H is defined as the product of the absorbed dose D of ionising radiation and a dimensionless factor Q (the quality factor):

$$H = D \times Q$$

  - The SI unit of equivalent dose for humans is the **sievert (Sv)**:

$$1\ \text{Sv} = \text{RBE} \times 1\ \text{Gy}$$

  - An alternative unit of equivalent dose for humans is the **roentgen equivalent man (rem)**:

$$1\ \text{rem} = 0.01\ \text{Sv}$$

  - In summary: the **gray** is the unit for D, the **absorbed dose**. 1 Gy = 1 joule/kilogram, a physical quantity. 1 Gy is the deposit of one joule of radiation energy per kilogram of matter or tissue. The **sievert** is the unit for H, the **equivalent dose**. 1 Sv = 1 joule/kilogram, a biological effect. 1 sievert represents the equivalent biological effect of the deposit of one joule of radiation energy in a kilogram of human tissue. The ratio of equivalent dose to absorbed dose is denoted by Q, where $Q = H / D$.

Tutorial 18B: Radioactive Decay

Self-Review Questions

S1 J03/II/16

(a) A student provided with a freshly prepared sample of a radioactive material and the count rate C is found to vary with time t as shown in Fig. 6.1 (a) below.

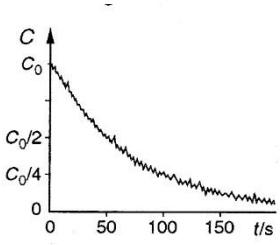


Fig. 6.1 *(a)*

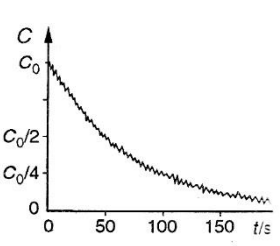


Fig. 6.1 *(b)*

A second similar sample of the radioactive material is then prepared and the student repeats the experiment, but with the sample at a higher temperature. The variation with time of the count rate of the second sample is shown in Fig. 6.1 (b).

State the evidence that is provided by these two experiments for

(i) the random nature of radioactive decay,

(ii) the spontaneous nature of radioactive decay. [2]

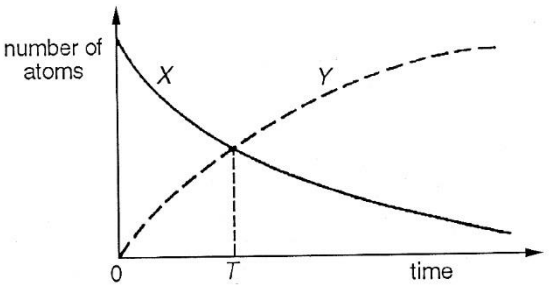
(b) The radioactive source in **(a)** is an isotope of radon ($^{220}_{86}\mathrm{Rn}$) that emits an α-particle to become polonium (Po).

(i) State the number of neutrons in one nucleus of radon-220.

(ii) Write down a nuclear equation to represent the radioactive decay of a nucleus of radon. [4]

S2 N90/I/30.

The graph on the right represents the decay of a newly prepared sample of radioactive nuclide X to a stable nuclide Y. The half-life of X is τ. The growth curve for Y intersects the decay curve for X after time T.



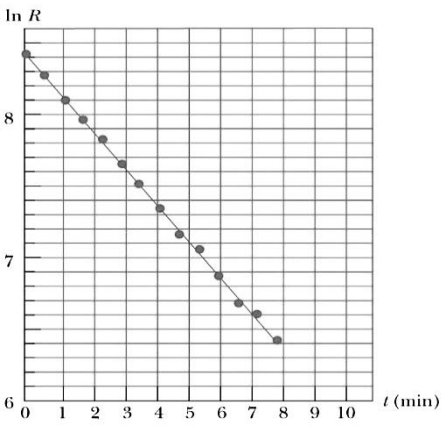
What is the time T?

A $\tau/2$ **B** $\ln(\tau/2)$ **C** τ **D** $\ln(2\tau)$ **E** 2τ

S3 Serway PSE 6th Edition. P44.75. Page 1478.

Student determination of the half-life of ^{137}Ba. The radioactive barium isotope ^{137}Ba has a relatively short half-life and can be easily extracted from a solution containing its parent caesium (^{137}Cs). This barium isotope is commonly used in an undergraduate laboratory exercise for demonstrating the radioactive decay law. Undergraduate students, using modest experimental equipment, took the data presented in the figure on the right.

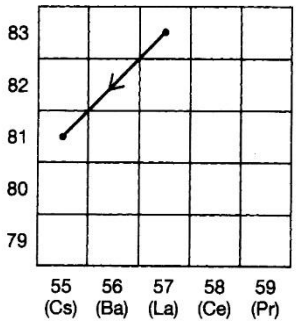
Determine the half-life for the decay of ^{137}Ba using their data.



S4 Akrill, Bennet and Millar. Page 132, Problem 13.16.

Copy the grid on the right, which gives the number of neutrons *N* in a nucleus (vertical axis) against the number of protons *Z* (horizontal axis) in the region around *N* = 81 and *Z* = 57.

Explain why the arrow shows an *α*-decay and write the nuclear equation for this decay.



S5 J85/II/11 (part)

A $^{238}_{92}\mathrm{U}$ nucleus, originally at rest, spontaneously decays to form a thorium (Th) nucleus and an alpha particle. A gamma ray is not emitted.

(a) Using conventional notation for mass/atomic numbers, write down the equation which describes this disintegration.

(b) The alpha particle produced in this disintegration travelled 25 mm in a cloud chamber. Given that, on average, an alpha particle creates 5.0 x 10^3 ion pairs per mm of track in the cloud chamber and that the energy require to produce an ion pair is 5.2 x 10^{-18} J, find the kinetic energy with which the alpha particle was emitted.

(c) Hence, deduce the initial velocities of the alpha particle and the thorium nucleus.

(d) Calculate the difference between the rest mass of the original uranium nucleus and the sum of the rest masses of the products of the disintegration.

Discussion Questions

D1 N91/I/30; J95/I/30

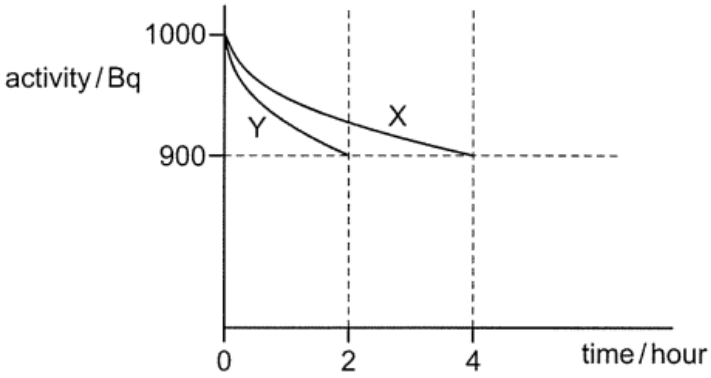
In a cancer therapy unit, patients are given treatment from a certain radioactive source. This source has a half-life of 4 years. A particular treatment requires 10 minutes of irradiation when the source is first used.

How much time is required for this treatment, using the same source, 2 years later?

A 7 minutes **B** 10 minutes **C** 14 minutes **D** 20 minutes

D2 N12/1/39

The graph shows how the activities of two isotopes X and Y vary with time.



What will be the ratio $\frac{\text{activity of X}}{\text{activity of Y}}$ after each has decayed for 8 hours?

A $\frac{9}{10}$ **B** $\frac{10}{9}$ **C** $\frac{81}{100}$ **D** $\frac{100}{81}$

D3 N11/I/40

At time t, a sample of a radioactive substance contains N atoms of a particular nuclide.

At time $(t + \Delta t)$, where Δt is a short period of time, the number of atoms of the nuclide is $(N - \Delta N)$.

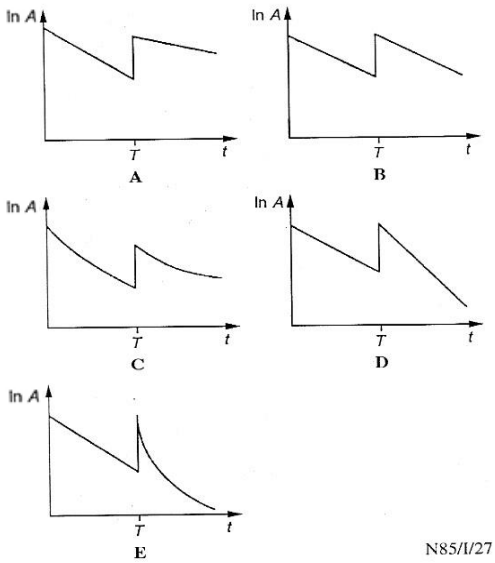
Which expression is equal to the decay constant of the nuclide?

A $\frac{\Delta N}{N}$ **B** $\frac{\Delta N}{\Delta t}$ **C** $\frac{\Delta N}{N\Delta t}$ **D** $\frac{N\Delta N}{\Delta t}$

D4 N90/I/30

At time $t = 0$, some radioactive gas is injected into a sealed vessel. At time T some more of the same gas is injected into the same vessel.

Which of the graphs on the right best represents the variation of the logarithm of the activity A of the gas with time?



N85/I/27

D5 N99/II/7. Potassium-Argon Dating

One isotope of potassium, $^{40}_{19}K$ has a half-life of 1.4 x 10^9 years and decays to form argon, $^{40}_{18}Ar$, which is stable. A sample of rock taken from the Sea of Tranquillity on the Moon contains both potassium and argon in the ratio

$$\frac{\text{number of Potassium-40 atoms}}{\text{number of Argon-40 atoms}} = \frac{1}{7}.$$

(a) Define *half-life*. [2]

(b) The decaying potassium nucleus emits a particle X.

(i) Write down the nuclear equation representing this decay.

(ii) Suggest the identity of X. [2]

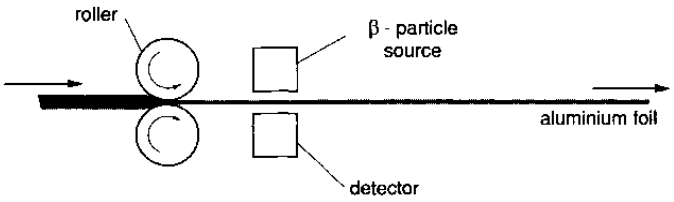
(c) Assume that when the rock was formed, there was no argon-40 present in the sample and none has escaped subsequently

(i) Estimate the age of the rock.

(ii) State, with a reason, whether your answer in **(i)** is an overestimate or an underestimate of the age of the rock if some escape of argon has occurred. [4]

D6 J97/II/8 (part), J92/III/6 (modified)

The thickness of a sheet of aluminium foil is to be monitored using β-radiation as illustrated.



The separation of the rollers is controlled by the output from the detector with the intention of maintaining constant foil thickness.

(i) State what adjustment should be made to the separation of the rollers if the output from the detector were to increase.

(ii) Suggest why a γ-radiation source would not be satisfactory for monitoring changes in thickness of the foil.

(iii) A β-particle source of half-life 14 days is installed in the monitor and then used for a working day of 8.0 hours.

1. Calculate the ratio

$$\frac{\text{activity of source at end of working day}}{\text{activity of source at start of working day}}$$

2. Estimate the percentage change in foil thickness during one working day if no allowance is made for radioactive decay. State whether the change is an increase or decrease. [7]

(iv) Discuss the advisability, or otherwise, of using a β-emitting source which has

1. a high activity, giving count rate many times that of background.

2. a short half-life. [6]

D7 N2000/II/8(part) Data Analysis

(a) Thorium-231 ($^{231}_{90}\mathrm{Th}$) decays with a half-life of 25 hours to form the daughter product protactinium-231 ($^{231}_{91}\mathrm{Pa}$), which has a half-life of 3.4 x 10^4 years.

(i) State the change in the composition of a nucleus of thorium-231 when it decays to become protactinium-231.

(ii) A fresh sample of radioactive material contains N_o nuclei of thorium-231 and no protactinium-231 at time t = 0. Assume that protactinium-231 is stable. On Fig. 8.1, sketch graphs to show the variation with time t of the number N of nuclei in the sample which are

1. thorium nuclei (label this line T),
2. protactinium nuclei (label this line P).

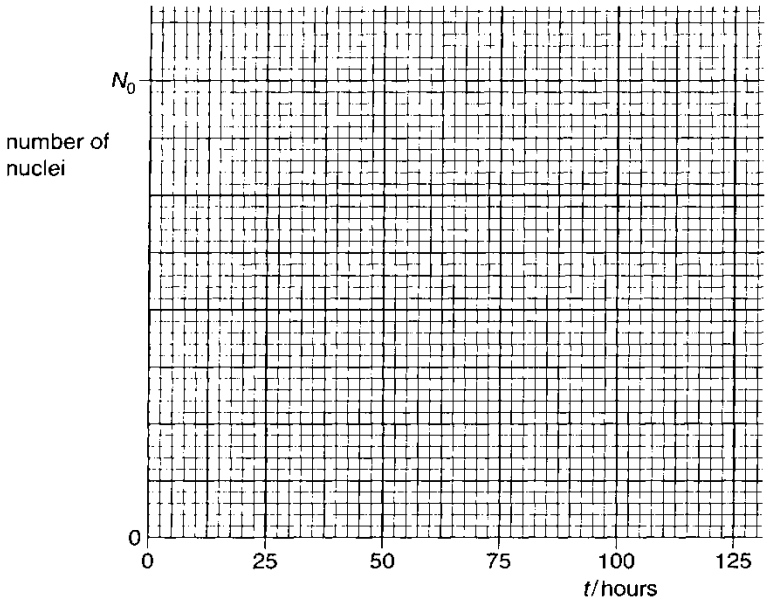


Fig. 8.1

(iii) Using Fig. 8.1 or otherwise, deduce the age of the sample of radioactive material when

$$\frac{\text{number of Thorium-231 in sample}}{\text{number of Proactinium-231 nuclei in sample}} = \frac{1}{3}$$

[6]

(b) For many unstable parent nuclei, the daughter product is itself radioactive. This gives rise to a radioactive series, where there may be ten or more different radioactive daughter products. The variation with time t of the percentage P of different nuclei in a radioactive sample is illustrated in Fig. 8.2.

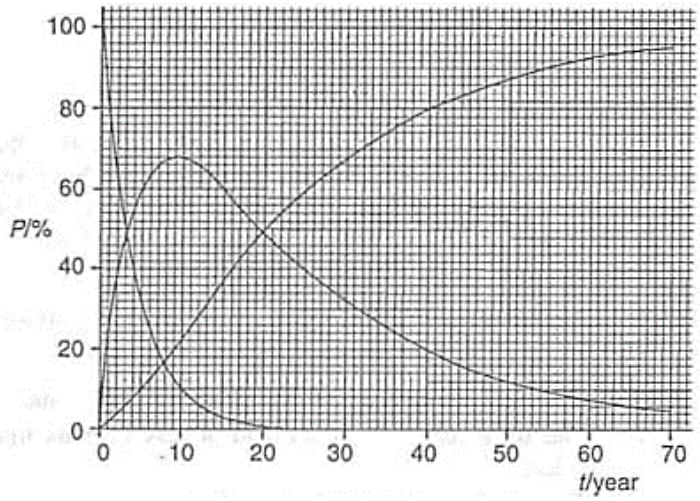


Fig 8.2

The parent nucleus has a daughter nucleus and, in turn, this daughter produces a further daughter.

Initially, there are 1.2×10^{15} nuclei of the parent isotope in the sample and the daughter products are not present. The parent isotope has a half-life of 3.0 years, and the daughter product D has a half-life of 15 years, decaying to form the further daughter S, which is stable.

(i) On Fig. 8.2, label with the letter S the line representing the variation of the number of nuclei of the stable further daughter S. Give an explanation for your choice. [1]

(ii) 1. State the time after which the daughter product D has its maximum activity.

2. Calculate the maximum activity, in microcuries (μCi), of the daughter product D, given that 1.0 μCi = 3.7×10^4 Bq. [4]

(iii) Suggest why the number of nuclei of the daughter product D increases to a maximum and then decreases. [4]

(iv) The relative activities of the parent isotope and the daughter products may be used as a means of determining the age of the sample. Suggest why this technique may provide reliable results for ages up to about 30 years, but for an age of about 100 year the method will be far less reliable. [2]

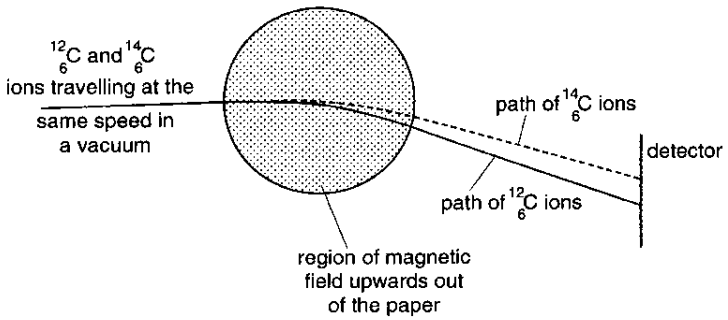
D8 N02/III/5 (parts) Carbon Dating

(a) Explain why the random nature of radioactive decay makes it difficult to measure the values of the *half-life* to a high degree of accuracy. [2]

(b) Measurements are made of the activity of a specimen of carbon from pieces of wood found in a fireplace at an archaeological site. The specimen is found to contain one carbon-14 atom per 8.6×10^{10} carbon-12 atoms. In similar carbon from a modern fire, the concentration of carbon-14 atoms is greater at one carbon-14 atom per 3.3×10^{10} carbon-12 atoms. The difference between these two figures is due to the fact that carbon-14 is radioactive and some atoms have decayed over the years.

Calculate the age of the wood from the ancient fire. The half-life of carbon-14 is 1.8×10^{11} s. Carbon-12 is stable. [6]

(c) The technique of dating described in (b) is difficult to do accurately. This difficulty can be minimised by using all the carbon-14 atoms rather than just those which happen to undergo radioactive decay when the dating is being carried out. Carbon atoms from the wood can be ionised by removing one electron from each atom. They are then formed into a beam which is passed through a magnetic field as shown.



(i) Explain why the paths of the two ion types are different.

(ii) Explain why the three sections of each path followed by the ions are straight, circular, straight and why there is no change in the speed of an ion throughout.

(iii) Suggest why this method of measuring the ratio of carbon-14 to carbon-12 atoms is more reliable. [6]

D9 N03/II/8. Data Analysis (continued from Tutorial 18A D4)

(d) The fission products are usually radioactive and give rise to a series of radioactive decay products. Each decay product has its own half-life, but eventually a stable nuclide is reached.

Two such fission products with their decay products and half-lives are shown below.

$$^{99}_{42}\text{Mo} \xrightarrow{67\text{ hour}} {}^{99}_{43}\text{Tc} \xrightarrow{2\times10^5\text{ year}} {}^{99}_{44}\text{Ru (stable solid)}$$

$$^{140}_{54}\text{Xe} \xrightarrow{16\text{ second}} {}^{140}_{55}\text{Cs} \xrightarrow{1.1\text{ minute}} {}^{140}_{56}\text{Ba} \xrightarrow{13\text{ day}} {}^{140}_{57}\text{La} \xrightarrow{40\text{ hour}} {}^{140}_{58}\text{Ce (stable solid)}$$

Consider equal amounts of these two fission products.

Suggest why there are very different problems for the storage for this nuclear waste. [3]

D10 Radioactivity (N2011/2/5 partial)

Plutonium-239 decays by α-emission to form the isotope uranium-X. The half-life of plutonium-239 is 2.4 x 10^4 years. The half-life of uranium-X is 7.1 x 10^8 years.

(a) (i) Explain the term isotope.

(ii) Define half-life.

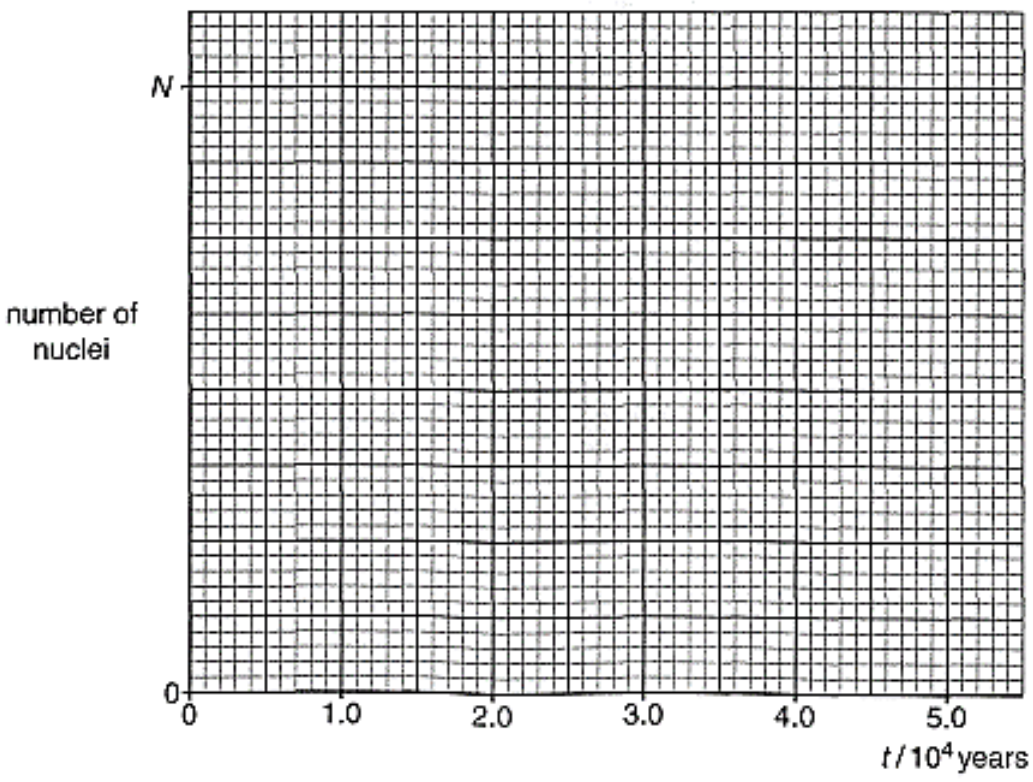
(iii) Complete the nuclear equation for the nuclear reaction $^{239}_{\cdots}\text{Pu} \rightarrow {}^{\cdots}_{92}\text{U} + {}^{\cdots}_{\cdots}\alpha$.

(b) A pure sample of plutonium-239 has N nuclei at time t = 0. On the figure below, sketch graphs to show the variation with t of the number of nuclei of

(i) plutonium-239 (label this graph P),

(ii) uranium-X (label this graph U).

Your graphs should cover the range from t = 0 to t = 5.0 x 10^4 years.



(c) In the nuclear reaction in (a), 5.26 MeV of energy is released. The energy of the emitted α-particle is 5.15 MeV. Suggest and explain why these two values are different.

Challenging Question

C1 N88/II/12 (part)

Radioactive iron, ^{59}Fe, is a radioactive nuclide with a half-life of 46 days. It is used medically in the diagnosis of blood disorders. Measurements are complicated by the fact that iron is excreted, i.e., removed from the blood, at a rate such that 69 days after administering a dose, half of the iron atoms in the dose have been excreted. That is to say, iron has a biological half-life of 69 days.

(i) If the count rate from a blood sample is 690 counts per minute, what will it be from a similar blood sample taken 138 days later? [4]

(ii) How long after the first sample was taken would a similar sample give a count rate of 480 counts per minute? [4]

(iii) Derive the relationship between

T, the effective half-life of the radioactive nuclide in the body,

T_b the biological half-life, and

T_r, the radioactive half-life.

Selected Answers to Tutorial 18B

S1 **(b)** **(i)** 134
S2 C
S3 2.8 minutes
S5 **(b)** 6.5 x 10^{-13} J
(c) 1.4 x 10^{7} m s^{-1}, 2.4 x 10^{5} m s^{-1}
(d) 7.3 x 10^{-30} kg

D5 **(b)** $^{40}_{19}K \rightarrow {}^{40}_{18}Ar + {}^{0}_{1}e$
(c) 4.2 x 10^{9} years
D6 **(iii)** **1.** 0.984
2. ~ 2%
D7 **(a)** **(iii)** 50 hours
(b) **(ii)** **1.** 9.0 years
2. 32 μCi
D8 **(b)** 7900 years

C1 **(i)** 22 counts per minute
(ii) 14 days