FOR THE
IB DIPLOMA
PROGRAMME

IN COOPERATION WITH

THIRD EDITION

Physics

John Allum

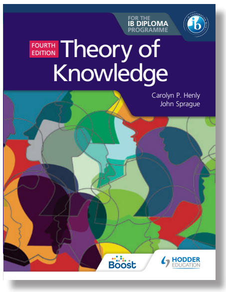
Paul Morris

Boost

HODDER
EDUCATION

Available for Diploma Core

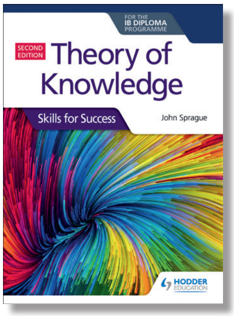
Theory of Knowledge for the IB Diploma Fourth Edition 9781510474314



Confidently navigate the Theory of Knowledge Guide with a set of rich and engaging resources, grounded in conceptual considerations and illustrated with real-world examples.

- Guide students by helping them examine the nature of knowledge and their own status as a knower.
- Develop diverse and balanced arguments with a variety of activities, case studies and Deeper Thinking features.
- Aid understanding with in-depth discussions of the twelve course concepts and detailed definitions of all key terms.
- Provide assessment support with guidance relating to the TOK Exhibition and Essay.

Theory of Knowledge for the IB Diploma: Skills for Success Second Edition 9781510474956

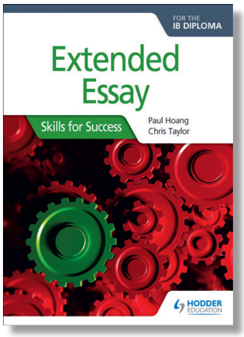


Build confidence in a range of key Theory of Knowledge skills with this practical companion, full of advice and guidance from an experienced TOK expert.

- Learn to apply analytical skills with Deeper Thinking, showing you how to go beyond simply identifying and explaining.
- Develop awareness of the practical application of knowledge with In Practice pointers, offering guidance on how topics can be used in TOK activities.
- Improve your ability to respond to knowledge questions, a crucial part of assessment success.
- Avoid making the mistakes that others make in the assessments with TOK Traps that highlight common errors and misconceptions.

www.hoddereducation.com/tok

Also available



Build confidence in a range of key essay writing techniques and skills with this practical companion, full of advice and guidance from experienced EE experts.

- Build essay writing techniques and skills through a range of strategies, serving as a useful companion throughout the writing process – from the development of a research question, critical-thinking, referencing and citation to reflecting on the process and final essay.
- Concise, clear explanations help you navigate the IB requirements, including advice on assessment objectives and academic honesty.
- Learn what is required to get the best EE grades and write an excellent essay with detailed examiner advice and expert tips and hints, including common mistakes to avoid.
- Explicit reference to the IB Learner profile and the importance of reflection.

www.hoddereducation.com/ibdiploma

FOR THE
IB DIPLOMA
PROGRAMME

THIRD EDITION

Physics

This page intentionally left blank

FOR THE
IB DIPLOMA
PROGRAMME

THIRD EDITION

Physics

John Allum
Paul Morris

Boost
HODDER EDUCATION
AN HACHETTE UK COMPANY

Publisher's acknowledgements

IB advisers: The Publishers would like to thank the following for their advice and support in the development of this project: Arno Dirks and Aurora Vicens. We would like to offer special thanks to Paul Morris for his invaluable review and feedback during the writing process.

The Publishers would also like to thank the International Baccalaureate Organization for permission to re-use their past examination questions in the online materials.

Every effort has been made to trace all copyright holders, but if any have been inadvertently overlooked, the Publishers will be pleased to make the necessary arrangements at the first opportunity.

Although every effort has been made to ensure that website addresses are correct at time of going to press, Hodder Education cannot be held responsible for the content of any website mentioned in this book. It is sometimes possible to find a relocated web page by typing in the address of the home page for a website in the URL window of your browser.

Hachette UK's policy is to use papers that are natural, renewable and recyclable products and made from wood grown in well-managed forests and other controlled sources. The logging and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

Orders: please contact Hachette UK Distribution, Hely Hutchinson Centre, Milton Road, Didcot, Oxfordshire, OX11 7HH. Telephone: +44 (0)1235 827827. Email education@hachette.co.uk Lines are open from 9 a.m. to 5 p.m., Monday to Friday. You can also order through our website: www.hoddereducation.com

ISBN: 978 1 3983 6991 7

© John Allum 2023
First published in 2012
Second edition published in 2014
This edition published in 2023 by
Hodder Education,
An Hachette UK Company
Carmelite House
50 Victoria Embankment
London EC4Y 0DZ

www.hoddereducation.com

Impression number 10 9 8 7 6 5 4 3 2 1

Year 2027 2026 2025 2024 2023

All rights reserved. Apart from any use permitted under UK copyright law, no part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or held within any information storage and retrieval system, without permission in writing from the publisher or under licence from the Copyright Licensing Agency Limited. Further details of such licences (for reprographic reproduction) may be obtained from the Copyright Licensing Agency Limited, www.cla.co.uk

Cover photo © salita2010/stock.adobe.com

Illustrations by Aptara Inc., Pantek Media and Barking Dog Art

Typeset in Times 10/14pt by DC Graphic Design Limited, Hextable, Kent

Printed in Bosnia & Herzegovina

A catalogue record for this title is available from the British Library.

FSC
www.fsc.org
MIX
Paper | Supporting responsible forestry
FSC™ C104740

Contents

A Space, time and motion

B The particulate nature of matter

C Wave behaviour

D Fields

E Nuclear and quantum

Free online content

Go to our website www.hoddereducation.com/ib-extras for free access to the following:

- Practice exam-style questions for each chapter
- Glossary
- Answers to self-assessment questions and practice exam-style questions
- Tools and Inquiries reference guide
- Internal Assessment – the scientific investigation

Introduction

Welcome to *Physics for the IB Diploma Third Edition*, updated and designed to meet the criteria of the new International Baccalaureate (IB) Diploma Programme Physics Guide. This coursebook provides complete coverage of the new IB Physics Diploma syllabus, with first teaching from 2023. Differentiated content for SL and HL students is clearly identified throughout.

The aim of this syllabus is to integrate concepts, topic content and the nature of science through inquiry. Approaches to learning in the study of physics are integrated with the topics, along with key scientific inquiry skills. This book comprises five main themes:

- **Theme A**: Space, time and motion
- **Theme B**: The particulate nature of matter
- **Theme C**: Wave behaviour
- **Theme D**: Fields
- **Theme E**: Nuclear and quantum physics

Each theme is divided into syllabus topics.

The book has been written with a sympathetic understanding that English is not the first language of many students.

No prior knowledge of physics by students has been assumed, although many will have taken an earlier course (and they will find some useful reminders in the content).

In keeping with the IB philosophy, a wide variety of approaches to teaching and learning has been included in the book (not just the core physics syllabus). The intention is to stimulate interest and motivate beyond the confines of the basic physics content. However, it is very important students know what is the essential knowledge they have to take into the examination room. This is provided by the Key information boxes. If this information is well understood, and plenty of self-assessment questions have been done (and answers checked), then a student will be well-prepared for their IB Physics examination.

The online Glossary is another useful resource. It's aim is to list and explain basic terminology used in physics, but it is not intended as a list of essential information for students. Many of the terms in the Glossary are highlighted in the book as 'Key terms' and also emphasized in the nearby margins.

The ‘In cooperation with IB’ logo signifies that this coursebook has been rigorously reviewed by the IB to ensure it fully aligns with the current IB curriculum and offers high-quality guidance and support for IB teaching and learning.

How to use this book

The following features will help you consolidate and develop your understanding of physics, through concept-based learning.

Guiding questions

- There are guiding questions at the start of every chapter, as signposts for inquiry.
- These questions will help you to view the content of the syllabus through the conceptual lenses of the themes.

SYLLABUS CONTENT

- This coursebook follows the order of the contents of the IB Physics Diploma syllabus.
- Syllabus understandings are introduced naturally throughout each topic.

Key information

Throughout the book, you will find some content in pink boxes like this one. These highlight the essential Physics knowledge you will need to know when you come to the examination. Included in these boxes are the key equations and constants that are also listed in the IBDP Physics data booklet for the course.

Tools

The Tools features explore the skills and techniques that you require and are integrated into the physics content to be practiced in context. These skills can be assessed through internal and external assessment.

Key terms

◆ Definitions appear throughout the margins to provide context and help you understand the language of physics. There is also a glossary of all key terms online.

Inquiry process

The application and development of the Inquiry process is supported in close association with the Tools.

Common mistake

These detail some common misunderstandings and typical errors made by students, so that you can avoid making the same mistakes yourself.

Nature of science

Nature of science (NOS) explores conceptual understandings related to the purpose, features and impact of scientific knowledge. It can be examined in Physics papers. NOS explores the scientific process itself, and how science is represented and understood by the general public. NOS covers 11 aspects: Observations, Patterns and trends, Hypotheses, Experiments, Measurements, Models, Evidence, Theories, Falsification, Science as a shared endeavour, and Global impact of science. It also examines the way in which science is the basis for technological developments and how these modern technologies, in turn, drive developments in science.

Content from the IBDP Physics data booklet is indicated with this icon and shown in bold. The data booklet contains electrical symbols, equations and constants that you need to familiarize yourself with as you progress through the course. You will have access to a copy of the data booklet during your examination.

ATL ACTIVITY

Approaches to learning (ATL) activities, including learning through inquiry, are integral to IB pedagogy. These activities are contextualized through real-world applications of physics.

Top tip!

This feature includes advice relating to the content being discussed and tips to help you retain the knowledge you need.

WORKED EXAMPLE

These provide a step-by-step guide showing you how to answer the kind of quantitative and other questions that you might encounter in your studies and in the assessment.

International mindedness is indicated with this icon. It explores how the exchange of information and ideas across national boundaries has been essential to the progress of science and illustrates the international aspects of physics.

Self-assessment questions appear throughout the chapters, phrased to assist comprehension and recall, but also to help familiarize you with the assessment implications of the command terms. These command terms are defined in the online glossary. Practice exam-style questions and their answers, together with answers to most self-assessment questions are on the accompanying website, IB Extras: **www.hoddereducation.com/ib-extras**

The IB learner profile icon indicates material that is particularly useful to help you towards developing in the following attributes: to be inquirers, knowledgeable, thinkers, communicators, principled, open-minded, caring, risk-takers, balanced and reflective. When you see the icon, think about what learner profile attribute you might be demonstrating – it could be more than one.

LINKING QUESTIONS

These questions are introduced throughout each topic. They are to strengthen your understanding by making connections across the themes. The linking questions encourage you to apply broad, integrating and discipline-specific concepts from one topic to another, ideally networking your knowledge. Practise answering the linking questions first, on your own or in groups. The links in this coursebook are not exhaustive, you may also encounter other connections between concepts, leading you to create your own linking questions.

TOK

Links to Theory of Knowledge (TOK) allow you to develop critical thinking skills and deepen scientific understanding by bringing discussions about the subject beyond the scope of the content of the curriculum.

About the author

John Allum taught physics to pre-university level in international schools for more than thirty years (as a head of department). He has now retired from teaching, but lives a busy life in a mountainside village in South East Asia. He has also been an IB examiner for many years.

Adviser, writer and reader

Paul Morris is Deputy Principal and IB Diploma Coordinator at the International School of London. He has taught IB Physics and IB Theory of Knowledge for over 20 years, has led teacher workshops internationally and has examined Theory of Knowledge. As an enthusiast for the IB concept-based continuum, Paul designed and developed Hodder Education's 'MYP by Concept' series and was author and co-author of the Physics and Sciences titles in the series. He has also advised on publishing projects for the national sciences education programmes for Singapore and Qatar.

Tools and Inquiry

Skills in the study of physics

The skills and techniques you must experience through the course are encompassed within the tools. These support the application and development of the inquiry process in the delivery of the physics course.

Tools

- **Tool 1:** Experimental techniques
- **Tool 2:** Technology
- **Tool 3:** Mathematics

Inquiry process

- **Inquiry 1:** Exploring and designing
- **Inquiry 2:** Collecting and processing data
- **Inquiry 3:** Concluding and evaluating

Throughout the programme, you will be given opportunities to encounter and practise the skills; and instead of stand-alone topics, they will be integrated into the teaching of the syllabus when they are relevant to the syllabus topics being covered.

You can see what the Tools and Inquiry boxes look like in the *How to use this book* section on page vi.

The skills in the study of physics can be assessed through internal and external assessment. The Approaches to learning provide the framework for the development of these skills.

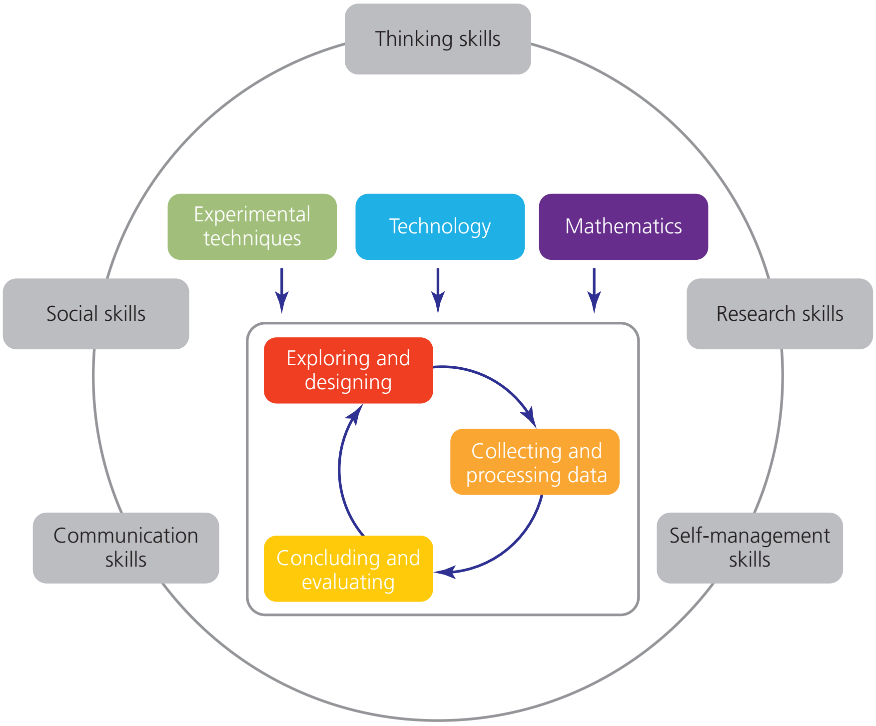


Figure 0.01 Tools for physics

Visit the link in the QR code or this website to view the Tools and Inquiry reference guide: **www.hoddereducation.com/ib-extras**

Tools

Tool 1: Experimental techniques

Skill	Description
Addressing safety of self, others and the environment	• Recognize and address relevant safety, ethical or environmental issues in an investigation.
Measuring variables	Understand how to accurately measure the following to an appropriate level of precision: • Mass • Time • Length • Volume • Temperature • Force • Electric current • Electric potential difference • Angle • Sound and light intensity

Tool 2: Technology

Skill	Description
Applying technology to collect data	• Use sensors. • Identify and extract data from databases. • Generate data from models and simulations. • Carry out image analysis and video analysis of motion.
Applying technology to process data	• Use spreadsheets to manipulate data. • Represent data in a graphical form. • Use computer modelling.

Tool 3: Mathematics

Skill	Description
Applying general mathematics	• Use basic arithmetic and algebraic calculations to solve problems. • Calculate areas and volumes for simple shapes. • Carry out calculations involving decimals, fractions, percentages, ratios, reciprocals, exponents and trigonometric ratios. • Carry out calculations involving logarithmic and exponential functions. • Determine rates of change. • Calculate mean and range. • Use and interpret scientific notation (for example, 3.5×10^6). • Select and manipulate equations. • Derive relationships algebraically. • Use approximation and estimation. • Appreciate when some effects can be neglected and why this is useful. • Compare and quote ratios, values and approximations to the nearest order of magnitude. • Distinguish between continuous and discrete variables.

Skill	Description
	• Understand direct and inverse proportionality, as well as positive and negative relationships or correlations between variables. • Determine the effect of changes to variables on other variables in a relationship. • Calculate and interpret percentage change and percentage difference. • Calculate and interpret percentage error and percentage uncertainty. • Construct and use scale diagrams. • Identify a quantity as a scalar or vector. • Draw and label vectors including magnitude, point of application and direction. • Draw and interpret free-body diagrams showing forces at point of application or centre of mass as required. • Add and subtract vectors in the same plane (limited to three vectors). • Multiply vectors by a scalar. • Resolve vectors (limited to two perpendicular components).
Using units, symbols and numerical values	• Apply and use SI prefixes and units. • Identify and use symbols stated in the guide and the data booklet. • Work with fundamental units. • Use of units (for example, eV, eVc^{-2}, ly, pc, h, day, year) whenever appropriate. • Express derived units in terms of SI units. • Check an expression using dimensional analysis of units (the formal process of dimensional analysis will not be assessed). • Express quantities and uncertainties to an appropriate number of significant figures or decimal places.
Processing uncertainties	• Understand the significance of uncertainties in raw and processed data. • Record uncertainties in measurements as a range (±) to an appropriate precision. • Propagate uncertainties in processed data in calculations involving addition, subtraction, multiplication, division and raising to a power. • Express measurement and processed uncertainties—absolute, fractional (relative) and percentage—to an appropriate number of significant figures or level of precision.
Graphing	• Sketch graphs, with labelled but unscaled axes, to qualitatively describe trends. • Construct and interpret tables, charts and graphs for raw and processed data including bar charts, histograms, scatter graphs and line and curve graphs. • Construct and interpret graphs using logarithmic scales. • Plot linear and non-linear graphs showing the relationship between two variables with appropriate scales and axes. • Draw lines or curves of best fit. • Draw and interpret uncertainty bars. • Extrapolate and interpolate graphs. • Linearize graphs (only where appropriate). • On a best-fit linear graph, construct lines of maximum and minimum gradients with relative accuracy (by eye) considering all uncertainty bars. • Determining the uncertainty in gradients and intercepts. • Interpret features of graphs including gradient, changes in gradient, intercepts, maxima and minima, and areas under the graph.

Inquiry process

Inquiry 1: Exploring and designing

Skill	**Description**
Exploring	• Demonstrate independent thinking, initiative and insight. • Consult a variety of sources. • Select sufficient and relevant sources of information. • Formulate research questions and hypotheses. • State and explain predictions using scientific understanding.
Designing	• Demonstrate creativity in the designing, implementation and presentation of the investigation. • Develop investigations that involve hands-on laboratory experiments, databases, simulations and modelling. • Identify and justify the choice of dependent, independent and control variables. • Justify the range and quantity of measurements. • Design and explain a valid methodology. • Pilot methodologies.
Controlling variables	Appreciate when and how to: • calibrate measuring apparatus, including sensors • maintain constant environmental conditions of systems • insulate against heat loss or gain • reduce friction • reduce electrical resistance • take background radiation into account.

Inquiry 2: Collecting and processing data

Skill	**Description**
Collecting data	• Identify and record relevant qualitative observations. • Collect and record sufficient relevant quantitative data. • Identify and address issues that arise during data collection.
Processing data	• Carry out relevant and accurate data processing.
Interpreting results	• Interpret qualitative and quantitative data. • Interpret diagrams, graphs and charts. • Identify, describe and explain patterns, trends and relationships. • Identify and justify the removal or inclusion of outliers in data (no mathematical processing is required). • Assess accuracy, precision, reliability and validity.

Inquiry 3: Concluding and evaluating

Skill	**Description**
Concluding	• Interpret processed data and analysis to draw and justify conclusions. • Compare the outcomes of an investigation to the accepted scientific context. • Relate the outcomes of an investigation to the stated research question or hypothesis. • Discuss the impact of uncertainties on the conclusions.
Evaluating	• Evaluate hypotheses. • Identify and discuss sources and impacts of random and systematic errors. • Evaluate the implications of methodological weaknesses, limitations and assumptions on conclusions. • Explain realistic and relevant improvements to an investigation.

A.1 Kinematics

◆ **Kinematics** Study of motion.

◆ **Classical physics** Physics theories that pre-dated the paradigm shifts introduced by quantum physics and relativity.

◆ **Uniform** Unchanging.

◆ **Magnitude** Size.

◆ **Scalars** Quantities that have only magnitude (no direction).

◆ **Vector** A quantity that has both magnitude and direction.

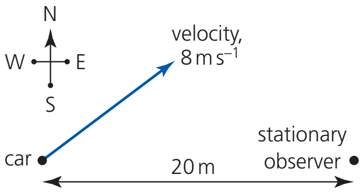
Guiding questions

- How can the motion of a body be described quantitatively and qualitatively?
- How can the position of a body in space and time be predicted?
- How can the analysis of motion in one and two dimensions be used to solve real-life problems?

Kinematics is the study of moving objects. In this topic we will describe motion by using graphs and equations, but the causes of motion (forces) will be covered in the next topic, A.2 Forces and Momentum. The ideas of **classical physics** presented in this chapter can be applied to the movement of all masses, from the very small (freely moving atomic particles) to the very large (stars).

To completely describe the motion of an object at any one moment we need to state its position, how fast it is moving, the direction in which it is moving and whether its motion is changing. For example, we might observe that a car is 20 m to the west of an observer and moving northeast with a constant (**uniform**) velocity of 8 m s^{-1}. See Figure A1.1.

Of course, any or all, of these quantities might be changing. In real life the movement of many objects can be complicated; they do not often move in straight lines and they might even rotate or have different parts moving in different directions.



■ **Figure A1.1** Describing the position and motion of a car

In this chapter we will develop an understanding of the basic principles of kinematics by dealing first with objects moving in straight lines, and calculations will be confined to those objects that have a uniform (unchanging) motion.

Tool 3: Mathematics

Identify a quantity as a scalar or a vector

Everything that we measure has a magnitude and a unit. For example, we might measure the mass of a book to be 640 g. Here, 640 g is the **magnitude** (size) of the measurement, but mass has no direction.

Quantities that have only magnitude, and no direction, are called **scalars**.

All physical quantities can be described as scalars or **vectors**.

Quantities that have both magnitude and direction are called vectors.

For example, force is a vector quantity because the direction in which a force acts is important.

Most quantities are scalars. Some common examples of scalars used in physics are mass, length, time, energy, temperature and speed.

However, when using the following quantities, we need to know both the magnitude and the direction in which they are acting, so they are vectors:

- displacement (distance in a specified direction)
- velocity (speed in a given direction)
- force (including weight)
- acceleration
- momentum and impulse
- field strength (gravitational, electric and magnetic).

In diagrams, all vectors are shown with straight arrows, pointing in a certain direction from the correct point of application.

The lengths of the arrows are proportional to the magnitudes of the vectors.

Distance and displacement

SYLLABUS CONTENT

- The motion of bodies through space and time can be described and analysed in terms of position, velocity and acceleration.
- The change in position is the displacement.
- The difference between distance and displacement.

◆ **Distance** Total length travelled, without consideration of directions.

◆ **Displacement, linear** Distance in a straight line from a fixed reference point in a specified direction.

◆ **Metre, m** SI unit of length (fundamental).

The term **distance** can be used in different ways, for example we might say that the distance between New York City and Boston is 300 km, meaning that a straight line between the two cities has a length of 300 km. Or, we might say that the (travel) distance was 348 km, meaning the length of the road between the cities.

We will define distance as follows:

Distance (of travel) is the total length of a specified path between two points. SI unit: **metre, m**

In physics, **displacement** (change of position) is often more important than distance:

The displacement of an object is the distance in a straight line from a fixed reference point in a specified direction.

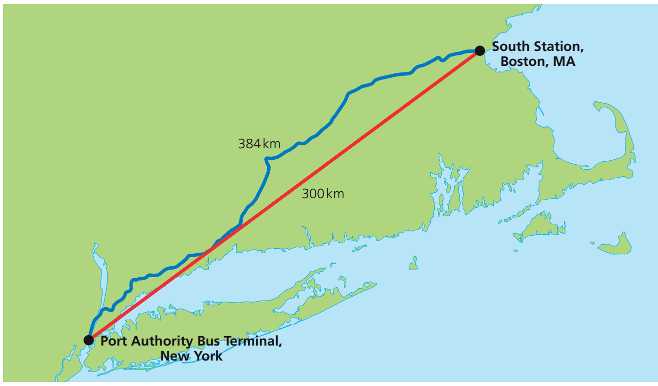


Figure A1.2 Boston is a travel distance of 384 km and a displacement of 300 km northeast from New York City

Continuing the example given above, if a girl travels from New York to Boston, her displacement will be 300 km to the northeast (see Figure A1.2).

Both distance and displacement are given the symbol *s* and the SI unit metres, m. Kilometres, km, centimetres, cm, and millimetres, mm, are also in widespread use. We often use the symbol *h* for heights and *x* for small displacements.

Figure A1.3 shows the route of some people walking around a park. The total distance walked was 4 km, but the displacement from the reference point varied and is shown every few minutes by the vector arrows (***a***–***e***). The final displacement was zero because the walkers returned to their starting place.

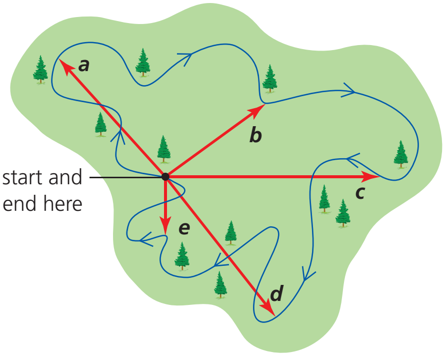


Figure A1.3 A walk in the park

Speed and velocity

SYLLABUS CONTENT

- Velocity is the rate of change of position.
- The difference between instantaneous and average values of velocity, speed and acceleration, and how to determine them.

Speed

The displacement of Wellington from Auckland, New Zealand, is 494 km south (Figure A1.4). The road distance is 642 km and it is predicted that a car journey between the two cities will take 9.0 hours.

■ **Figure A1.4** Distance and displacement from Auckland to Wellington

If we divide the total distance by the total time (642 / 9.0) we determine a speed of 71 km h^{-1}. In this example it should be obvious that the speed will have changed during the journey and the calculated result is just an **average speed** for the whole trip. The value seen on the speedometer of the car is the speed at any particular moment, called the **instantaneous speed**.

◆ **Speed, v** Average speed = distance travelled/time taken. Instantaneous speed is determined over a very short time interval, during which it is assumed that the speed does not change.

◆ **Reaction time** The time delay between an event occurring and a response. For example, the delay that occurs when using a stopwatch.

◆ **Sensor** An electrical component that responds to a change in a physical property with a corresponding change in an electrical property (usually resistance). Also called a transducer.

◆ **Light gate** Electronic sensor used to detect motion when an object interrupts a beam of light.

Tool 1: Experimental techniques

Understand how to accurately measure quantities to an appropriate level of precision: time

Accurate time measuring instruments are common, but the problem with obtaining accurate measurements of time is starting and stopping the timers at exactly the right moments.

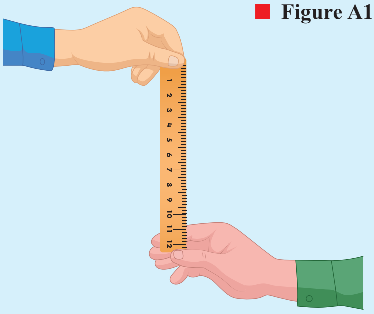
Whenever we use stopwatches or timers operated by hand, the results will have an unavoidable and variable uncertainty because of the delays between seeing an event and pressing a button to start or stop the timer. The delay between seeing something happen and responding with some kind of action is known as **reaction time**. For example, for car drivers it is usually assumed that a driver takes about 0.7 s to press the brake pedal after they have seen a problem. (But some drivers will be able to react quicker than this.) A car will travel about 14 m in this time if it is moving at 50 km h^{-1}. Reaction times will increase if the driver is distracted, tired, or under the influence of any type of drug, such as alcohol.

A simple way of determining a person's reaction time is by measuring how far a metre ruler falls before it can be caught between thumb and finger (see Figure A1.5). The time, t, can then be calculated using the equation for distance, $s = 5t^2$ (explained later in this topic).

If the distance the ruler falls $s = 0.30$

Rearranging for t, $t = \sqrt{\frac{s}{5}} = \sqrt{\frac{0.30}{5}}$

So, reaction time $t = 0.25$ s.



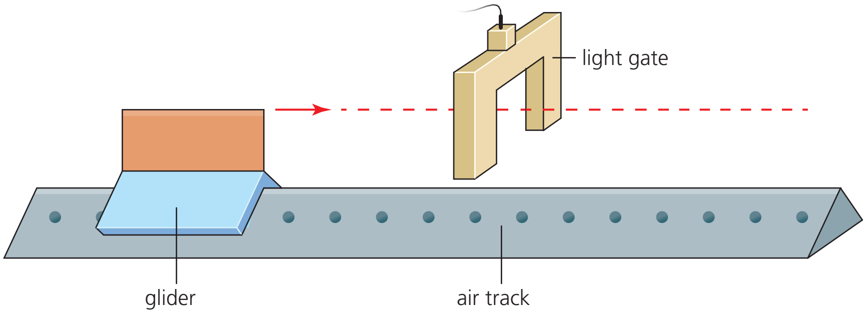
■ **Figure A1.5** Determining reaction time

Under these conditions a typical reaction time is about 0.25 s, but it can vary considerably depending on the conditions involved. The measurement can be repeated with the person tested being blindfolded to see if the reaction time changes if the stimulus (to catch the ruler) is either sound or touch, rather than sight.

In science experiments it is sensible to make time measurements as long as possible to decrease the effect of this problem. (This reduces the percentage uncertainty.) Repeating measurements and calculating an average will also reduce the effect of random uncertainties. If a stopwatch is started late because of the user's reaction time, it may be offset by also stopping the stopwatch late for the same reason.

Electronics **sensors**, such as **light gates**, are very useful in obtaining accurate time measurements. See below.

There are a number of different methods in which speed can be measured in a school or college laboratory. Figure A1.6 shows one possibility, in which a glider is moving on a frictionless air track at a constant velocity. The time taken for a card of known length (on the glider) to pass through the light gate is measured and its speed can be calculated from length of card / time taken.



■ **Figure A1.6** Measuring speed in a laboratory

Tool 2: Technology

Use sensors

An electronic sensor is an electronic device used to convert a physical quantity into an electrical signal. The most common sensors respond to changes in light level, sound level, temperature or pressure.

A light gate contains a source of light that produces a narrow beam of light directed towards a sensor on the other side of a gap. When an object passes across the light beam, the unit behaves as a switch which turns a timer on or off very quickly.

Tool 3: Mathematics

Determine rates of change

The Greek capital letter delta, Δ, is widely used in physics and mathematics to represent a change in the value of a quantity.

For example, $\Delta x = (x_2 - x_1)$, where x_2 and x_1 are two different values of the variable x.

The change involved is often considered to be relatively small.

◆ **Second, s** SI unit of time (fundamental).

Most methods of determining speed involve measuring the small amount of time (Δt) taken to travel a certain distance (Δs). The SI unit for time is the **second**, s.

$$\text{speed} = \frac{\text{distance travelled}}{\text{time taken}} \quad (\text{SI unit m s}^{-1})$$

This calculation determines an average speed during time Δt, but if Δt is small enough, we may assume that the calculated value is a good approximation to an instantaneous speed.

Speed is a scalar quantity. Speed is given the same symbol, (v), as velocity.

■ **Figure A1.7** The peregrine falcon is reported to be the world's fastest animal (speeds measured up to 390 km h^{-1})

◆ **At rest** Stays stationary in the same position.

◆ **Milky Way** The galaxy in which our Solar System is located.

Nature of science: Observations

Objects at rest

It is common in physics for people to refer to an object being **at rest**, meaning that it is not moving. But this is not as simple as it may seem. A stone may be at rest on the ground, meaning that it is not moving when compared with the ground: it appears to us to have no velocity and no acceleration. However, when the same stone is thrown upwards, at the top of its path its instantaneous speed may be zero, but it has an acceleration downwards.

We cannot assume that an object which is at rest has no acceleration; its velocity may be changing – either in magnitude, in direction, or both.

We may prefer to refer to an object being *stationary*, suggesting that an object is not moving over a period of time.

Of course, the surface of the Earth is moving, the Earth is orbiting the Sun, which orbits the centre of the **Milky Way** galaxy, which itself exists in an expanding universe. So, at a deeper level, we must understand that *all* motion is relative and nowhere is truly stationary. This is the starting point for the study of Relativity (Topic A.5).

Velocity

Velocity, v, is the rate of change of position. It may be considered to be speed in a specified direction.

◆ **Velocity, v** Rate of change of position.

$$\text{velocity, } v = \frac{\text{displacement}}{\text{time taken}} = \frac{\Delta s}{\Delta t} \quad (\text{SI unit m s}^{-1})$$

The symbol Δs represents a change of position (displacement).

Velocity is a vector quantity. $12\,\text{m s}^{-1}$ is a speed. $12\,\text{m s}^{-1}$ to the south is a velocity. We use positive and negative signs to represent velocities in opposite directions. For example, $+12\,\text{m s}^{-1}$ may represent a velocity upwards, while $-12\,\text{m s}^{-1}$ represents the same speed downwards, but we may choose to reverse the signs used.

Speed and velocity are both represented by the same symbol (v) and their magnitudes are calculated in the same way $\left(v = \frac{\Delta s}{\Delta t}\right)$ with the same units. It is not surprising that these two terms are sometimes used interchangeably and this can cause confusion. For this reason, it may be better to define these two quantities in words, rather than symbols.

As with speed, we may need to distinguish between average velocity over a time interval, or instantaneous velocity at a particular moment. As we shall see, the value of an instantaneous velocity can be determined from the gradient of a displacement–time graph.

Top tip!

When a direction of motion is clearly stated (such as ‘up’, ‘to the north’, ‘to the right’ and so on), it is very clear that a velocity is being discussed. However, we may commonly refer to the ‘velocity’ of a car, for example, without stating a direction. Although this is casual, it is usually acceptable because an unchanging direction is implied, even if it is not specified. For example, we may assume that the direction of the car is along a straight road.

WORKED EXAMPLE A1.1

A satellite moves in circles along the same path around the Earth at a constant distance of 6.7×10^3 km from the Earth's centre. Each **orbit** takes a time of 90 minutes.

a Calculate the average speed of the satellite.

b Describe the instantaneous velocity of the satellite.

c Determine its displacement from the centre of the Earth after

- **i** 360 minutes
- **ii** 405 minutes.

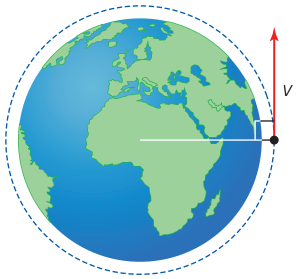
Answer

a $v = \frac{\text{circumference}}{\text{time for orbit}} = \frac{2\pi r}{\Delta t}$

$= \frac{(2 \times \pi \times 6.7 \times 10^6)}{(90 \times 60)}$

$= 7.8 \times 10^3\,\text{m}\,\text{s}^{-1}$

b The velocity also has a constant magnitude of $7.8 \times 10^3\,\text{m}\,\text{s}^{-1}$, but its direction is continuously changing. Its instantaneous velocity is always directed along a **tangent** to its circular orbit. See Figure A1.8.



■ **Figure A1.8** Satellite's instantaneous velocity

◆ **Orbit** The curved path (may be circular) of a mass around a larger central mass.

◆ **Tangent** Line which touches a given curve at a single point.

c **i** 360 minutes is the time for four complete orbits. The satellite will have returned to the same place. Its displacement from the centre of the Earth compared to 360 minutes earlier will be the same. (But the Earth will have rotated.)

ii In the extra 45 minutes the satellite will have travelled half of its orbit. It will be on the opposite side of the Earth's centre, but at the same distance. We could represent this as -6.7×10^3 km from the Earth's centre.

1 Calculate the average speed ($\text{m}\,\text{s}^{-1}$) of an athlete who can run a marathon (42.2 km) in 2 hours, 1 minute and 9 seconds. (The men's world record at the time of writing.)

■ **Figure A1.9** Eliud Kipchoge, world record holder for the men's marathon

2 A small ball dropped from a height of 2.0 m takes 0.72 s to reach the ground.

- **a** Calculate $\frac{2.0}{0.72}$
- **b** What does your answer represent?
- **c** The speed of the ball just before it hits the ground is $5.3\,\text{m}\,\text{s}^{-1}$. This is an instantaneous speed. Distinguish between an instantaneous value and an average value.
- **d** State the instantaneous velocity of the ball just before it hits the ground.
- **e** After bouncing, the ball only rises to a lower height. Give a rough estimate of the instantaneous velocity of the ball as it leaves the ground.

3 A magnetic field surrounds the Earth and it can be detected by a compass. State whether it is a scalar or a vector quantity. Explain your answer.

4 On a flight from Rome to London, a figure of $900\,\text{km}\,\text{h}^{-1}$ is displayed on the screen.

- **a** State whether this is a speed or a velocity.
- **b** Is it an average or instantaneous value?
- **c** Convert the value to $\text{m}\,\text{s}^{-1}$.
- **d** Calculate how long it will take the aircraft to travel a distance of 100 m.

Acceleration

SYLLABUS CONTENT

- Acceleration is the rate of change of velocity.
- Motion with uniform and non-uniform acceleration.

◆ **Acceleration, *a*** Rate of change of velocity with time. Acceleration is a vector quantity.

◆ **Deceleration** Term commonly used to describe a decreasing speed.

Any variation from moving at a constant speed in a straight line is described as an **acceleration**.

Going faster, going slower and/or changing direction are all different kinds of acceleration (changing velocities).

When the velocity (or speed) of an object changes during a certain time, the symbol u is used for the initial velocity and the symbol v is used for the final velocity. These velocities are not necessarily the beginning and end of the entire motion, just the velocities at the start and end of the period of time that is being considered.

Acceleration, a, is defined as the rate of change of velocity with time:

$$a = \frac{\Delta v}{\Delta t} = \frac{(v - u)}{t} \quad \text{(SI unit m s}^{-2}\text{)}$$

One way to determine an acceleration is to measure two velocities and the time between the measurements. Figure A1.10 shows an example.

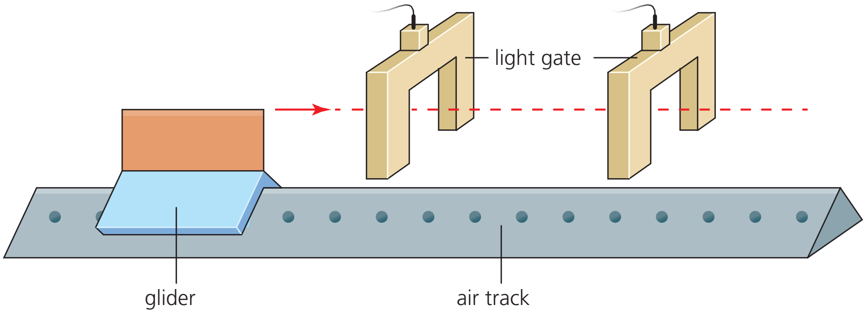


Figure A1.10 Measuring two velocities to determine an acceleration

Acceleration is a vector quantity. For a typical motion in which displacement and velocity are both given positive values, a positive acceleration means increasing speed in the same direction ($+\Delta v$), while a negative acceleration means decreasing speed in the same direction ($-\Delta v$). In everyday speech, a reducing speed is often called a **deceleration**.

For a motion in which displacement and velocity are given negative values, a positive acceleration means a decreasing speed. For example, a velocity change from $-6\,\text{m s}^{-1}$ to $-4\,\text{m s}^{-1}$ in 0.5 s corresponds to an acceleration:

$$a = \frac{\Delta v}{\Delta t} = \frac{([-4]-[-6])}{0.5} = +\,4\,\text{ms}^{-2}$$

As with speed and velocity, we may need to distinguish between average acceleration over a time interval, or instantaneous acceleration at a particular moment.

WORKED EXAMPLE A1.2

A high-speed train travelling with a velocity of 84 m s^{-1} needs to slow down and stop in a time of one minute.

a Determine the necessary average acceleration.

b Calculate the distance that the train will travel in this time assuming that the acceleration is uniform.

Answer

a $a = \frac{\Delta v}{\Delta t} = \frac{(0 - 84)}{60} = -1.4\,\text{m s}^{-2}$

The acceleration is negative. The negative sign shows that the velocity is decreasing.

b average speed $= \frac{(84 - 0)}{2} = 42\,\text{m s}^{-1}$

distance = average speed × time = 42 × 60 = 2.5×10^3 m

5 A car moving at 12.5 m s^{-1} accelerates uniformly on a straight road at a rate of 0.850 m s^{-2}.

- **a** Calculate its velocity after 4.60 s.
- **b** What uniform rate of acceleration will reduce the speed to 5.0 m s^{-1} in a further 12 s?

6 An athlete accelerates uniformly from rest at the start of a race at a rate of 4.3 m s^{-2}. How much time is needed before her speed has reached 8.0 m s^{-1}?

7 A trolley takes 3.62 s to accelerate from rest uniformly down a slope at a rate of 0.16 m s^{-2}. A light gate at the bottom of the slope records a velocity of 0.58 m s^{-1}. What was the speed about halfway down the slope, 1.2 s earlier?

Inquiry 1: Exploring and designing

Designing

Suppose that the Principal of your school or college is worried about safety from traffic on the nearby road. He has asked your physics class to collect evidence that he can take to the police. He is concerned that the traffic travels too fast and that the vehicles do not slow down as they approach the school.

1. Using a team of students, working over a period of one week, with tape measures and stop watches, develop an investigation which will produce sufficient and accurate data that can be given in a report to the Principal. Explain how you would ensure that the investigation was carried out safely.
2. What is the best way of presenting a summary of this data?

Tool 3: Mathematics

Interpret features of graphs

In order to analyse and predict motions we have two methods: graphical and algebraic. Firstly, we will look at how motion can be represented graphically.

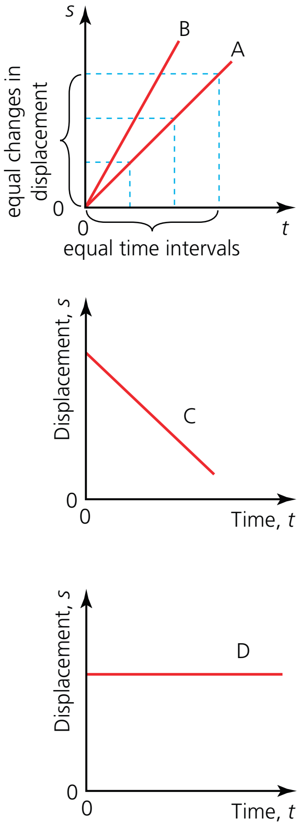
Graphs can be drawn to represent any motion and they provide extra understanding and insight (at a glance) that very few of us can get from written descriptions or equations. Furthermore, the gradients of graphs and the areas under graphs often provide additional useful information.

Displacement–time graphs and distance–time graphs

◆ **Linear relationship** One which produces a straight line graph.

Displacement–time graphs, similar to those shown in Figure A1.11, show how the displacements of objects from a known reference point vary with time. All the examples shown in Figure A1.11 are straight lines and are representing **linear relationships** and constant velocities.

- Line A represents an object moving away from the reference point (zero displacement) such that equal displacements occur in equal times. That is, the object has a constant velocity.



■ **Figure A1.11** Constant velocities on displacement–time graphs

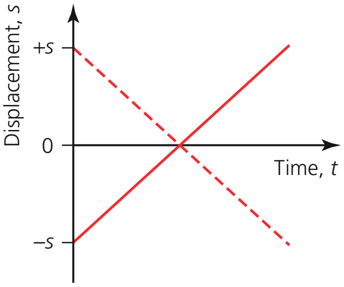
Any linear displacement–time graph represents a constant velocity (it does not need to start or end at the origin).

- Line B represents an object moving with a greater velocity than A.
- Line C represents an object that is moving back towards the reference point.
- Line D represents an object that is stationary (at rest). It has zero velocity and stays at the same distance from the reference point.

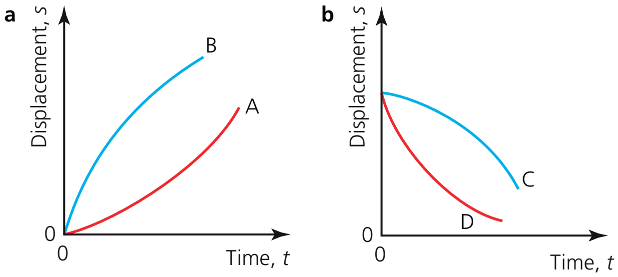
Figure A1.12 shows how we can represent displacements in opposite directions from the same reference point.

The solid line represents the motion of an object moving with a constant (positive) velocity. The object moves towards a reference point (where the displacement is zero), passes it, and then moves away from the reference point with the same velocity. The dotted line represents an identical speed in the opposite direction (or it could also represent the original motion if the directions chosen to be positive and negative were reversed).

Any curved (non-linear) line on a displacement–time graph represents a changing velocity, in other words, an acceleration. This is illustrated in Figure A1.13.



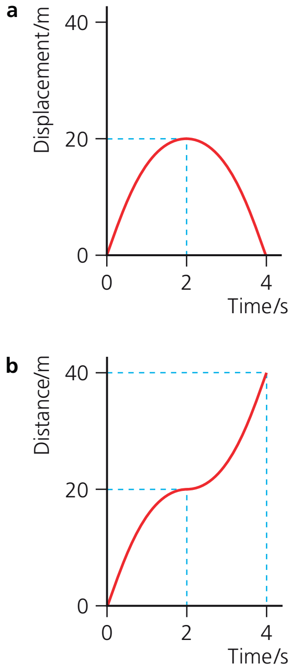
■ **Figure A1.12** Motion in opposite directions represented on a displacement–time graph



■ **Figure A1.13** Accelerations on displacement–time graphs

Figure A1.13a shows motion away from a reference point. Line A represents an object accelerating. Line B represents an object decelerating. Figure A1.13b shows motion towards a reference point. Line C represents an object accelerating. Line D represents an object decelerating. The values of the accelerations represented by these graphs may, or may not, be constant. (This cannot be determined without a more detailed analysis.)

In physics, we are usually more concerned with displacement–time graphs than distance–time graphs. In order to explain the difference, consider Figure A1.14.



■ **Figure A1.14** **a** Displacement–time and **b** distance–time graphs for an object moving up and then down

Figure A1.14a shows a displacement–time graph for an object thrown vertically upwards with an initial speed of $20\,\mathrm{m\,s^{-1}}$ (without air resistance). It takes 2 s to reach a maximum height of 20 m. At that point it has an instantaneous velocity of zero, before returning to where it began after 4 s and regaining its initial speed. Figure A1.14b is a less commonly used graph showing how the same motion would appear on an overall distance–time graph.

Tool 3: Mathematics

Interpret features of graphs: gradient

In this topic we will need to repeatedly use the following information:

- The gradient of a displacement–time graph equals velocity.
- The gradient of a velocity–time graph equals acceleration.

In the following section we will explore how to measure and interpret gradients.

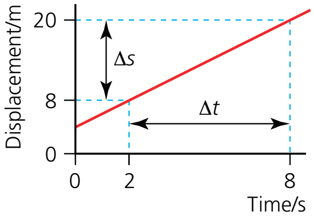


Figure A1.15 Finding a constant velocity from a displacement–time graph

Gradients of displacement–time graphs

Consider the motion at *constant* velocity represented by Figure A1.15.

The **gradient** of the graph $= \dfrac{\Delta s}{\Delta t}$, which is the velocity of the object. A downwards sloping graph would have a negative gradient (velocity).

In this example,

$$\text{constant velocity, } v = \frac{\Delta s}{\Delta t} = \frac{(20 - 8.0)}{(8.0 - 2.0)} = 2.0\,\text{m}\,\text{s}^{-1}$$

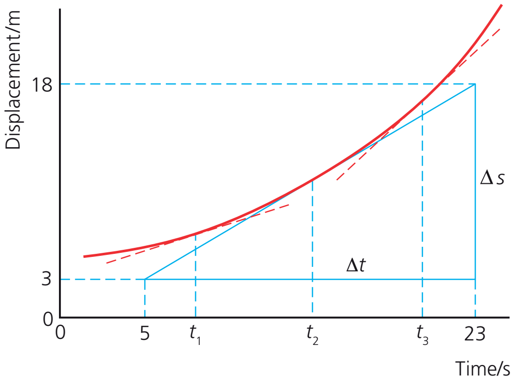


Figure A1.16 Finding an instantaneous velocity from a curved displacement–time graph

Figure A1.16 represents the motion of an object with a *changing* velocity, that is, an accelerating object.

The gradient of this graph varies, but at any point it is still equal to the velocity of the object at that moment, that is, the instantaneous velocity.

The gradient (velocity) can be determined at any time by drawing a tangent to the curve, as shown.

The triangle used to calculate the gradient should be large, in order to make this process as accurate as possible. In this example:

$$\text{velocity at time } t_2 = \frac{(18 - 3.0)}{(23 - 5.0)} = 0.83\,\text{m}\,\text{s}^{-1}$$

A tangent drawn at time t_1 would have a smaller gradient and represent a smaller velocity. A tangent drawn at time t_3 would represent a larger velocity.

◆ **Gradient** The rate at which one physical quantity changes in response to changes in another physical quantity. Commonly, for an y–x graph, gradient $= \dfrac{\Delta y}{\Delta x}$.

We have been referring to the object's displacement and velocity, although no direction has been stated. This is acceptable because that information would be included when the origin of the graph was explained. If information was presented in the form of a distance–time graph, the gradient would represent the speed.

In summary:

The gradient of a displacement–time graph represents velocity.

The gradient of a distance–time graph represents speed.

WORKED EXAMPLE A1.3

Figure A1.17 represents the motion of a train on a straight track between two stations.

a Describe the motion.

b State the distance between the two stations.

c Calculate the maximum speed of the train.

d Determine the average speed of the train.

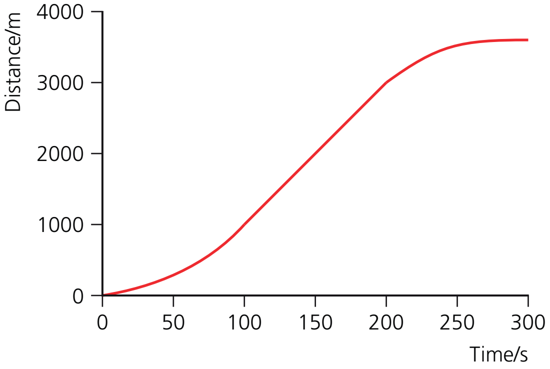


Figure A1.17 Distance–time graph for train on a straight track

Answer

a The train started from rest. For the first 90 s the train was accelerating. It then travelled with a constant speed until a time of 200 s. After that, its speed decreased to become zero after 280 s.

b 3500 m

c From the steepest, straight section of the graph:

$$v = \frac{\Delta s}{\Delta t} = \frac{(3000 - 800)}{(200 - 90)} = 20\,\text{m}\,\text{s}^{-1}$$

d $\text{average speed} = \dfrac{\text{total distance travelled}}{\text{time taken}} = \dfrac{3500}{300} = 11.7\,\text{m}\,\text{s}^{-1}$

8 Draw a displacement–time graph for a swimmer swimming a total distance of 100 m at a constant speed of $1.0\,\text{m}\,\text{s}^{-1}$ in a swimming pool of length 50 m.

9 Describe the motion of a runner as shown by the graph in Figure A1.18.

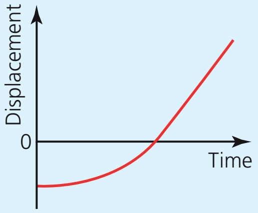


Figure A1.18 Displacement–time graph for a runner

10 Sketch a displacement–time graph for the following motion: a stationary car is 25 m away; 2 s later it starts to move further away in a straight line from you with a constant acceleration of $1.5\,\text{m}\,\text{s}^{-2}$ for 4 s; then it continues with a constant velocity for another 8 s.

11 Figure A1.19 is a displacement–time graph for an object.

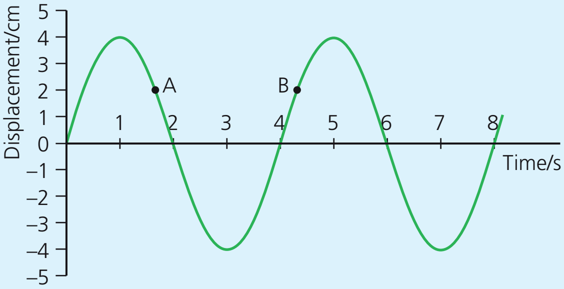


Figure A1.19 A displacement–time graph for an object

- **a** Describe the motion represented by the graph in Figure A1.19.
- **b** Compare the velocities at points A and B.
- **c** When is the object moving with its maximum and minimum velocities?
- **d** Estimate values for the maximum and minimum velocities.
- **e** Suggest what kind of object could move in this way.

Velocity–time graphs and speed–time graphs

Figure A1.20, shows how the velocity of four objects changed with time. Any straight (linear) line on any velocity–time graph shows that equal changes of velocity occur in equal times – that is, it represents *constant* acceleration.

- Line A shows an object that has a constant positive acceleration.
- Line B represents an object moving with a greater positive acceleration than A.
- Line C represents an object that has a negative acceleration.
- Line D represents an object moving with a constant velocity – that is, it has zero acceleration.

Curved lines on velocity–time graphs represent *changing* accelerations.

Velocities in opposite directions are represented by positive and negative values.

We will return to the example shown in Figure A1.14 to illustrate the difference between velocity–time and speed–time graphs. Figure A1.21a shows how the speed of an object changes as it is thrown up in the air (without air resistance), reaches its highest point, where its speed has reduced to zero, and then returns downwards. Figure A1.21b shows the same information in terms of velocity. Positive velocity represents motion upwards, negative velocity represents motion downwards. In most cases, the velocity graph is preferred to the speed graph.

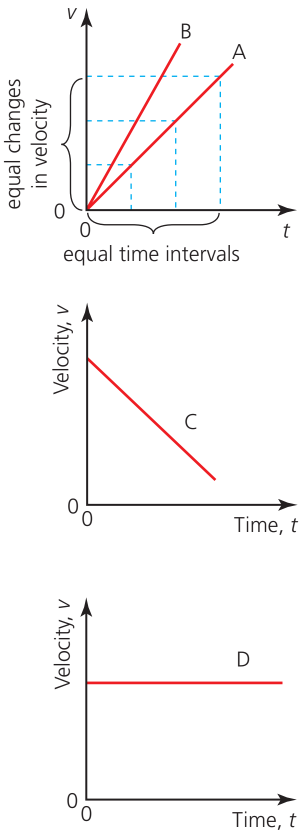


Figure A1.20 Constant accelerations on velocity–time graphs

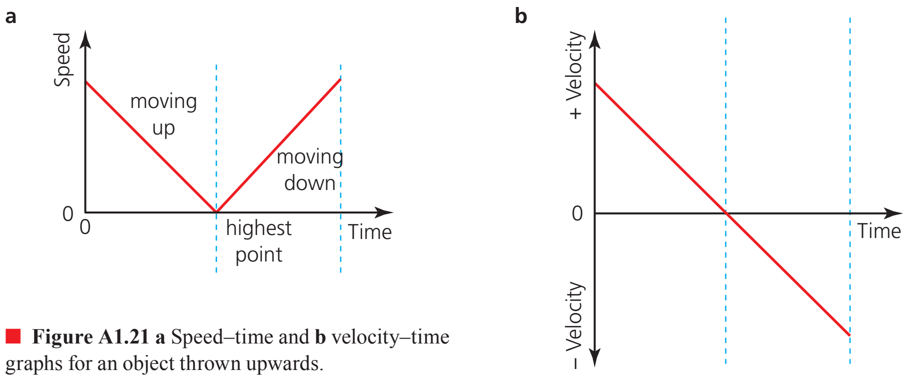
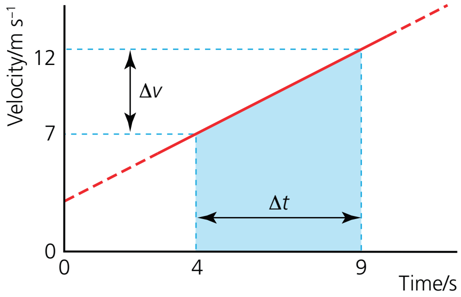


Figure A1.21 **a** Speed–time and **b** velocity–time graphs for an object thrown upwards.

Gradients of velocity–time graphs

Consider the motion at constant acceleration shown by the straight line in Figure A1.22.



■ **Figure A1.22** Finding the gradient of a velocity–time graph

The gradient of the graph $= \frac{\Delta v}{\Delta t}$, which is equal to the acceleration of the object.

In this example, the constant acceleration:

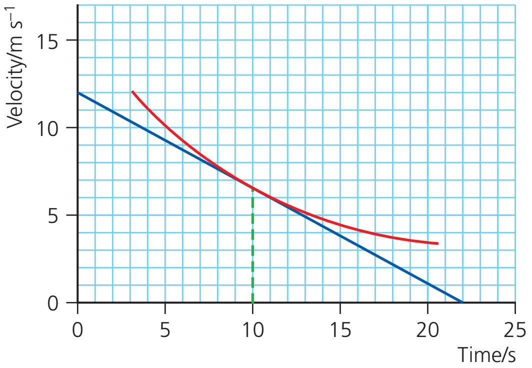
$$a = \frac{\Delta v}{\Delta t} = \frac{(12.0 - 7.0)}{(9.0 - 4.0)} = + 1.0\,\mathrm{m\,s^{-2}}$$

The acceleration of an object is equal to the gradient of the velocity–time graph.

A *changing* acceleration will appear as a curved line on a velocity–time graph. A numerical value for the acceleration at any time can be determined from the gradient of the graph at that moment. See Worked example A1.4.

WORKED EXAMPLE A1.4

The red line in Figure A1.23 shows an object decelerating (a decreasing negative acceleration). Use the graph to determine the instantaneous acceleration at a time of 10.0 s.



■ **Figure A1.23** Finding an instantaneous acceleration from a velocity–time graph

Answer

Using a tangent to the curve drawn at $t = 10$ s.

$$\text{Acceleration, } a = \frac{\Delta v}{\Delta t} = \frac{(0 - 12)}{(22 - 0)} = -0.55\,\mathrm{m\,s^{-2}}$$

The negative sign indicates a deceleration. In this example the large triangle used to determine the gradient accurately was drawn by extending the tangent to the axes for convenience.

Tool 3: Mathematics

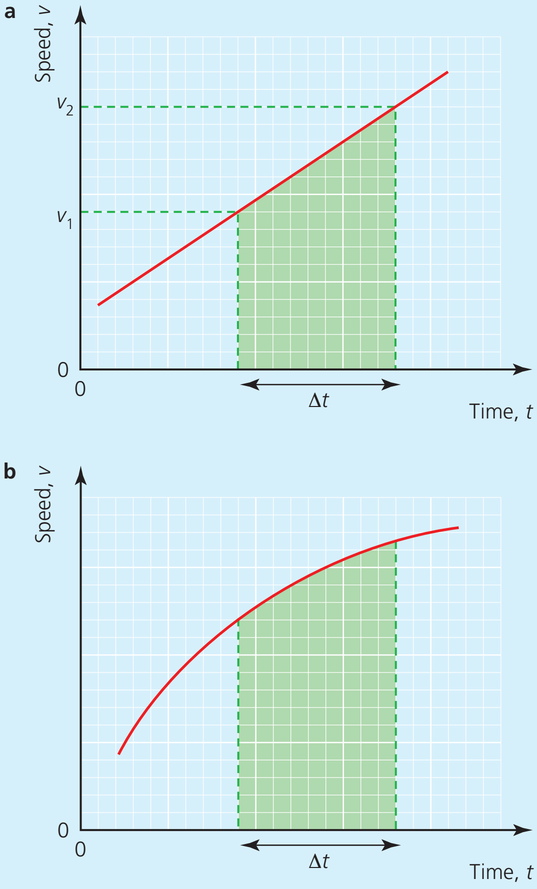
Interpret features of graphs: areas under the graph

The area under many graphs has a physical meaning. As an example, consider Figure A1.24a, which shows part of a speed–time graph for a vehicle moving with constant acceleration. The area under the graph (the shaded area) can be calculated from the average speed, given by $\frac{(v_1 + v_2)}{2}$, multiplied by the time, Δt.

The area under the graph is therefore equal to the distance travelled in time Δt. In Figure A1.24b a vehicle is moving with a changing (decreasing) acceleration, so that the graph is curved, but the same rule applies – the area under the graph (shaded) represents the distance travelled in time Δt.

The area in Figure A1.24b can be estimated in a number of different ways, for example by counting small squares, or by drawing a rectangle that appears (as judged by eye) to have the same area. (If the equation of the line is known, it can be calculated using the process of **integration**, but this is *not* required in the IB course.)

In the following section, we will show how a change in displacement can be calculated from a velocity–time graph.



■ **Figure A1.24** Area under a speed–time graph for **a** constant acceleration and **b** changing acceleration

◆ **Integration** Mathematical process used to determine the area under a graph.

Areas under velocity–time and speed–time graphs

As an example, consider again Figure A1.22. The change of displacement, Δs, between the fourth and ninth seconds can be found from (average velocity) × time.

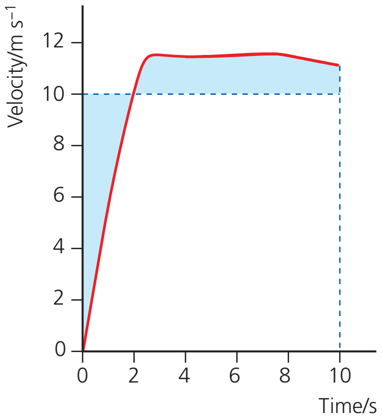
$$\Delta s = \frac{(12.0 + 7.0)}{2} \times (9.0 - 4.0) = 47.5\,\text{m}$$

This is numerically equal to the area under the line between $t = 4.0\,\text{s}$ and $t = 9.0\,\text{s}$ (as shaded in Figure A1.22). This is always true, whatever the shape of the line.

The area under a velocity–time graph is always equal to the change of displacement.

The area under a speed–time graph is always equal to the distance travelled.

As an example, consider Figure A1.21a. The two areas under the speed–time graph are equal and they are both positive. Each area equals the vertical height travelled by the object. The total area = total distance = twice the height. Each area under the velocity graph also represents the height, but the total area is zero because the areas above and below the time axis are equal, indicating that the final displacement is zero – the object has returned to where it started.



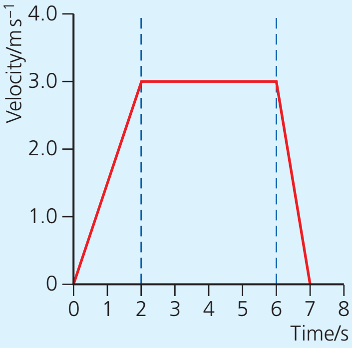
■ **Figure A1.25** Velocity–time graph for an athlete running 100 m

Figure A1.25 shows a velocity–time graph for an athlete running 100 m in 10.0 s. The area under the curve is equal to 100 m and it equals the area under the dotted line. (The two shaded areas are judged by sight to be equal.) The initial acceleration of the athlete is very important, and in this example, it is about 5 m s^{-2}.

■ **Figure A1.26** Elaine Thompson-Herah (Jamaica) won the women's 100 m in the Tokyo Olympics in 2021 in a time of 10.54 s

12 Look at the graph in Figure A1.27.

- **a** Describe the straight-line motion represented by the graph.
- **b** Calculate accelerations for the three parts of the journey.
- **c** What was the total distance travelled?
- **d** What was the average velocity?



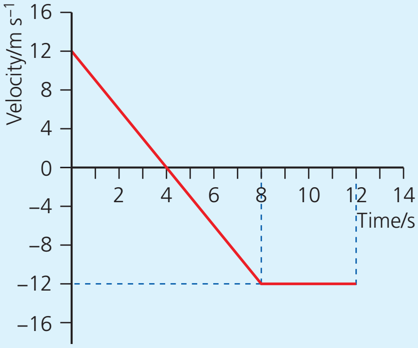
■ **Figure A1.27** Velocity–time graph

13 The velocity of a car was read from its speedometer at the moment it started and every 2 s afterwards. The successive values (converted to m s^{-1}) were: 0, 1.1, 2.4, 6.9, 12.2, 18.0, 19.9, 21.3 and 21.9.

- **a** Draw a graph of these readings.
- **b** Use the graph to estimate
  - **i** the maximum acceleration
  - **ii** the distance covered in 16 s.

14 Look at the graph in Figure A1.28.

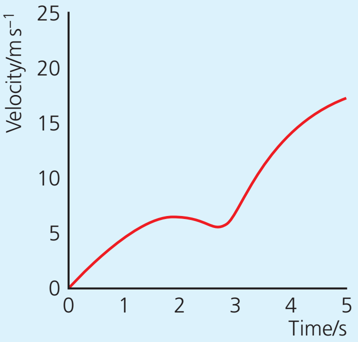
- **a** Describe the straight-line motion of the object represented by the graph.
- **b** Calculate the acceleration during the first 8 s.
- **c** What was the total distance travelled in 12 s?
- **d** What was the total displacement after 12 s?
- **e** What was the average velocity during the 12 s interval?



■ **Figure A1.28** Velocity–time graph

15 Sketch a velocity–time graph of the following motion: a car is 100 m away and travelling along a straight road towards you at a constant velocity of 25 m s^{-1}. Two seconds after passing you, the driver decelerates uniformly and the car stops 62.5 m away from you.

16 Figure A1.29 shows how the velocity of a car, moving in a straight line, changed in the first 5 s after starting. Use the area under the graph to show that the distance travelled was about 40 m.



■ **Figure A1.29** Determining the displacement of a car during acceleration

Tool 2: Technology

Use spreadsheets to manipulate data

Figure A1.30 represents how the velocities of two identical cars changed from the moment that their drivers saw danger in front of them and tried to stop their cars as quickly as possible. It has been assumed that both drivers have the same reaction time (0.7 s) and both cars decelerate at the same rate ($-5.0\,\mathrm{m\,s^{-2}}$).

The distance travelled at constant velocity before the driver reacts and depresses the brake pedal is known as the 'thinking distance'. The distance travelled while decelerating is called the 'braking distance'. The total stopping distance is the sum of these two distances.

Car B, travelling at twice the velocity of car A, has twice the thinking distance. That is, the thinking distance is proportional to the velocity of the car. The distance travelled when braking, however, is proportional to the velocity squared. This can be confirmed from the areas under the v–t graphs. The area under graph B is four times the area under graph A (during the deceleration). This has important implications for road safety and most countries make sure that people learning to drive must understand how stopping distances change with the vehicle's velocity. Some countries measure the reaction times of people before they are given a driving licence.

Set up a **spreadsheet** that will calculate the total stopping distance for cars travelling at initial speeds, u, between 0 and $40\,\mathrm{m\,s^{-1}}$ with a deceleration of $-6.5\,\mathrm{m\,s^{-2}}$. (Make calculations every $2\,\mathrm{m\,s^{-1}}$.) The thinking distance can be calculated from $s_t = 0.7u$ (reaction time 0.7 s).

In this example the braking time can be calculated from:

$$t_b = \frac{u}{6.5}$$

and the braking distance can be calculated from:

$$s_b = \left(\frac{u}{2}\right)t_b$$

Use the data produced to plot a computer-generated graph of stopping distance (y-axis) against initial speed (x-axis).

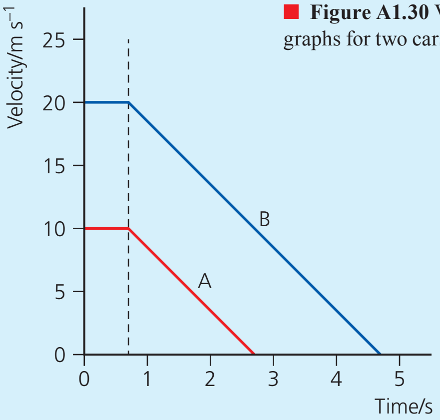


Figure A1.30 Velocity–time graphs for two cars braking

◆ **Spreadsheet (computer)** Electronic document in which data is arranged in the rows and columns of a grid, and can be manipulated and used in calculations.

Acceleration–time graphs

In this topic, we are mostly concerned with constant accelerations. The graphs in Figure A1.31 show five straight lines representing *constant* accelerations. A *changing* acceleration would be represented by a curved line on the graph.

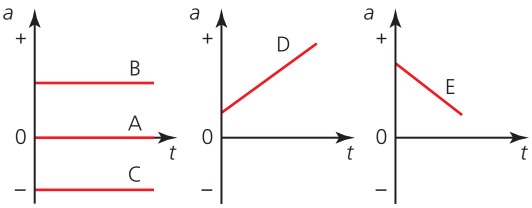


Figure A1.31 Graphs of constant acceleration

- Line A shows zero acceleration, constant velocity.
- Line B shows a constant positive acceleration (uniformly increasing velocity).
- Line C shows the constant negative acceleration (deceleration) of an object that is slowing down at a uniform rate.
- Line D shows a (linearly) increasing positive acceleration.
- Line E shows an object that is accelerating positively, but at a (linearly) decreasing rate.

Areas under acceleration–time graphs

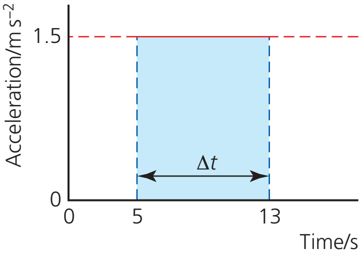


Figure A1.32 Calculating change of velocity from an acceleration–time graph

Figure A1.32 shows the constant acceleration of a moving car.

Using $a = \frac{\Delta v}{\Delta t}$, between the fifth and thirteenth seconds, the velocity of the car increased by:

$$\Delta v = a\Delta t = 1.5 \times (13.0 - 5.0) = 12\,\mathrm{m\,s^{-1}}$$

The change in velocity is numerically equal to the area under the line between $t = 5$ s and $t = 13$ s (the shaded area in Figure A1.32). This is always true, whatever the shape of the line.

The area under an acceleration–time graph is equal to the change of velocity.

17 Draw an acceleration–time graph for a car that starts from rest, accelerates at 2 m s^{-2} for 5 s, then travels at constant velocity for 8 s, before decelerating uniformly to rest again in a further 2 s.

18 Figure A1.33 shows how the acceleration of a car changed during a 6 s interval. If the car was travelling at 2 m s^{-1} after 1 s, estimate a suitable area under the graph and use it to determine the approximate speed of the car after another 5 s.

19 Sketch displacement–time, velocity–time and acceleration–time graphs for a bouncing ball that was dropped from rest. Continue the sketches until the third time that the ball contacts the ground.

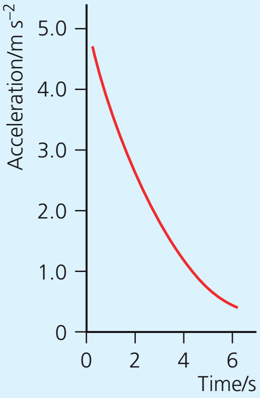


Figure A1.33 Acceleration–time graph for an accelerating car

◆ **Calculus** Branch of mathematics which deals with continuous change.

◆ **Differentiate** Mathematically determine an equation for a rate of change.

TOK

Mathematics and the arts

- Why is mathematics so important in some areas of knowledge, particularly the natural sciences?

If you study Mathematics: Analysis and Approaches (SL or HL) or Mathematics: Applications and Interpretations (HL) you will explore how **calculus** is used to mathematically describe changing functions. The gradient of a function is found using the process of **differentiation** and the area under a curve is found using the process of integration. The mathematical procedures for calculus were developed by Isaac Newton and he first published his ‘method of fluxions’ as an appendix to his book *Opticks* in 1704. Newton is usually therefore credited with the ‘invention’ of calculus – although historians of science point to the earlier work of Gottfried Wilhelm Leibniz, published in 1684. Newton accused Leibniz of plagiarism, even though Leibniz’s work was published first! In fact, it is Leibniz’s notation that we still use today. So, who invented calculus?

◆ **Equations of motion** Equations that can be used to make calculations about objects that are moving with uniform acceleration.

Equations of motion for uniformly accelerated motion

SYLLABUS CONTENT

- The **equations of motion** for solving problems with uniformly accelerated motion as given by:

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

  $v = u + at$

  $s = ut + \frac{1}{2}at^2$

  $v^2 = u^2 + 2as$

The five quantities u, v, a, s and t are all that is needed to fully describe the motion of an object that is moving with *uniform* acceleration.

- u = velocity (speed) at the start of time t
- v = velocity (speed) at the end of time t
- a = acceleration (constant)
- s = displacement occurring in time t
- t = time taken for velocity (speed) to change from u to v and to travel a distance s.

If any three of the quantities are known, the other two can be calculated using the first two equations highlighted below.

If we know the initial velocity u and the uniform acceleration a of an object, then we can determine its final velocity v after a time t by rearranging the equation used to define acceleration:

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

This gives:

$$\boldsymbol{v = u + at}$$

If an object moving with velocity u accelerates uniformly to a velocity v, then its average velocity is:

$$\frac{(u + v)}{2}$$

Then, since distance = average velocity × time:

$$\boldsymbol{s = \frac{(u + v)}{2}t}$$

These two equations can be combined mathematically to give two further equations, shown below. These very useful equations do not involve any further physics theory, they just express the same physics principles in a different way.

$$\boldsymbol{s = ut + \frac{1}{2}at^2}$$

$$\boldsymbol{v^2 = u^2 + 2as}$$

LINKING QUESTION

- How are the equations for rotational motion related to those for linear motion?

This question links to understandings in Topic A.4.

WORKED EXAMPLE A1.5

A Formula One racing car (see Figure A1.34) accelerates from rest at 18 m s^{-2}.

Figure A1.34 Formula One racing cars at the starting grid

a Calculate its speed after 3.0 s.

b Calculate how far it travels in this time.

c If it continues to accelerate at the same rate, determine its velocity after it has travelled 200 m from the start.

Answer

a $v = u + at = 0 + (18 \times 3.0) = 54\,\text{m s}^{-1}$

b $s = \frac{(u + v)}{2}t = \frac{(0 + 54)}{2} \times 30 = 81\,\text{m}$

But note that the distance can be calculated directly, without first calculating the final velocity, as follows:

$$s = ut + \frac{1}{2}at^2 = (0 \times 3.0) + (0.5 \times 18 \times 3.0^2) = 81\,\text{m}$$

c $v^2 = u^2 + 2as = 0^2 + (2 \times 18 \times 200) = 7200$

$v = 85\,\text{m s}^{-1}$

WORKED EXAMPLE A1.6

A train travelling at 50 m s^{-1} (180 km h^{-1}) needs to decelerate uniformly so that it stops at a station 2.0 kilometres away.

a Determine the necessary deceleration.

b Calculate the time needed to stop the train.

Answer

a $v^2 = u^2 + 2as$

$0^2 = 50^2 + (2 \times a \times 2000)$

$a = -0.63\,\text{m s}^{-2}$

b $v = u + at$

$0 = 50 + (-0.63) \times t$

$t = 80\text{ s}$

Alternatively, you could use $s = \frac{(u + v)}{2}t$

In the following questions, assume that all accelerations are uniform.

20 A ball rolling down a slope passes a point P with a velocity of 1.2 m s^{-1}. A short time later it passes point Q with a velocity of 2.6 m s^{-1}.

a What was its average velocity between P and Q?

b If it took 1.4 s to go from P to Q, determine the distance PQ.

c Calculate the acceleration of the ball.

21 An aircraft accelerates from rest along a runway and takes off with a velocity of 86.0 m s^{-1}. Its acceleration during this time is 2.40 m s^{-2}.

a Calculate the distance along the runway that the aircraft needs to travel before take-off.

b Predict how long after starting its acceleration the aircraft takes off.

22 An ocean-going cruiser can decelerate no quicker than $0.0032\,\mathrm{m\,s^{-2}}$.

Figure A1.35 Ocean-going cruise liner

a Determine the minimum distance needed to stop if the ship is travelling at 10 knots. ($1\text{ knot} = 0.514\,\mathrm{m\,s^{-1}}$)

b How much time does this deceleration require?

23 An advertisement for a new car states that it can travel 100 m from rest in 8.2 s.

a Discuss why the car manufacturers express the acceleration in this way (or the time needed to reach a certain speed).

b Calculate the average acceleration.

c Calculate the velocity of the car after this time.

24 A car travelling at a constant velocity of $21\,\mathrm{m\,s^{-1}}$ (faster than the speed limit of $50\,\mathrm{km\,h^{-1}}$) passes a stationary police car. The police car accelerates after the other car at $4.0\,\mathrm{m\,s^{-2}}$ for 8.0 s and then continues with the same velocity until it overtakes the other car.

a When did the two cars have the same velocity?

b Determine if the police car has overtaken the other car after 10 s.

c By equating two equations for the same distance at the same time, determine exactly when the police car overtakes the other car.

25 A car brakes suddenly and stops 2.4 s later, after travelling a distance of 38 m.

a Calculate its deceleration.

b What was the velocity of the car before braking?

26 A spacecraft travelling at $8.00\,\mathrm{km\,s^{-1}}$ accelerates at $2.00 \times 10^{-3}\,\mathrm{m\,s^{-2}}$ for 100 hours.

a How far does it travel during this acceleration?

b What is its final velocity?

27 Combine the first two equations of motion (given on page 17) to derive the second two equations:

$$v^2 = u^2 + 2as$$

$$s = ut + \frac{1}{2}at^2$$

Acceleration due to gravity

The motions of objects through the air are common events and deserve special attention.

At the start, we will consider only objects that are moving vertically up, or down, under the effects of gravity only. That is, we will assume (to begin with) that **air resistance** has no significant effect.

◆ **Air resistance** Resistive force opposing the motion of an object through air. A type of drag force.

When an object held up in the air is released from rest, it will accelerate downwards because of the force of gravity. Figure A1.36 shows a possible experimental arrangement that could be used to determine a value for this acceleration.

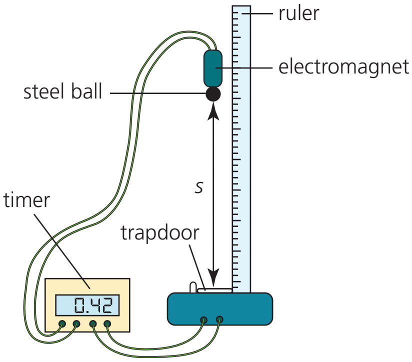


Figure A1.36 An experiment to measure the acceleration due to gravity

Inquiry 2: Collecting and processing data

Collecting data

Figure A1.36 shows how the time for a steel ball to fall a certain distance can be determined experimentally.

Describe how this apparatus can be used to collect and record sufficient, relevant quantitative data which will enable an accurate value for the acceleration of free fall to be determined from a suitable graph.

In the absence of air resistance, all objects (close to the Earth's surface) fall towards the Earth with the same acceleration, $\boldsymbol{g = 9.8\,\mathrm{m\,s^{-2}}}$

g is known as the **acceleration of free fall** due to gravity (sometimes called acceleration due to **free fall**).

g is not a true constant. Its value varies very slightly at different locations around the world. Although, to 2 significant figures (9.8) it has the same value everywhere on the Earth's surface. A convenient value of $g = 10\,\mathrm{m\,s^{-2}}$ is commonly used in introductory physics courses.

The acceleration of free fall (g) reduces with distance from the Earth. (For example, at a height of 100 km above the Earth's surface the value of g is $9.5\,\mathrm{m\,s^{-2}}$.) We will return to this subject in Topic D.1.

◆ **Acceleration due to gravity, *g*** Acceleration of a mass falling freely towards Earth. On, or near the Earth's surface, $g = 9.8 \approx 10\,\mathrm{m\,s^{-2}}$. Also called **acceleration of free fall**.

◆ **Free fall** Motion through the air under the effects of gravity but without air resistance.

◆ **Negligible** Too small to be significant.

WORKED EXAMPLE A1.7

A ball is dropped vertically from a height of 18.3 m. Assuming that the acceleration of free fall is $9.81\,\mathrm{m\,s^{-2}}$ and air resistance is **negligible**, calculate:

a its velocity after 1.70 s

b its height after 1.70 s

c its velocity when it hits the ground

d the time for the ball to reach the ground.

Answer

a $v = u + at = 0 + (9.81 \times 1.70) = 16.7\,\mathrm{m\,s^{-1}}$

b $s = ut + \frac{1}{2}at^2 = 0 + \left(\frac{1}{2} \times 9.81 \times 1.70^2\right) = 14.2\,\mathrm{m}$

So, height above ground = (18.3 – 14.2) = 4.1 m

c $v^2 = u^2 + 2as = 0^2 + (2 \times 9.81 \times 18.3) = 359$

$v = 18.9\,\mathrm{m\,s^{-1}}$

d $v = u + at$

$18.9 = 0 + (9.81 \times t)$

$t = 1.93\,\mathrm{s}$

Tool 3: Mathematics

Appreciate when some effects can be neglected and why this is useful

When studying physics, you may be advised to make assumptions when answering numerical questions. For example: 'assume that air resistance is **negligible** / is insignificant'. It is possible that this is a true statement, for example, air resistance will have no noticeable effect on a solid rubber ball falling 50 cm to the ground. However, the usual reason for advising you to ignore an effect is to make the calculation simpler, and not go beyond what is required in your course.

Calculating the time for a table-tennis ball dropped 50 cm to the ground will result in an underestimate if air resistance is ignored, but the answer can be interpreted as a lower limit to the time taken, and you may be questioned on your understanding of that.

Other examples will be found in all topics. Examples include: assuming friction between surfaces is negligible (Topic A.2); assuming thermal energy losses are negligible (Topic B.1); assuming the internal resistance of a battery is negligible (Topic B.5).

Moving up and down

If gravity is the only force acting, all objects close to the Earth's surface have the same acceleration (9.8 m s^{-2} downwards), whatever their mass and whether they are moving down, moving up or moving sideways.

The velocity of an object moving freely vertically downwards will increase by 9.8 m s^{-1} every second. The velocity of an object moving freely vertically upwards will decrease by 9.8 m s^{-1} every second.

Top tip!

Displacement, velocity and acceleration are all vector quantities and the signs used for motions up and down can be confusing.

If displacement measured up from the ground is considered to be positive, then the acceleration due to gravity is always negative. Velocity upwards is positive, while velocity downwards is negative.

If displacement measured down from the highest point is considered to be positive, then the acceleration due to gravity is always positive. Velocity upwards is negative, while velocity downwards is positive.

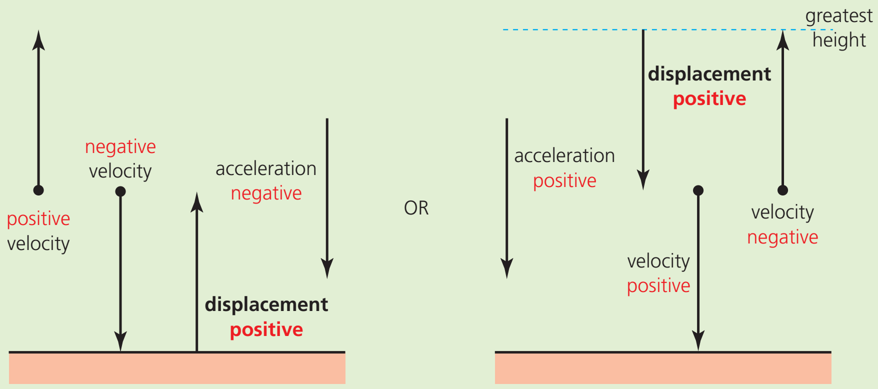


Figure A1.37 Directions of vectors

WORKED EXAMPLE A1.8

A ball is thrown vertically upwards and reaches a maximum height of 21.4 m. For the following questions, assume that $g = 9.81$ m s^{-2}.

a Calculate the speed with which the ball was released.

b State any assumption that you made in answering **a**.

c Determine where the ball will be 3.05 s after it was released.

d Calculate its velocity at this time.

Answer

a $v^2 = u^2 + 2as$

$0^2 = u^2 + (2 \times [-9.81] \times 21.4)$

$u^2 = 419.9$

$u = 20.5\ \mathrm{m\,s^{-1}}$

In this example, the vector quantities directed upwards (u, v, s) are considered positive and the quantity directed downwards (a) is negative. The same answer would be obtained by reversing all the signs.

b It was assumed that there was no air resistance.

c $s = ut + \frac{1}{2}at^2 = (20.5 \times 3.05) + \left(\frac{1}{2} \times [-9.81] \times 3.05^2\right)$

$s = +16.9$ m (above the ground)

d $v = u + at = 20.5 + (-9.81 \times 3.05)$

$= -9.42\ \mathrm{m\,s^{-1}}$ (moving downwards)

In the following questions, ignore the possible effects of air resistance.

Use $g = 9.81\,\mathrm{m\,s^{-2}}$.

28 Discuss possible reasons why the acceleration due to gravity is not exactly the same everywhere on or near the Earth's surface.

29 **a** How long does it take a stone dropped from rest from a height of 2.1 m to reach the ground?

b If the stone was thrown downwards with an initial velocity of $4.4\,\mathrm{m\,s^{-1}}$, calculate the speed with which it hits the ground.

c If the stone was thrown vertically upwards with an initial velocity of $4.4\,\mathrm{m\,s^{-1}}$, with what speed would it hit the ground?

30 A small rock is thrown vertically upwards with an initial velocity of $22\,\mathrm{m\,s^{-1}}$.

a Calculate when its velocity will be $10\,\mathrm{m\,s^{-1}}$.

b Explain why there are two possible answers to **a**.

31 A falling ball has a velocity of $12.7\,\mathrm{m\,s^{-1}}$ as it passes a window 4.81 m above the ground.
Predict when the ball will hit the ground.

32 A ball is thrown vertically upwards with a velocity of $18.5\,\mathrm{m\,s^{-1}}$ from a window that is 12.5 m above the ground.

a Determine when it will pass the same point moving down.

b With what velocity will it hit the ground?

c Calculate how far above the ground the ball was after exactly 2.00 s.

33 Two balls are dropped from rest from the same height. If the second ball is released 0.750 s after the first, and assuming they do not hit the ground, calculate the distance between the balls:

a 3.00 s after the second ball was dropped

b 2.00 s later.

34 A stone is dropped from rest from a height of 34 m. Another stone is thrown downwards 0.5 s later.
If they both hit the ground at the same time, show that the second stone was thrown with a velocity of $5.5\,\mathrm{m\,s^{-1}}$.

Projectile motion

SYLLABUS CONTENT

- The behaviour of projectiles in the absence of fluid resistance, and the application of the equations of motion resolved into vertical and horizontal components.
- The qualitative effect of fluid resistance on projectiles, including time of flight, trajectory, velocity, acceleration, range and terminal speed.

◆ **Projectile** An object that has been projected through the air and which then moves only under the action of the forces of gravity and air resistance.

◆ **Resolve (a vector)** To express a single vector as components (usually two components which are perpendicular to each other).

In our discussion of objects moving through the air, we have so far only considered motion vertically up or down. Now we will extend that work to cover objects moving in any direction. A **projectile** is an object that has been projected through the air (for example: fired, launched, thrown, kicked or hit) and which then moves only under the action of the force of gravity (and air resistance, if significant). A projectile has no ability to power or control its own motion.

Tool 3: Mathematics

Resolve vectors

This process occurs in several places during the course, but the most prominent examples are **resolving** velocities (as below) and forces.

Components of a projectile's velocity

The instantaneous velocity of a projectile at any time can conveniently be resolved into vertical and horizontal components, v_V and v_H, as shown in Figure A1.38.

Common mistake

When using these equations make sure that the angle θ is the angle between the velocity and the horizontal.

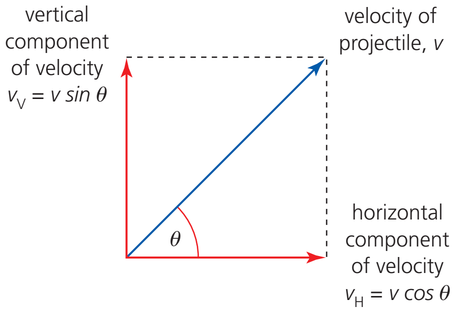


Figure A1.38 Vertical and horizontal components of velocity

Vertical and horizontal components of velocity, v:

$$\boldsymbol{v_V = v \sin\theta}$$

$$\boldsymbol{v_H = v \cos\theta}$$

WORKED EXAMPLE A1.9

A tennis player strikes the ball so that it leaves the racket with a velocity of $64.0\,\mathrm{m\,s^{-1}}$ at an angle of 6.0° below the horizontal. Calculate the vertical and horizontal components of this velocity.

Figure A1.39 A tennis player serving a ball

Answer

$v_H = v \cos\theta = 64.0 \times \cos 6.0 = 64\,\mathrm{m\,s^{-1}}$ (63.649... seen on calculator display)

$v_V = v \sin\theta = 64.0 \times \sin 6.0 = 6.7\,\mathrm{m\,s^{-1}}$ downwards

◆ **Stroboscope** Apparatus used for observing rapid motions. It produces regular flashes of light at an appropriate frequency chosen by the user.

◆ **Trajectory** Path followed by a projectile.

◆ **Parabolic** In the shape of a parabola. The trajectory of a projectile is parabolic in a gravitational field if air resistance is negligible.

◆ **Range (of a projectile)** Horizontal distance travelled before impact with the ground.

Components perpendicular to each other can be analysed separately

The vertical and horizontal components of velocity can be treated separately (independently) in calculations.

- Earlier in this topic, we stated that any object (close to the Earth's surface) which is affected only by gravity (no air resistance) will accelerate towards the Earth with an acceleration of $9.8\,\mathrm{m\,s^{-2}}$. This remains true even if the object is projected sideways (so that its velocity has a horizontal component).
- If there is no air resistance, the horizontal component of a projectile's velocity will remain constant (until it comes into contact with something else).

Figure A1.40 Parabolic trajectory of a bouncing ball

Figure A1.40 shows a **stroboscopic** picture of a bouncing ball. The time intervals between each image of the ball are all the same.

The horizontal separations of successive images of the ball are all the same because the horizontal component of velocity is constant. The vertical separations of successive images of the ball increase as the ball accelerates as it falls, and the separations decrease as the ball decelerates as it moves upwards after bouncing on the ground.

The path followed by a projectile (as seen in Figure A1.40) is called its **trajectory**. The typical shape of a freely moving projectile is **parabolic**. The horizontal distance covered is called the **range** of the projectile.

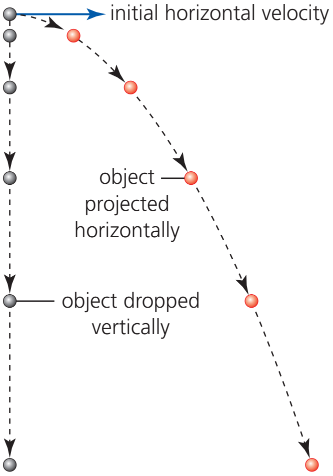


Figure A1.41 The parabolic trajectory of an object projected horizontally compared with an object dropped vertically

Figure A1.41 compares the trajectory of an object dropped vertically to the trajectory of an object projected horizontally at the same time. Note that both objects fall equal distances in the same time. This is true whatever the horizontal component of velocity (assuming negligible air resistance)

WORKED EXAMPLE A1.10

Object projected horizontally

A bullet was fired horizontally with a speed of $524\,\mathrm{m\,s^{-1}}$ from a height of 22.0 m above the ground. Calculate where it hit the ground. Assume that air resistance was negligible.

Answer

First, we need to calculate how long the bullet is in the air. We can do this by finding the time that the same bullet would have taken to fall to the ground if it had been dropped vertically from rest (so $u = 0$):

$$s = ut + \frac{1}{2}at^2$$

$$22.0 = 0 + (0.5 \times 9.81 \times t^2)$$

$$t = 2.12\,\mathrm{s}$$

Without air resistance the bullet will continue to travel with the same horizontal component of velocity ($524\,\mathrm{m\,s^{-1}}$) until it hits the ground 2.12 s later. Therefore:

horizontal distance travelled = horizontal velocity × time

horizontal distance = $524 \times 2.12 = 1.11 \times 10^3\,\mathrm{m}$ (1.11 km)

ATL A1A: Thinking skills

Providing a reasoned argument to support conclusions

Figure A1.42 shows an experimental arrangement in which a steel ball can be projected horizontally from a table top.

Sketch a graph to show the pattern of results that you would expect to see when the range *x* was measured for different heights, *h*. Explain your reasoning.

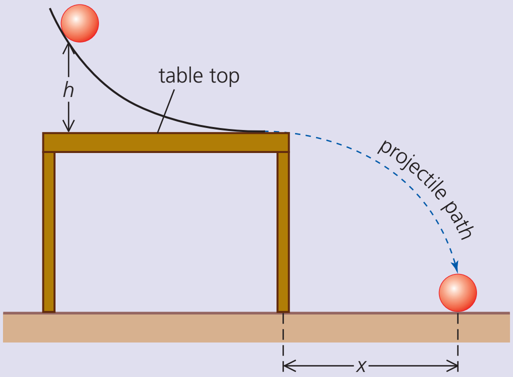


Figure A1.42 Investigating range, *x*, travelled by a projectile

WORKED EXAMPLE A1.11

Object projected at an angle to the horizontal

A stone was thrown upwards from a height 1.60 m above the ground with a speed of $18.0\,\mathrm{m\,s^{-1}}$ at an angle of 52.0° to the horizontal. Assuming that air resistance is negligible, calculate:

a its maximum height

b the vertical component of velocity when it hits the ground

c the time taken to reach the ground

d the horizontal distance to the point where it hits the ground

e the velocity of **impact**.

◆ **Impact** Collision involving relatively large forces over a short time.

Top tip!

If we know the velocity and position of a projectile, we can always use its vertical component of velocity to determine:

- the time taken before it reaches its maximum height, and the time before it hits the ground
- the maximum height reached (assuming its velocity has an upwards component).

The horizontal component can then be used to determine the range.

Answer

First, we need to know the two components of the initial velocity:

$v_V = v \sin\theta = 18.0 \sin 52.0° = 14.2\,\text{m s}^{-1}$

$v_H = v \cos\theta = 18.0 \cos 52.0° = 11.1\,\text{m s}^{-1}$

a Using $v^2 = u^2 + 2as$ for the upwards vertical motion (with directions upwards considered to be positive), and remembering that at the maximum height $v = 0$, we get:

$0 = 14.2^2 + [2 \times (-9.81) \times s]$

$s = +10.3\,\text{m}$ above the point from which it was released; a total height of 11.9 m.

b Using $v^2 = u^2 + 2as$ for the complete motion gives:

$v^2 = 14.2^2 + [2 \times (-9.81) \times (-1.60)]$

$v = 15.27 = 15.3\,\text{m s}^{-1}$ downwards

c Using $v = u + at$ gives:

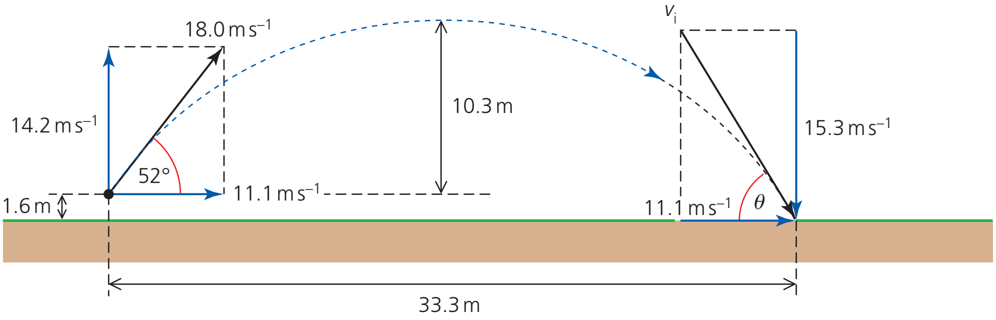
$-15.27 = 14.2 + (-9.81)t$

$t = 3.00\,\text{s}$

d Using $s = vt$ with the horizontal component of velocity gives:

$s = 11.1 \times 3.00 = 33.3\,\text{m}$

e Figure A1.43 illustrates the information we have so far, and the unknown angle, θ, and velocity, v_i.



■ **Figure A1.43** Object projected at an angle to the horizontal

From looking at the diagram (Figure A1.43), we can use Pythagoras's theorem to calculate the velocity of impact.

$(\text{velocity of impact})^2 = (\text{horizontal component})^2 + (\text{vertical component})^2$

$v_i^2 = 11.1^2 + 15.3^2$

$v_i = 18.9\ \text{m s}^{-1}$

The angle of impact with the horizontal, θ, can be found using trigonometry:

$$\tan\theta = \frac{15.3}{11.1}$$

$\theta = 54.0°$

◆ **Imagination** Formation of new ideas that are not related to direct sense perception or experimental results.

◆ **Intuition** Immediate understanding, without reasoning.

◆ **Inspiration** Stimulation (usually to be creative).

TOK

The natural sciences

- What is the role of **imagination** and **intuition** in the creation of hypotheses in the natural sciences?

The independence of horizontal and vertical motion in projectile motion may seem unexpected and counterintuitive. It requires imagination (some would say genius) to propose ideas and theories which are contrary to accepted wisdom and 'common sense'. This is especially true in understanding the worlds of relativity and quantum physics, where relying on everyday experiences for **inspiration** is of little or no use.

It is worth remembering that many of the well-established concepts and theories of classical physics that are taught now in introductory physics lessons would have seemed improbable to many people at the time they were first proposed. For example, many people would say (incorrectly) that a force is needed to keep an object moving at constant speed (see Topic A.2).

Fluid resistance and terminal speed

So far, we have only considered projectile motion in which air resistance is negligible. We will now broaden the discussion.

◆ **Drag** Force(s) opposing motion through a fluid; sometimes called fluid resistance.

◆ **Fluid** Liquid or gas.

◆ **Fluid resistance (friction)** Force(s) opposing motion through a fluid; sometimes called drag.

As any object moves through air, the air is forced to move out of the path of the object. This causes a force opposing the motion called air resistance, also known as **drag**. Drag forces will oppose the motion of an object moving in any direction through any gas or liquid. (Gases and liquids are both described as **fluids** because they can flow.) Such forces opposing motion are generally described as **fluid resistance**.

Figure A1.44 gives a visual impression of air resistance. It shows the movement of air (marked by streamers) past a model of a car. (The picture was taken in a wind tunnel, in which moving air was directed towards the vehicle.)

■ **Figure A1.44** Air flow over a clay aerodynamic model of a high-performance sports vehicle

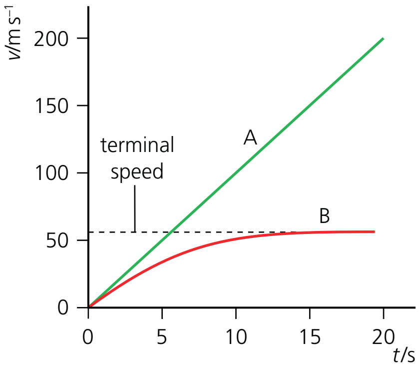


Figure A1.45 An example of a graph of velocity against time for an object falling under the effect of gravity, with (B) and without (A) air resistance

Figure A1.45 represents the motion of an object falling towards Earth.

Line A shows the motion without air resistance and with a constant acceleration of 9.8 m s^{-2} (≈ 10). Line B shows the motion more realistically, with air resistance.

When any object first starts to fall, there is no air resistance. As the object falls faster, the air resistance increases, so that the rate of increase in velocity becomes less. This is shown in the Figure A1.45 by line B becoming less steep. Eventually the object reaches a constant, maximum speed known as the **terminal speed** or **terminal velocity** ('terminal' means final).

Objects falling through fluids (such as air) have a maximum speed, called terminal speed, which occurs when their acceleration has reduced to zero because of increasing fluid resistance (as their velocity increases).

◆ **Terminal speed (velocity)** The greatest downwards speed of a falling object that is experiencing resistive forces (for example, air resistance). It occurs when the object's weight is equal to the sum of resistive forces (+ upthrust).

Top tip!

The concept of a top (terminal) speed can also be applied to the horizontal motion of vehicles, like trains, cars and aircraft. As they travel faster, increasing air resistance reduces their acceleration to zero.

The value of an object's terminal speed will depend on its cross-sectional area, shape and weight, as discussed in Topic A.2. The terminal speed of skydivers (Figure A1.46) is usually quoted at about 200 km h^{-1} (56 m s^{-1}).

Terminal speed also depends on the density of the air. In October 2012 Felix Baumgartner (Figure A1.47), an Austrian skydiver, reached a world record speed of 1358 km h^{-1} by starting his jump from a height of about 39 km above the Earth's surface, where the density of air is about 250 times less than near the Earth's surface. In 2014 Alan Eustace completed a jump from greater altitude, but at 1323 km h^{-1} he did not break Baumgartner's speed record.

Figure A1.46 Skydivers at their terminal speed

Figure A1.47 Felix Baumgartner about to jump from a height of 39 km

Effect of fluid (air) resistance on projectiles

Without air resistance we assume that the horizontal component of a projectile's velocity is constant, but with air resistance it decreases. Without air resistance the vertical motion always has a downwards acceleration of 9.8 m s^{-2}, but with air resistance the acceleration will be reduced for falling objects and the deceleration increased for objects moving upwards.

Figure A1.48 shows typical trajectories with and without air resistance (for the same initial velocity).

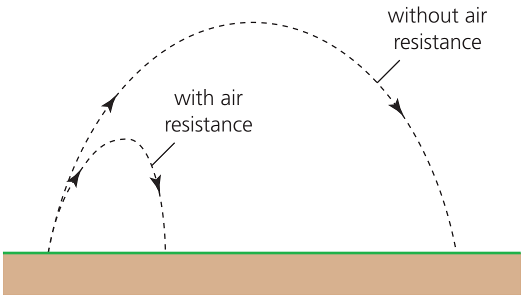


Figure A1.48 Effect of air resistance on the trajectory of a projectile

Air resistance reduces the range of a projectile and its trajectory will not be parabolic.

Tool 2: Technology

Carry out image analysis and video analysis of motion

Video-capture technology is used in sports, such as tennis and soccer. Capturing the trajectory of a projectile on video allows us to analyse its motion frame-by-frame. For example, the cameras used in VAR in football usually capture 50 frames per second, so the motion of the projectile (the ball) can be observed at time intervals of 0.02 s.

Explain how you could use **video analysis** of motion to investigate the motion of a shuttlecock in a game of badminton.

◆ **Video analysis** Analysis of motion by freeze-frame or slow-motion video replay.

Figure A1.49 Consider how video analysis could be used to investigate the motion of a badminton shuttlecock.

In the following questions, ignore the possible effects of air resistance. Use $g = 9.81\,\text{m}\,\text{s}^{-2}$.

35 At an indoor rifle range, a bullet was fired horizontally at the centre of a target 36 m away. If the speed of the bullet was $310\,\text{m}\,\text{s}^{-1}$, predict where the bullet will strike the target.

36 Repeat Worked example A1.11 for a stone thrown with a velocity of $26\,\text{m}\,\text{s}^{-1}$ at an angle of 38° to the horizontal from a cliff top. The point of release was 33 m vertically above the sea.

37 It can be shown that the maximum theoretical range of a projectile occurs when it is projected at an angle of 45° to the ground (once again, ignoring the effects of air resistance). Calculate the maximum distance a golf ball will travel before hitting the ground if its initial velocity is $72\,\text{m}\,\text{s}^{-1}$. (Because you need to assume that there is no air resistance, your answer should be much higher than the actual ranges achieved by top-class golfers. Research to determine the actual ranges achieved in competition golf.)

38 A jet of water from a hose is aimed directly at the base of a flower, as shown in Figure A1.50. The water emerges from the hose with a speed of $3.8\,\text{m}\,\text{s}^{-1}$.

a Calculate the vertical and horizontal components of the initial velocity of the water.

b How far away from the base of the plant does the water hit the ground?

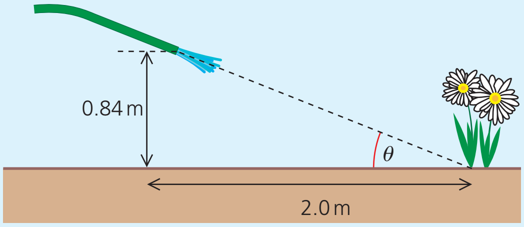


Figure A1.50 Water from a hose aimed at the base of a flower

39 If the maximum distance a man can throw a ball is 78 m, what is the minimum speed of release of the ball? (Assume that the ball lands at the same height from which it was thrown and that the greatest range for a given speed is when the angle is 45°.)

40 Figure A1.51 shows a player making a basketball shot.

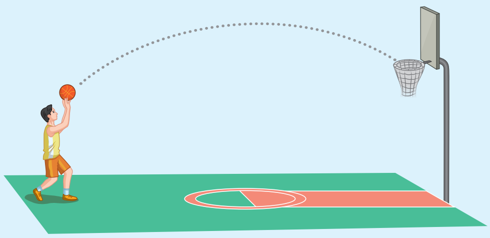


Figure A1.51 Basketball player making a shot

a In practice, air resistance can be considered negligible for a basketball. Suggest a reason why.

b Make a copy of the figure and add to it two other possible trajectories which will result in the ball arriving at the basket.

c Suggest which trajectory is best and explain your reasoning.

d Add to your drawing a possible trajectory that would enable a light-weight sponge ball to reach the basket.

Nature of science: Models

The motions of all projectiles are affected – often considerably – by air resistance. But the mathematics we have used to make predictions has assumed that air resistance is negligible. This is a recurring theme in physics: when theories are first developed, or when you are first introduced to a topic, the ideas are simplified. A 'complete' understanding of projectile motion may be expected at university level, but the topic is important enough that you should be introduced to the basic ideas at an earlier age.

In Worked example A1.10, the calculated answer predicts that a bullet will travel 1.1 km before striking the ground, although we should stress that this 'assumes that there is no air resistance'. In reality, it should be well understood that air resistance cannot be ignored, and the bullet will not travel as far as calculated. This should not suggest that the calculation was not useful.

As your knowledge and experience increase, mathematical theories of projectile motion can be expanded to include the effects of air resistance – but this is beyond the limits of the IB Course. Similar comments can be applied to all areas of physics. This simplifying approach to gaining knowledge is not unique to physics but it is, perhaps, most obvious in the sciences.

LINKING QUESTION

- How does the motion of an object change within a gravitational field?

This question links to understandings in Topics A.3 and D.1.

Ballistics

The study of the use of projectiles is known as ballistics. Because of its close links to hunting and fighting, this is an area of science with a long history, going all the way back to spears, and bows and arrows. Figure A1.52 shows a common medieval misconception about the motion of cannon balls: they were thought to travel straight until they ran out of energy.

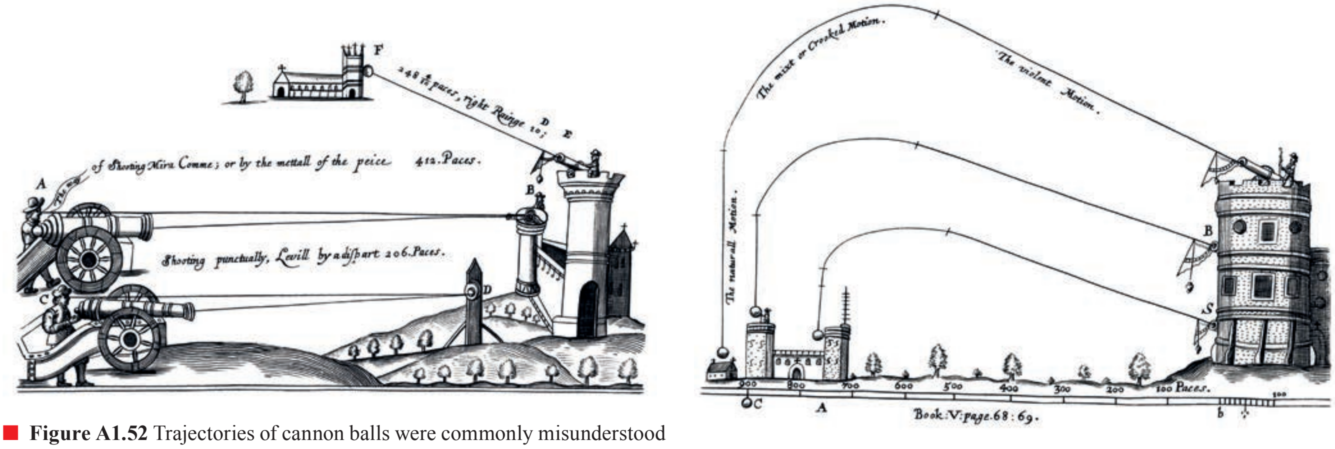


Figure A1.52 Trajectories of cannon balls were commonly misunderstood

Photographs taken in quick succession became useful in analysing many types of motion in the nineteenth century, but the trajectories of very rapidly moving projectiles were difficult to determine until they could be filmed or illuminated by lights flashing very quickly (stroboscopes). The photograph of the bullet from a gun shown in Figure A1.53 required high technology, such as a very high-speed flash and very sensitive image recorders, in order to 'freeze' the projectile (bullet) in its rapid motion (more than 500 $m s^{-1}$).

Figure A1.53 A bullet 'frozen' by high-speed photography

◆ **Thought experiment** An experiment that is carried out in the mind, rather than actually being done, normally because it is otherwise impossible.

'Newton's cannonball' is a famous **thought experiment** concerning projectiles, in which Newton imagined what would happen to a cannonball fired (projected) horizontally at various very high speeds from the top of a very high mountain (in the absence of air resistance). See Figure A1.54.

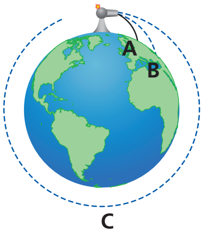


Figure A1.54 Newton's cannonball thought experiment

The balls labelled A and B will follow parabolic paths to the Earth's surface. B has a greater range than A because it was fired with greater velocity. Cannonball C has exactly the correct velocity that it never falls back to the Earth's surface and never moves further away from the Earth. (The required velocity would be about 7 $km s^{-1}$, but remember that we are assuming that there is no air resistance.) These ideas are developed further in Topic D.1.

Nature of science: Models

In a thought experiment, we use our imagination to answer scientific 'what if…?' type questions. Known principles or a possible theory are applied to a precise scenario, and the consequences thought through in detail. Usually, but not always, it would not be possible to actually carry out the experiment.

At the time of 'Newton's cannonball' thought experiment (published in 1728) it would have been impossible to make any object move at 7 $km s^{-1}$ and, even if that had been possible, air resistance would have quickly reduced its speed. Nevertheless, the thought processes involved advanced understanding and led to ideas of satellite motion. The first satellite to orbit the Earth was the Russian Sputnik 1 in 1957, which had a maximum speed of about 8 $km s^{-1}$ and avoided air resistance by being above most of the Earth's atmosphere.

Another (possible) thought experiment connected to this topic, and involving an assumption of no air resistance, is the dropping of two spheres of different masses from the same height on the Tower of Pisa. See Figure A1.55. Most historians doubt if there was an actual experiment at the Tower of Pisa that confirmed Galileo's theory that both masses would fall at the same rate.

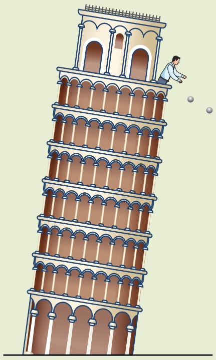


Figure A1.55 Galileo's famous experiment to demonstrate acceleration due to gravity

Two further famous thought experiments in physics are *Maxwell's demon* and *Schrödinger's cat*. Research online to find out how these thought experiments prompted new hypotheses and theories in physics.

A.2 Forces and momentum

Guiding questions

- How can we represent the forces acting on a system both visually and algebraically?
- How can Newton's laws be modelled mathematically?
- How can knowledge of forces and momentum be used to predict the behaviour of interacting bodies?

The nature of force

◆ **Interaction** Any event in which two or more objects exert forces on each other.

◆ **Newton, N** Derived SI unit of force. $1\,\text{N} = 1\,\text{kg}\,\text{m}\,\text{s}^{-2}$.

SYLLABUS CONTENT

- Forces as interactions between bodies.

In everyday life we may describe a force as a push or a pull but, more generally, a force can be considered to be *any* type of **interaction** / influence on an object which will tend to make it start moving or change its motion if it is already moving (assuming that the force is unopposed). Many forces do not cause changes of motion, simply because the objects on which they are acting are not able to move freely. Forces also change the shapes of objects.

Scientists refer to forces 'acting' on objects, 'exerting' forces on objects and 'applying' forces to objects. If objects 'interact', this means there are forces between them.

The size of a force is measured in the SI unit **newton**, **N**. The direction in which a force acts on an object is important:

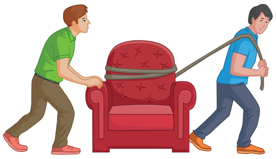


Figure A2.1 Pushing and pulling a chair

Forces, F, are vector quantities and are represented in drawings by arrows of scaled length, direction and point of application. All forces should be labelled with commonly accepted symbols, or names.

(The vectors displacement, velocity and acceleration were introduced in Topic A.1.)

Most situations, such as the two boys moving a chair in Figure A2.1, involve several forces, not just the obvious forces arising from the boys' actions.

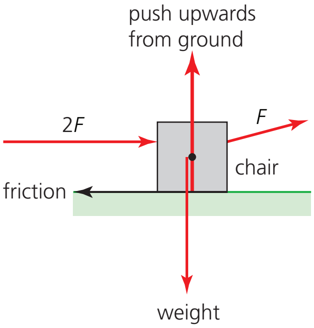


Figure A2.2 Representing the forces in Figure A2.1

Figure A2.2 shows all the forces acting on the chair. These include the weight of the chair, the friction opposing its movement and the push upwards from the floor which is supporting the chair. The boy on the left is pushing the chair with a force which is twice the size of the force, F, that the boy on the right is using.

We will return to force diagrams later, but first we need to identify and explain different types of force.

Different types of force

In general, we can classify all forces as one of two kinds.

- Forces that involve physical contact. Examples include everyday pushes and pulls, friction and air resistance.

- Forces that act 'at a distance' across empty space. Examples include magnetic forces and the force of gravity. These forces are more difficult to understand and can be described as 'field forces'.

We will now explore some important types of force in greater detail.

◆ **Mass** A measure of an object's resistance to a change of motion (inertia).

◆ **Kilogramme, kg** SI unit of mass (fundamental).

◆ **Weight, F_g** Gravitational force acting on a mass. $F_g = mg$.

◆ **Gravitational field strength, g** The gravitational force per unit mass (that would be experienced by a small test mass placed at that point).

$g = F_g/m$ (SI unit: $N\,kg^{-1}$).

Numerically equal to the acceleration due to gravity.

Weight

SYLLABUS CONTENT

▶ Gravitational force F_g as the weight of the body and calculated as given by: $F_g = mg$

The **mass** of an object may be considered to be a measure of the quantity of matter it contains. Mass has the SI unit **kilogramme**, kg. Mass does not change with location. This definition may seem rather vague, but this is because mass is such a fundamental concept it is difficult to explain in terms of other things. However, later in this topic we will provide an improved definition.

The **weight**, F_g, of a mass, m, is the gravitational force that pulls it towards the centre of the Earth (or any other planet). Weight and mass are connected by the simple relationship:

$$\textbf{weight, } F_g = mg$$

Where g is the weight : mass ratio, which is called the **gravitational field strength**. It has the units $N\,kg^{-1}$.

g is numerically equal to the acceleration due to gravity (see Topic A.1). An explanation is given later in this topic.

Clearly, in principle, the weight of an object is not constant, but varies with location (where the value of g changes). The value of g varies with a planet's or a moon's mass and radius, and with distance from the planet's centre of mass. For example, it has a value of $9.8\,N\,kg^{-1}$ on the Earth's surface, $1.6\,N\,kg^{-1}$ on the surface of the Moon and $3.7\,N\,kg^{-1}$ on Mars.

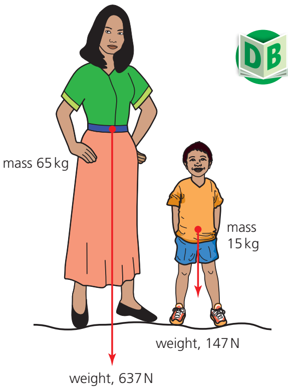


Figure A2.3 Weight acts downwards from the centre of mass

Weight is represented in a diagram by a vector arrow vertically downwards from the **centre of mass** of the object. See Figure A2.3. When an object is subjected to forces, it will behave as if all of its mass was at a single point: its centre of mass. (In a gravitational field, the same point is sometimes called its 'centre of gravity'.)

◆ **Centre of mass** Average position of all the mass of an object. The mass of an object is distributed evenly either side of any plane through its centre of mass.

WORKED EXAMPLE A2.1

An astronaut has a mass of 58.6 kg. Calculate her weight using data from the preceding paragraphs:

a on the Earth's surface

b in a satellite 250 km above the surface ($g = 9.1\,N\,kg^{-1}$)

c on the surface of the Moon

d on the surface of Mars

e in 'deep space', a very long way from any planet or star.

Answer

a $F_g = mg = 58.6 \times 9.8 = 5.7 \times 10^2\,N$

b $F_g = mg = 58.6 \times 9.1 = 5.3 \times 10^2\,N$, which is only 7% lower than on the Earth's surface

c $F_g = mg = 58.6 \times 1.6 = 94\,N$

d $F_g = mg = 58.6 \times 3.7 = 217\,N$

e 0 N, truly weightless

1 Calculate the weight of the following objects on the surface of the Earth:

- **a** a car of mass 1250 kg
- **b** a new-born baby of mass 3240 g
- **c** one pin in a pile of 500 pins that has a total mass of 124 g.

2 It is said that 'an A380 airplane has a maximum take-off weight of 570 tonnes' (Figure A2.4). A tonne is the name given to a mass of 1000 kg.

- **a** What is the maximum weight of the aircraft (in newtons) during take off?
- **b** The aircraft can take a maximum of about 850 passengers. Estimate the total mass of all the passengers and crew.
- **c** What percentage is this of the total mass of the airplane on take off?
- **d** The maximum landing weight is '390 tonnes'. Suggest a reason why the aircraft needs to be less massive when landing than when taking off.
- **e** Calculate the difference in mass and explain where the 'missing' mass has gone.

Figure A2.4 The Airbus A380 is the largest passenger airplane in the world

3 A mass of 50 kg would have a weight of 445 N on the planet Venus. What is the strength of the gravitational field there? Compare it with the value of *g* on Earth.

4 Consider two solid spheres made of the same metal. Sphere A has twice the radius of sphere B. Calculate the ratio of the two spheres' circumferences, surface areas, volumes, masses and weights.

◆ **Force meter** Instrument used to measure forces. Also sometimes called a newton meter or a spring balance.

◆ **Calibrate** Put numbered divisions on a scale.

◆ **Weigh** Determine the weight of an object. In everyday use the word 'weighing' usually means quoting the result as the equivalent mass: 'my weight is 60 kg' actually means I have the weight of a 60 kg mass (about 590 N).

Tool 1: Experimental techniques

Understand how to accurately measure quantities to an appropriate level of precision: force, weight and mass

Forces are easily measured by the changes in length they produce when they squash or stretch a spring (or something similar). Such instruments are called **force meters** (also called *newton meters* or *spring balances*) – see Figure A2.5. In this type of instrument, the spring usually has a change of length proportional to the applied force. The length of the spring is shown on a linear scale, which can be **calibrated** (marked in newtons). The spring goes back to its original shape after it has measured the force.

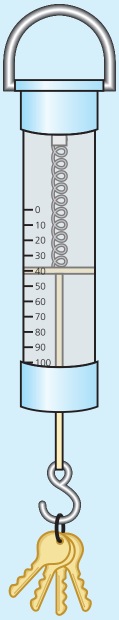


Figure A2.5 A spring balance force meter

Such instruments can be used for measuring forces acting in any direction, but they are also widely used for the measurement of the downwards force of weight. The other common way of measuring weight is with some kind of '*balance*' (scales). In an equal-arm balance, as shown in Figure A2.6, the beam will only balance if the two weights are equal. That is, the unknown weight equals the known weight. (Larger weights can be measured by positioning the pivot closer to the unknown weight and using the 'principle of moments' – mentioned in Topic A.4.)

Either of these methods can be used to determine (**weigh**) an unknown weight (N) and they rely on the force of gravity to do this, but such instruments may be calibrated to indicate mass (in kg or g) rather than weight. This is because we are usually more concerned with the quantity of something, rather than the effects of gravity on it. We usually assume that:

$$\text{mass (kg)} = \frac{\text{weight (N)}}{9.8}$$

anywhere on Earth because any variations in the acceleration due to gravity, *g*, are insignificant for most, but not all, purposes.

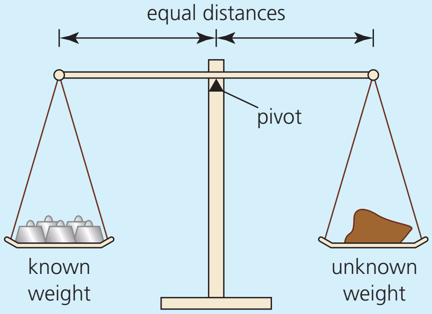


Figure A2.6 An equal-arm balance

Determining a mass without using its weight (gravity) is not so easy. Two ways we can do this are:

- If it is a solid and all the same material, its volume can be measured, then mass = volume × density (assuming that its density is known.)
- As we will see in Topic A.2, resultant force, mass and acceleration are connected by the equation $F = ma$, so that, if the acceleration produced by a known force can be measured, then the mass can be calculated.

Nature of science: Science as a shared endeavour

Science is a collaborative activity – scientists work together across the world to confirm (or dispute) findings by repeating experiments. Scientists review each other's work (**peer review**) to make sure that it is reliable before it is published. Communication is an essential part of science, and precision in communication is very important. Scientists must agree to use specific *terminology*, which is why scientific terminology sometimes differs from everyday use of the same words.

◆ **Peer review** Evaluation of scientific work by experts in the same field of study.

◆ **Tension (force)** Force that tries to stretch an object or material.

◆ **Compression (force)** Force that tries to squash an object or material.

◆ **Deformation** Change of shape.

◆ **Normal** Perpendicular to a surface.

Contact forces

Apart from obvious everyday pushes and pulls, the following terms should be understood:

Tension: pulling forces are acting tending to cause stretching.

Compression: forces are pushing inwards on an object (See Figure A2.7).

Both of these types of force will tend to change the shape of an object (**deformation**).

In the following sections we will discuss the following contact forces in more detail: normal forces, buoyancy forces, elastic restoring forces, surface friction and fluid friction.

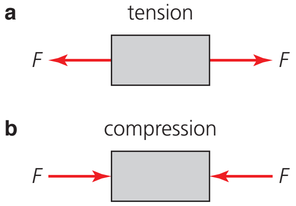


Figure A2.7 Object under **a** tension and **b** compression

Normal forces

SYLLABUS CONTENT

- Normal force F_N is the component of the contact force acting perpendicular to the surface that counteracts the body.

When two objects come in contact, they will exert forces on each other. This is because the particles in the surfaces resist getting closer together. A simple example is a book on a table, as shown in Figure A2.8. The book presses down on the table with its weight, and the table pushes up on the book with an equal force (so that the book is stationary). This force from the table is called a **normal** force, F_N. 'Normal' in this sense means that it is perpendicular to the surface.

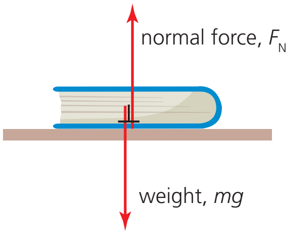


Figure A2.8 Normal force acting upwards on a book

If a force acts on a surface, the surface pushes back. The component of that force which is perpendicular to the surface is called a normal force.

Top tip!

Many students find the idea that solid and hard objects like walls, tables and floors can exert forces, difficult to comprehend, whereas forces from cushions, or trampolines, are easier to visualize and understand. Remember that solid materials will resist any deformation and push back, even if the change of shape is very, very small and not noticeable.

A normal force does not need to be vertical, nor equal to weight, as the two examples in Figure A2.9 illustrate.

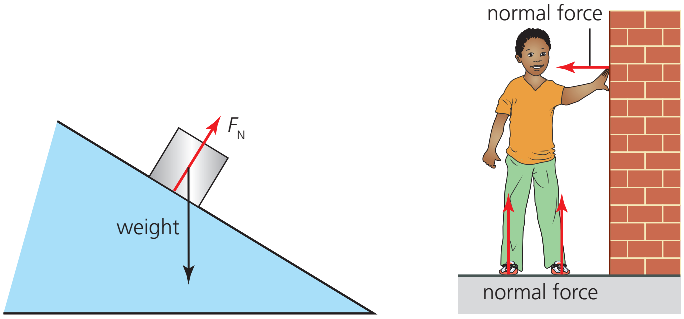


Figure A2.9 Other examples of normal forces

Buoyancy forces

SYLLABUS CONTENT

- Buoyancy force, F_b, acting on a body due to the displacement of the fluid as given by: $F_b = \rho V g$, where V is the volume of fluid displaced.

We have discussed the normal contact forces which act upwards on objects placed on solid horizontal surfaces. Liquids also provide vertical forces upwards on objects placed in, or on them. Gases, too, provide some support, although it is often insignificant.

◆ **Buoyancy force** Vertical upwards force on an object placed in or on a fluid. Sometimes called **upthrust**.

◆ **Density** $\frac{\text{mass}}{\text{volume}}$.

Buoyancy is the ability of any fluid (liquid or gas) to provide a vertical upwards force on an object placed in, or on it. This force is sometimes called **upthrust**. (Buoyancy can be explained by considering the difference in fluid pressures on the upper and lower surfaces of the object. Pressure is explained in Topic B.3.)

The magnitude of an upthrust will be greater in fluids of greater **density**.

Density is a concept with which you may be familiar, although it is not introduced in this course until Topic B.1.

$$\text{density (SI unit: kg m}^{-3}\text{)} = \frac{\text{mass}}{\text{volume}} \qquad \rho = \frac{m}{V}$$

g cm^{-3} is also widely used as a unit for density. A density of $1000\,\text{kg m}^{-3}$ (the density of pure water at $0\,°\text{C}$) is equivalent to $1.000\,\text{g cm}^{-3}$. It is also useful to know that one litre (1l) of water has a volume of $1000\,\text{cm}^3$ and has a mass of $1.00\,\text{kg}$.

Figure A2.10 shows two forces acting on a rock immersed in water. Its weight is greater than the buoyancy force, so it is sinking.

■ **Figure A2.10** Forces on an object immersed in a fluid

◆ **Archimedes' principle** When an object is wholly or partially immersed in a fluid, it experiences buoyancy force equal to the weight of the fluid displaced.

This area of classical physics was first studied more than 2250 years ago in Syracuse, Italy by **Archimedes** (from Greece, who identified the following principle, which still bears his name):

> When an object is wholly or partially immersed in a fluid, it experiences a buoyancy force, F_b, equal to the weight of the fluid displaced. Since weight = mg, and density, $\rho = \frac{m}{V}$:
>
> $$F_b = \rho V g$$

TOK

The natural sciences

- What is the role of imagination in the natural sciences?

Myths, stories and science

The story of Archimedes' discovery of the principle of displacement is well known. The story is that Archimedes was asked by the king of Syracuse, Hiero, to check whether his goldsmith was trying to cheat him by mixing cheaper metals with the gold of a wreath in honour of the gods. Archimedes accepted the challenge, although was uncertain how to establish the true composition of the wreath crown. Reputedly, the idea came to him while sitting in the bath: if the wreath contained other metals, it would be less dense than gold, and as such would need to have a greater volume to achieve the same weight. Archimedes saw that he could test the composition of the wreath by measuring how much water was displaced by it, so measuring its volume and so allowing him to compare its density to that of gold. As the story relates, when Archimedes discovered this he shouted 'I have found it!' or 'Eureka!' in Greek and ran naked through the streets of Syracuse to give Hiero the news!

In fact, this story was never recorded by Archimedes himself and is found in the writings of a Roman architect from much later in the first century BCE called Vitruvius. Many who heard the story doubted it – including Galileo Galilei, who pointed out in his work 'The Little Balance' that Archimedes could have achieved a more precise result using a balance and the law of buoyancy he already knew. But the story persists, perhaps because it is a great way to visualize and so understand the concepts of displacement and density.

■ **Figure A2.11** A statue of Archimedes in a bathtub demonstrates the principle of the buoyant force. Located at Madatech, Israel's National Museum of Science, Technology and Space in Haifa

Consider: In what ways does the story of Archimedes resemble a thought experiment (see Topic A.1)? Do myths and stories serve always to obscure or confuse scientific truths? Can they sometimes enlighten us, too?

WORKED EXAMPLE A2.2

A piece of wood has a volume of 34 cm^3 and a mass of 29 g.

a Calculate its weight.

b Determine the volume of water that it will displace if it is completely under water.

c What buoyancy force will it experience while under water? (Assume density of water = 1000 kg m^{-3}.)

d What resultant force will act on the wood?

e State what will happen to the wood if it is free to move.

f Repeat a–e for the same piece of wood when it is surrounded by air (density 1.3 kg m^{-3}).

Answer

a weight = mg = $(29 \times 10^{-3}) \times 9.8 = 0.28$ N downwards

b 34 cm^3

c Weight of water displaced = $mg = V\rho g = (34 \times 10^{-6}) \times 1000 \times 9.8 = 0.33$ N upwards

d $0.33 - 0.28 = 0.05$ N upwards

e It will move (accelerate) up to the surface, where it will float.

f (see a–e below)

a Weight = 0.28 N downwards, as before

b 34 cm^3 as before

c Weight of air displaced = $mg = V\rho g = (34 \times 10^{-6}) \times 1.3 \times 9.8 = 4.3 \times 10^{-4}$ N upwards. Which is very small!

d $0.28 - (4.3 \times 10^{-4}) \approx 0.28$ N downwards. The buoyancy force in air has an insignificant effect on the wood.

e It will move (accelerate) down towards the Earth.

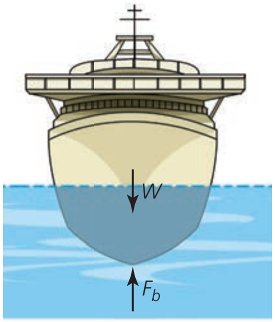


Figure A2.12 A floating object

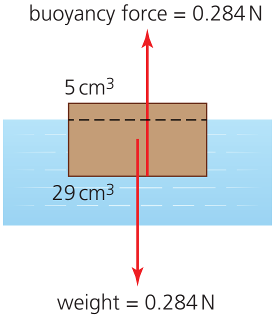


Figure A2.13 Floating wood

Floating

An object placed on the surface of water (or any other liquid) will move lower until it displaces its own weight of water. See Figure A2.12. Then there will be no overall force acting on it, because the buoyancy force upwards (upthrust) will be equal to its weight downwards. If that is not possible, it will sink.

Continuing the numerical Worked example A2.2:

The wood has a weight of 0.28 N, so when floating it will displace water of this weight. Density of pure water = 1000 kg m^{-3}.

Weight = $0.28 = V\rho g = V \times 1000 \times 9.8$

$V = 2.9 \times 10^{-5}$ m^3. That is, 29 cm^3. The wood will float with 29 cm^3 below the water surface and 5 cm^3 above the surface, as shown in Figure A2.13.

5 a Calculate the buoyancy force acting on a boy of mass 60 kg and volume 0.0590 m^3 (use $g = 9.81\,N\,kg^{-1}$)
   i in water of density 1000 $kg\,m^{-3}$
   ii in air of density 1.29 $kg\,m^{-3}$.
  b Will the boy sink or float in water? Explain your answer.
  c Suggest why he would float easily if he was in the Dead Sea. See Figure A2.14.

■ **Figure A2.14** Floating in the Dead Sea

  d Calculate a value for the ratio: boy's weight / buoyancy force in air.

6 A wooden cube with a density of 880 $kg\,m^{-3}$ is floating on water (density 1000 $kg\,m^{-3}$). If the sides of the cube are 5.5 cm long and the cube is floating with a surface parallel to the water's surface, show that the depth of wood below the surface is 4.8 cm.

7 After the rock shown in Figure A2.10 begins to move downwards (sink) another force will act on it. State the name of that force.

8 Outline the reasons why a balloon filled with helium will rise (in air), while a balloon filled with air will fall.

9 Learning to scuba dive involves being able to remain 'neutrally buoyant', so that the diver stays at the same level under water. Explain why breathing in and out affects the buoyancy of a diver.

■ **Figure A2.15** How much of an iceberg is submerged?

10 It is commonly said that about 10% of an iceberg is above the surface of the sea (Figure A2.15). Use this figure to estimate a value for the density of sea ice. Assume the density of sea water is 1025 $kg\,m^{-3}$.

Elastic restoring forces

SYLLABUS CONTENT

▶ Elastic restoring force, F_H, following Hooke's law as given by: $F_H = -kx$, where k is the spring constant.

◆ **Elastic behaviour** A material regains its original shape after a force causing deformation has been removed.

◆ **Elastic limit** The maximum force and/or extension that a material, or spring, can sustain before it becomes permanently deformed.

When a force acts on an object it can change its shape: then we say that there is a deformation. Sometimes the deformation will be obvious, such as when someone sits on a sofa; sometimes the deformation will be too small to be seen, such as when we stand on the floor.

If an object returns to its original shape after the force has been removed, we say that the deformation was **elastic**. We hope and expect that most of the objects we use in everyday life behave elastically, because after we use them, we want them to return to the same condition as before their use. If they do not, we say that they have passed their **elastic limit**.

Common mistake

Rubber bands behave elastically and they are useful because they can stretch a lot and exert inwards forces on the objects they are wrapped around. Because of this behaviour, the word ‘elastic’ in common usage has also come to mean ‘easy to stretch’ – which is different from its true meaning in science. Most people would be surprised to learn that steel usually behaves elastically.

How deformation depends on force

How any object, or material, responds when forces act on them is obviously very important information when considering their use in practical situations.

The deformation of a steel spring is a common starting investigation because it is easy to measure and it will usually stretch regularly and elastically (unless over-stretched). See Figure A2.16.

◆ **Extension** Displacement of the end of an object that is being stretched.

Figure A2.17 shows typical results. The weights provide the downwards force, *F*. In this case the deformation is called the **extension** of the spring, *x*, and it is usually plotted on the horizontal axis of graphs.

Top tip!

No material will behave elastically under all conditions. They all have their limits: *elastic limits*. For this reason, it is probably sensible not to describe a material as being ‘elastic’. It is better to say that it behaved elastically under the conditions at that time.

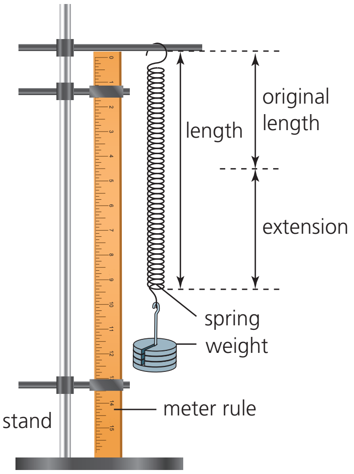


Figure A2.16 Steel spring investigation

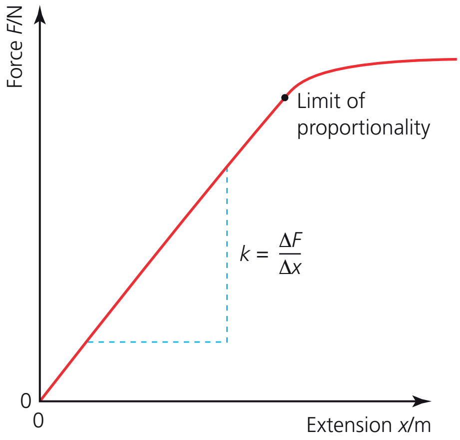


Figure A2.17 Results of stretching a steel spring

Most of the graph is a straight line passing through the origin. (The coils of the spring should not be touching each other at the beginning.) The conclusion is that the force, *F,* and the extension, *x*, are proportional to each other, up to a limit (as shown on the graph). The graph also shows that the spring gets easier to stretch after the limit of proportionality has been passed. For the linear part of the graph, starting at the origin: $F \propto x$.

The constant of proportionality is given the symbol *k*: $F = kx$.

◆ **Spring constant, *k*** The constant seen in **Hooke’s law** that represents the stiffness of a spring (or other material).

k is a measure of the ‘stiffness’ of the spring and is commonly called the **spring constant** (or the *force constant*). It can be determined from the gradient of the graph:

$$k = \frac{\Delta F}{\Delta x}$$

k has the SI units $\mathrm{N\,m^{-1}}$. ($\mathrm{N\,cm^{-1}}$ is also widely used.)

Hooke’s law

In the seventeenth century, Robert Hooke was famously the first to publish a quantitative study of springs. The physics law that describes his results is still used widely and bears his name:

Hooke’s law for elastic stretching: restoring force, $\boldsymbol{F_H = -kx}$

This is essentially the same as the equation $F = kx$, but the symbol F_H has been used for the force (to show that it is Hooke's law stretching), and the force refers to the **restoring force** within the spring, tending to return it to its original shape – this force is equal in size but opposite in direction to the externally applied force from the weights. The negative sign has been included to indicate that the restoring force acts in the opposite direction to increasing extension.

◆ **Restoring force** Force acting in the opposite direction to displacement, returning an object to its equilibrium position.

LINKING QUESTION

- How does the application of a restoring force acting on a particle result in simple harmonic motion?

This question links to understandings in Topic C.1.

Nature of science: Models

Obeying the law

Sometimes, everyday language differs from scientific terminology (for example, when speaking about 'weight'). So, what are 'laws' in science? If the extension of a stretched material is proportional to the force, we describe it as 'obeying' Hooke's law. In what way is that similar / different to 'obeying' a legal law?

Archimedes' description of buoyancy forces is described as a 'principle'. How are scientific 'principles' different from scientific 'laws'?

Research this online using search terms such as 'difference scientific principle and law'.

How might these concepts relate to theories and models in science?

Tool 3: Mathematics

Extrapolate and interpolate graphs

A curve of best fit is usually drawn to cover a specific range of measurements recorded in an experiment, as shown in Figure A2.18. The diagram indicates how values for y can be determined for a chosen values of x. If we want to predict other values within that range, we can usually do that with confidence. This is called **interpolation**.

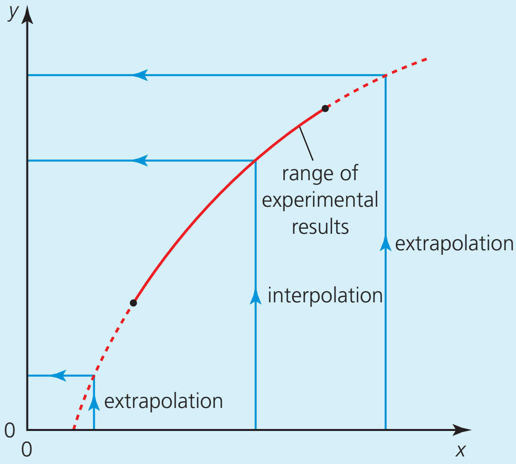


Figure A2.18 Interpolating and extrapolating to find values on the y-axis

If we want to predict what would happen outside the range of measurements (**extrapolation**) we need to extend the line of best fit. Lines are often extrapolated to see if they pass through the origin, or to find an intercept, as shown in Figure A2.18.

Predictions made by extrapolation should be treated with care, because it may be wrong to assume that the behaviour seen within the range of measurements also applies outside that range.

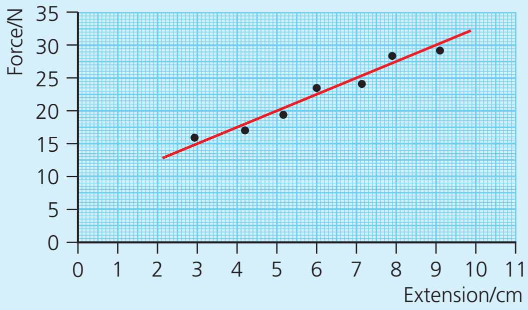


Figure A2.19 F–x graph for stretching a spring

Force–extension graphs, such as seen in Figure A2.19, are an interesting example.

a Use the graph to determine values for extensions when the force was 25 N, 10 N and 35 N.

b Use the graph to determine a possible value for the intercept on the force axis, and explain what it represents.

c Comment on your answers.

◆ **Interpolate** Estimate a value within a known data range.

◆ **Extrapolate** Predict behaviour that it outside of the range of available data.

WORKED EXAMPLE A2.3

When a weight of 12.7 N was applied to a spring its length was 15.1 cm. When the force increased to 18.3 N, the length increased to 18.1 cm because the extension was proportional to the force.

a Determine the spring constant.

b Calculate the length of the spring when the force was 15.0 N.

c Explain why it is impossible to be sure what the length of the spring would be if the force was 25 N.

Answer

a $k = \frac{\Delta F}{\Delta x} = \frac{(18.3 - 12.7)}{(18.1 - 15.1)} = 1.87\,\text{N}\,\text{cm}^{-1}$. Which is the same as $187\,\text{N}\,\text{m}^{-1}$.

b Consider the extension from a length of 15.1 cm:

$$\Delta x = \frac{\Delta F}{k} = \frac{(15.0 - 12.7)}{1.87} = 1.23\text{ cm}$$

So that, length = 15.1 + 1.23 = 16.3 cm

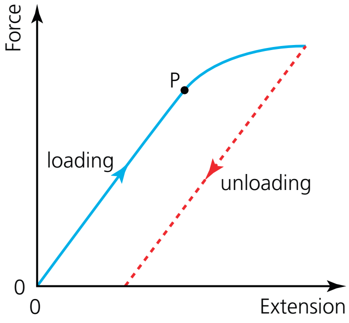
c The spring may have passed its limit of proportionality.

The results shown in Figure A2.19 were probably taken as the spring was *loaded* (as the weight was increased). If the extension is measured as the weight is *reduced* the results will be similar, but only if the elastic limit has not been exceeded.

The elastic limit of the spring is not shown on the graph, but it is often assumed to be close to, or the same as, the limit of proportionality. In other words, when a spring stretches, such as its extension is proportional to the force, we assume that it is behaving elastically. That may or may not be true for other materials.

Force–extension graphs and the concepts of elastic limits and ‘spring constants’ are not restricted to describing springs. They are widely used to represent the behaviour of many materials. Figure A2.20 shows a typical graph obtained when a metal wire is stretched and then the load is removed.

The force is proportional to the extension up until point P. During this time the particles in the metal are being pulled slightly further apart and we may assume that the metal is behaving elastically. But when the force is increased further, the wire begins to stretch more easily, the elastic limit is passed and a permanent deformation occurs. When the wire is unloaded the atoms move back closer together, so that the gradient of the graph is the same as for the loading graph, but the wire has a permanent deformation after all force has been removed.



■ **Figure A2.20** Stretching a metal wire

11 A spring has a spring constant of $125\,\text{N}\,\text{m}^{-1}$ and will become permanently deformed if its extension is greater than 20 cm.

  a Assuming that it behaves elastically, what extension results from a tensile force of 18.0 N?

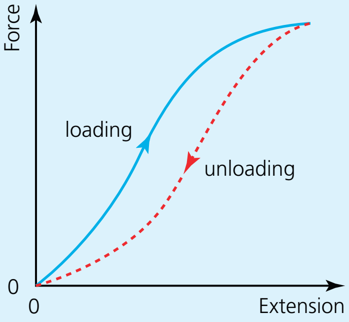
  b What is the maximum force that should be used with this spring?

12 When a mass of 200 g was hung on a spring its length increased from 4.7 cm to 5.3 cm.

  a Assuming that it obeyed Hooke’s law, what was its spring constant?

  b The spring behaves elastically if the force does not exceed 10 N. What is the length of the spring with that force?

13 Figure A2.21 shows a force–extension graph for a piece of rubber which was first loaded, then unloaded.



■ **Figure A2.21** Stretching rubber

  a Does the rubber behave elastically? Explain your answer.

  b Does the rubber obey Hooke’s law under the circumstances shown by the graph? Explain your answer.

Surface friction

SYLLABUS CONTENT

- Surface frictional force, F_f, acting in a direction parallel to the plane of contact between a body and a surface, on a stationary body as given by: $F_f \leqslant \mu_s F_N$, or a body in motion as given by: $F_f = \mu_d F_N$, where μ_s and μ_d are the coefficients of static and dynamic friction respectively.

◆ **Friction** Resistive forces opposing relative motion. Occurs between solid surfaces, but also with fluids. **Static friction** prevents movement, whereas **dynamic friction** occurs when there is already motion.

When we move an object over another surface (or try to move it), forces parallel to the surfaces will resist the movement. Collectively, these forces are known as surface **friction**. The causes of friction can be various, and it is well known that friction can often be difficult to analyse or predict. Figure A2.22 shows a typical simple frictional force diagram. (The frictional force acting on the ground is not shown.) The block is moving to the right and the frictional force is acting to the left.

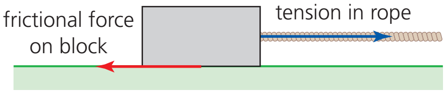


Figure A2.22 Frictional force on a block opposing its motion to the right

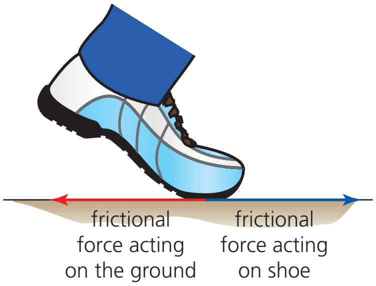


Figure A2.23 We need friction to walk

Friction is very useful: without friction we would not be able to walk. Similarly, a car's wheels would just spin on the same spot if there was no friction. Figure A2.23 explains why (the vertical forces are not shown). Because of friction, the shoe is able to push backwards, to the left, parallel to the ground, at the same time an equal frictional force pushes the shoe forward, to the right. (This is an example of Newton's third law of motion, which is discussed later in this topic.)

The roughness of both surfaces (see Figure A2.24) is certainly an important factor in producing friction: rougher surfaces generally increase friction, but this is not always true. For example, there may be considerable friction between very flat and smooth surfaces, like two sheets of glass. Friction can often be reduced by placing a lubricant, such as oil or water, between the surfaces.

Figure A2.25 shows a basic laboratory investigation of the frictional forces between a wooden block and a horizontal table top.

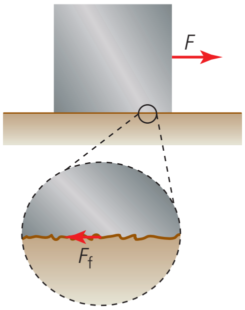


Figure A2.24 Even smooth surfaces have irregularities

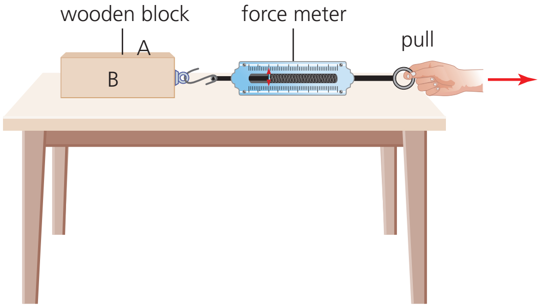


Figure A2.25 A simple experiment to measure frictional forces

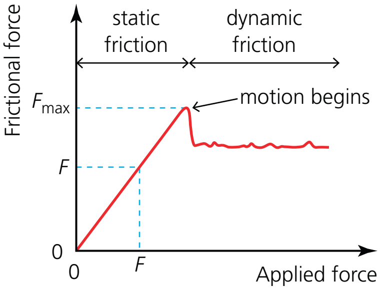
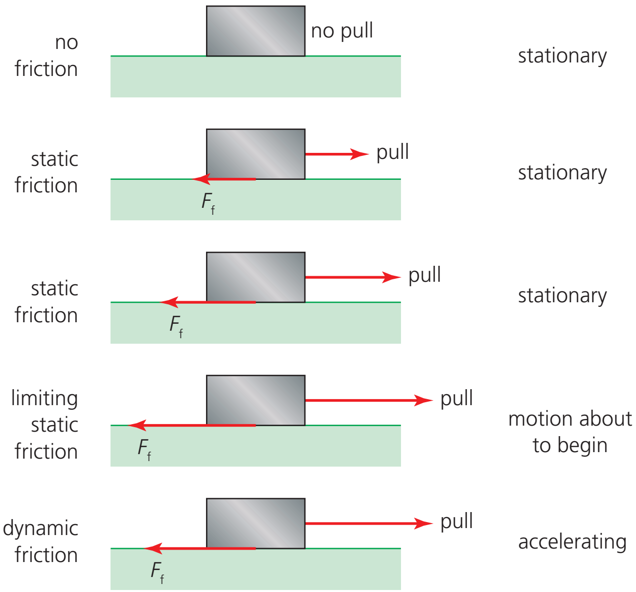


Figure A2.26 Variation of friction with applied force

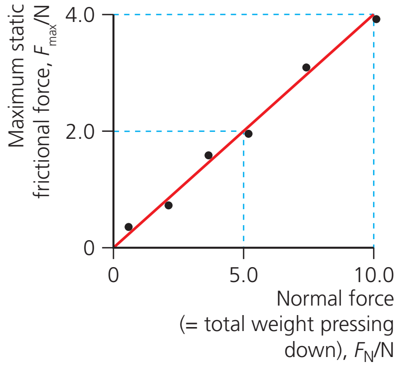
As the applied force (pull) is increased, the block will remain stationary until the force reaches a certain value, F_{max}. The block then starts to move, but after that, a steady force, which is less than F_{max}, will maintain a motion at constant speed. See Figure A2.26.

While the block is stationary (static) the force of friction adjusts, keeping equal to any applied force, but in the opposite direction. Under these circumstances the friction is called **static friction**. The size of the static friction force can increase from zero to a maximum value, F_{max}. Once an object is moving, the reduced friction is called **dynamic friction**, and its value is approximately constant at different speeds.

Figure A2.27 illustrates how frictional forces can change as a pulling force is increased.



■ **Figure A2.27** How frictional forces change as the force applied increases



■ **Figure A2.28** Typical variation of maximum static frictional force with normal force (a similar pattern of results will be obtained for dynamic friction)

The arrangement shown in Figure A2.25 can also be used to investigate how the *maximum* value of static friction depends on the force pushing the surfaces together: weights can be added on top of the block to increase the normal contact force, F_N. Figure A2.28 shows some typical results.

The graph shows that there is more static friction when there is a greater force pushing the surfaces together. In fact, frictional forces, F_f, are proportional to the normal contact forces, F_N. ($F_f \propto F_N$) The constant of proportionality equals the gradient of the graph and is called the **coefficient of friction**, μ (no units)

Just before motion begins: $F_f = F_{max} = \mu_s F_N$, where μ_s is the coefficient of static friction.

When there is no movement, static frictional force: $\boldsymbol{F_f \leq \mu_s F_N}$.

Table A2.1 shows some typical values for the coefficient of static friction between different materials.

◆ **Coefficient of friction, μ**
Constants used to represent the amount of friction between two different surfaces. Maybe static or dynamic.

Table A2.1 Approximate values for coefficients of static friction

Materials		**Approximate coefficients of static friction, μ_s**
steel	ice	0.03
ski	dry snow	0.04
Teflon™	steel	0.05
graphite	steel	0.1
wood	concrete	0.3
wood	metal	0.4
rubber tyre	grass	0.4
rubber tyre	road surface (wet)	0.5
glass	metal	0.6
rubber tyre	road surface (dry)	0.8
steel	steel	0.8
glass	glass	0.9
skin	metal	0.9

When there is movement, dynamic frictional force, $\boldsymbol{F}_f = \boldsymbol{\mu}_d \boldsymbol{F}_N$, where μ_d is the coefficient of dynamic friction.

◆ **Constant** A number which is assumed to have the same numerical value under a specified range of circumstances.

◆ **Fundamental constants** Numbers which are assumed to have exactly the same numerical values under all circumstances and all times.

◆ **Coefficient** A multiplying constant placed before a variable, indicating a physical property.

Tool 3: Mathematics

Applying general mathematics: constants

A number which is assumed to be **constant** always has the same value under the specified circumstances. For example, the spring constant described earlier in this topic represents the properties of a spring, but only up to its limit of proportionality. In Topic A.1, the acceleration due to gravity was assumed to be constant at 9.8 m s^{-2}, but only if we limit precision to 2 significant figures and only apply it to situations close to the Earth's surface.

However, there are a few constants which are believed to have exactly the same value in all locations and for all time. They are called the **fundamental constants**, or universal constants. Two examples are the speed of light and the charge on an electron.

In general, a **coefficient** is a number (usually a constant) placed before a variable in an algebraic expression. For example, in the expression $5a - 2 = 8$, the number 5 is described as a coefficient. In physics, a coefficient is used to characterize a physical process under certain specified conditions.

We have seen that: dynamic frictional force, F_f = *coefficient* of dynamic friction × F_N

Another example (which is not in the IB course): when a metal rod is heated it expands so that increase in length for each 1 °C temperature rise
= *coefficient* of thermal expansion × original length.

Objects also experience friction when they move through liquids and gases (fluids). This is discussed in the next section.

ATL A2A: Research skills

Using search engines and libraries effectively

Tyres and road safety

Much of road safety is dependent on the nature of road surfaces and the tyres on vehicles. Friction between the road and a vehicle provides the forces needed for any change of velocity – speeding up, slowing down, and changing direction. Smooth tyres will usually have the most friction in dry conditions, but when the roads are wet, ridges and grooves in the tyres are needed to disperse the water (Figure A2.29).

To make sure that road surfaces produce enough friction, they cannot be allowed to become too smooth and they may need to be resurfaced every few years. This is especially important on sharp corners and hills. Anything that gets between the tyres and the road surface – for example, oil, water, soil, ice and snow – is likely to affect friction and may have a significant effect on road safety. Increasing the area of tyres on a vehicle will change the pressure underneath them and this may alter the nature of the contact between the surfaces. For example, a farm tractor may have a problem about sinking into soft ground, and such a situation is more complicated than simple friction between two surfaces. Vehicles that travel over soft ground need tyres with large areas to help avoid this problem.

Using a search engine, research online to find what materials are used in the construction of tyres and road surfaces to produce high coefficients of friction. Organize your data in a table, making sure to credit your sources using a recognized, standard method of referencing and citation.

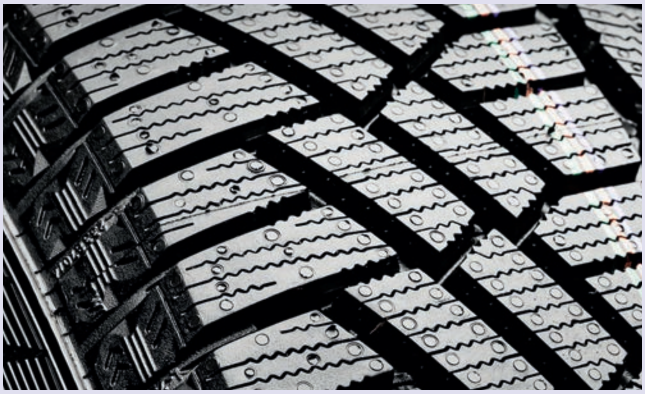


Figure A2.29 Tread on a car tyre

WORKED EXAMPLE A2.4

a Determine the coefficient of friction for the two surfaces represented in the graph shown in Figure A2.28.

b Assuming the results were obtained for apparatus like that shown in Figure A2.25, calculate the minimum force that would be needed to move a block of total mass:

- **i** 200 g
- **ii** 2000 g.
- **iii** Suggest why the answer to part **ii** is unreliable.

c Estimate a value for the dynamic frictional force acting on a mass of 200 g with the same apparatus:

- **i** for movement at $1.0\,\text{m}\,\text{s}^{-1}$
- **ii** for movement at $2.0\,\text{m}\,\text{s}^{-1}$.

Answer

a $\mu_s = \dfrac{F_{max}}{F_N} = \dfrac{4.0}{10.0} = 0.40$ (This is equal to the gradient of the graph.)

b **i** $F_f = \mu_s F_N = \mu_s mg = 0.40 \times 0.200 \times 9.8 = 0.78\,\text{N}$

ii $0.40 \times 2.000 \times 9.8 = 7.8\,\text{N}$

iii Because the answer is extrapolated from well outside the range of experimental results shown on the graph.

c **i** We would expect the dynamic frictional force to be a little less than the static frictional force, say about 0.6 N instead of 0.78 N.

ii The dynamic frictional force is usually assumed to be independent of speed, so the force would still be about 0.6 N at the greater speed.

Common mistake

Many students expect that, if the block in Figure A2.25 was rotated so that side B was in contact with the table (instead of the side parallel to A), there would be more friction because of the greater area of contact. However, the frictional force will remain (approximately) the same, because if, for example, the area doubles, the force acting down on each cm^2 will halve.

Use data from Table A2.1 where necessary.

14 If dynamic friction is 85% of the maximum static friction, estimate the frictional force acting on the steel skates of a 47 kg ice-skater moving across the ice.

15 A 54 kg wooden box is on a horizontal concrete floor.
- **a** Estimate the minimum force required to start it sliding sideways.
- **b** Suggest why your answer to part **a** may not be reliable.
- **c** If a force of 120 N keeps the box moving at a constant speed, what is the coefficient of dynamic friction?
- **d** What will happen to the box if the applied force increases above 120 N?

16
- **a** Predict the maximum frictional force possible between a dry road surface and each tyre of a stationary, 1500 kg four-wheeled family car.
- **b** Why will the force be less if the road is wet or icy?
- **c** Discuss how roads can be made safer under icy conditions.

17 Figure A2.30 shows the front of a Formula One racing car. Suggest how this design helps to increase the friction between the tyre and the race track.

Figure A2.30 Front of a Formula One racing car

18 A book of mass 720 g is being held in place next to a vertical wall as shown in Figure A2.31.
- **a** State the weight of the book.
- **b** Suggest an approximate value for the coefficient of static friction between the book and the wall.
- **c** Use your answer to part **b** to estimate the minimum force needed to keep the book stationary against the wall.

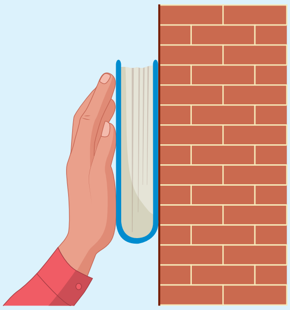


Figure A2.31 Book being held next to a vertical wall

Friction of objects with air and liquids

SYLLABUS COVERAGE

- Viscous drag force, F_d, acting on a small sphere opposing its motion through a fluid as given by: $F_d = 6\pi\eta rv$, where η is the fluid viscosity, r is the radius of the sphere and v is the velocity of the sphere through the fluid.

Air resistance was briefly discussed in Topic A.1. The word *drag* is widely used to describe friction in air and liquids. We will use the symbol F_d for this type of force.

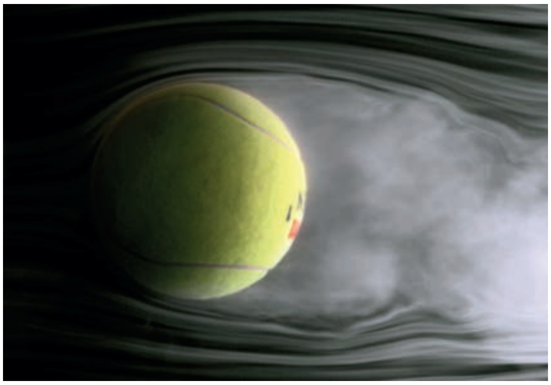


Figure A2.32 Flow of air past a tennis ball in a wind tunnel.

There are a great number of applications of this subject, including moving vehicles, sports and falling objects. Wind tunnels are useful in the study of drag: the object is kept stationary while the speed of air flowing past it is varied. The flow of the air can be marked as shown in Figure A2.32.

Drag can be a complicated subject because the amount of drag experienced by an object moving through air, or a liquid, depends on many factors, including the object's size and shape, the nature of its surface, its speed v, and the nature of the fluid. Drag will also depend on the cross-sectional area of the object (perpendicular to its movement).

Typically, for small objects moving slowly $F_{\mathrm{d}} \propto v$.

But for larger objects, moving more quickly, $F_{\mathrm{d}} \propto v^2$.

Viscosity and Stokes's law

◆ **Viscosity** Resistance of a fluid to movement.

◆ **Viscous drag** The drag force acting on a moving object due to the viscosity of the fluid through which it is moving.

◆ **Turbulence** Flow of a fluid which is erratic and unpredictable.

◆ **Stokes's law** Equation for the viscous drag acting on a smooth, spherical object undergoing non-turbulent motion.

When an object moves through a fluid it has to push the fluid out of its path. A fluid's resistance to such movement is called its **viscosity**. Clearly, greater viscosity will tend to increase drag, and when this is the dominant factor, we refer to **viscous drag**.

Viscosity is given the symbol η (eta) and has the SI unit of Pa s (kg m^{-1} s^{-1}). Some typical values at 20 °C are given in Table A2.2. Viscosities of liquids can be very dependent on temperature.

Table A2.2 Viscosities of some fluids

Fluid	**Viscosity η/Pa s**
'heavy' oil	0.7
'light' oil	0.1
water	1×10^{-3}
human blood	4×10^{-3}
gasoline (petrol)	6×10^{-4}
air	1.8×10^{-5}

In order to understand this further, we start by simplifying the situation, as is common in physics: by considering a smooth spherical object, of radius r, moving at a speed v, which is not great enough to cause **turbulence** (irregular movements) in the fluid.

Under these circumstances, the viscous drag, F_{d}, can be determined from the following equation (known as **Stokes's law**):

viscous drag $F_{\mathrm{d}} = 6\pi\eta r v$

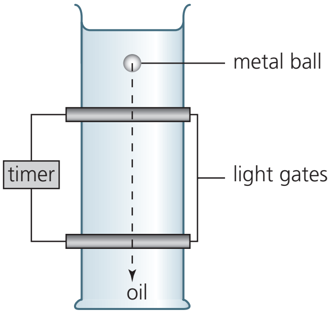


Figure A2.33 Experiment to determine the viscosity of a liquid

Dropping small spheres through fluids is a widely used method for determining their viscosities and how they may depend on temperature. A method is shown in Figure A2.33, in which an electronic timer is started and stopped as the metal ball passes through the two light gates.

Inquiry 1: Exploring and designing

Designing

Look at the apparatus setup in Figure A2.33. Apply what you know about terminal speed (Topic A.1) and viscous fluid flow to design and explain a valid methodology for an experiment to obtain a single set of measurements. Include an explanation of:

1. why the metal ball is released such that it passes through some oil before reaching the first timing gate
2. why the tube should be as wide as possible.

■ **Figure A2.34** Forces on a sphere falling with terminal speed

If a sphere of mass m and radius r is moving with a constant terminal speed, v_t, then the upwards and downwards forces on it are balanced, as shown in Figure A2.34.

viscous drag, F_d + buoyancy force, F_b = weight, mg:

$$6\pi\eta rv + \rho Vg = mg$$

but:

$$V = \frac{4}{3}\pi r^3$$

so:

$$6\pi\eta rv + \frac{4}{3}\rho g\pi r^3 = mg$$

If the mass and radius of the sphere are measured and the terminal speed determined as shown in Figure A2.33, then this equation can be used to determine a value for the viscosity of the liquid, assuming that its density is known.

Inquiry 3: Concluding and evaluating

Evaluating

The experimental determination of a viscosity discussed above involved just one set of measurements and a calculation.

Explain improvements to increase the accuracy of the determination of the viscosity of a liquid by collecting sufficient data to enable a graph of the results to be drawn.

WORKED EXAMPLE A2.5

Calculate the force of viscous drag on a sphere of radius 1.0 mm moving at 1.0 cm s^{-1} through ‘heavy’ oil.

Answer

$F_d = 6\pi \times \eta \times r \times v = 6 \times 3.14 \times 0.7 \times (1.0 \times 10^{-3}) \times (1.0 \times 10^{-2}) = 1.3 \times 10^{-4}$ N

19 The air resistance acting on a car moving at 5.0 m s^{-1} was 120 N.
Assuming that this force was proportional to the speed squared, what was the air resistance when the car's speed increased to:
- **a** 10 m s^{-1}
- **b** 15 m s^{-1}?

20 Show that the units of viscosity are Pa s.

21 Calculate the viscous drag force acting on a small metal sphere of radius 1.3 mm falling through oil of viscosity 0.43 Pa s at a speed of 7.6 cm s^{-1}.

22 A drop of water in a cloud had a mass of 0.52 g and radius of 0.50 mm (and volume of 0.52 mm^3).
- **a** Assuming that the density of the surrounding air is 1.3 kg m^{-3}, calculate and compare the size of the three forces acting on the drop if it has just started to fall with a speed of 5.0 cm s^{-1}.
- **b** Draw an annotated diagram to display your answers.
- **c** Determine the subsequent movement of the drop.

23 In an experiment similar to that shown in Figure A2.33, a sphere of radius 8.9 mm and mass 3.1 g reached a terminal speed of 7.6 cm s^{-1} when falling through an oil of density 842 kg m^{-3}.
Determine a value for the viscosity of the liquid.

24 Use the internet to find out how the design of golf balls reduces drag forces in flight. Write a 100 word summary of your findings.

ATL A2B: Thinking skills

Evaluating and defending ethical positions

Air travel

Aircraft use a lot of fuel moving passengers and goods from place to place quickly, but we are all becoming more aware of the effects of planes on global warming and air pollution. Some people think that governments should put higher taxes on the use of planes to discourage people from using them too much. Improving railway systems, especially by operating trains at higher speeds, will also attract some passengers away from air travel. Of course, engineers try to make planes more efficient so that they use less fuel, but the laws of physics cannot be broken and jet engines, like all other *heat engines*, cannot be made much more efficient than they are already.

Planes will use a lower fuel if there is a lower air resistance acting on them. This can be achieved by designing planes with **streamlined** shapes, and also by flying at greater heights where the air is less dense. Flying more slowly than their maximum speed can also reduce the amount of fuel used for a particular trip, as it does with cars, but people generally want to spend as little time travelling as possible.

The pressure of the air outside an aircraft at its typical cruising height is far too low for the comfort and health of the passengers and crew, so the air pressure has to be increased inside the airplane, but this is still much lower than the air pressure near the Earth's surface. The difference in air pressure between the inside and outside of the aircraft would cause problems if the airplane had not been designed to withstand the extra forces.

Aircraft generally carry a large mass of fuel, and the weight of an aircraft decreases during a journey as the fuel is used up. The upwards force supporting the weight of an aircraft in flight comes from the air that it is flying through and will vary with the speed of the airplane and the density of the air. When the aircraft is lighter towards the end of its journey it can travel higher, where it will experience less air resistance.

Debate the issue in class. Break into groups. One group can represent the airline operators, another group can represent passengers, a third group can represent an environmental campaign group, while a fourth group could represent the government. In your groups, allocate roles for researchers and a spokesperson. Using the information above and your understanding of air resistance prepare a proposal from the point of view of your assigned group detailing different ways in which we can reduce the environmental impact of air travel.

To help your research and calculations, refer to the following guiding questions:

- How do airlines hope that in the future they can become 'carbon neutral'. What is 'SAF'?
- Find out how much fuel is used on a long-haul flight of, say, 12 hours.
- Compare your answer with the capacity of the fuel tank on an average sized car.
- On a short-haul flight it is often claimed that as much of 50% of an aircraft's fuel might be used for taxiing, taking off, climbing and landing, but on longer flights this can reduce to under 15%. Explain the difference.

◆ **Streamlined** Having a shape that reduces the drag forces acting on an object that is moving through a fluid.

◆ **Field (gravitational, electric or magnetic)** A region of space in which a mass (or a charge, or a current) experiences a force due to the presence of one or more other masses (charges, or currents – moving charges).

LINKING QUESTION

- How can knowledge of electrical and magnetic forces allow the prediction of changes to the motion of charged particles?

This question links to understandings in Topic D.3.

Field forces

SYLLABUS CONTENT

- The nature and use of the following field forces.
  - Gravitational force, F_g, as the weight of the body and calculated as given by: $F_g = mg$
  - Electric force F_e
  - Magnetic force F_m

These three forces are very important in the study of physics but, apart from the gravitational force of weight, knowledge about them is not required in *this* topic.

These forces can act across empty space, without the need for any material in between. This can be difficult for the human mind to accept. One way of increasing our understanding is to develop the concept of force **fields** surrounding masses (gravitational fields), charges (electric fields) and magnets / electric currents (magnetic fields). Using this concept, we can give numerical values to points in space, for example, by stating that the gravitational field strength at the height of a particular satellite's orbit is $8.86\,\mathrm{N\,kg^{-1}}$.

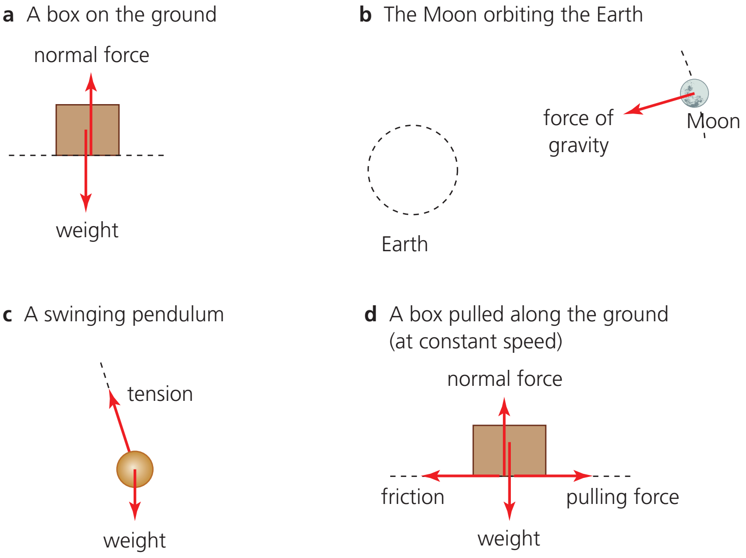
Free-body diagrams

SYLLABUS CONTENT

- Forces acting on a body can be represented in a free-body diagram.

Even the simplest of force diagrams can get confusing if all the forces are shown. To make the diagrams simpler we usually draw only one object and show only the forces acting on that one object. These drawings are called **free-body diagrams**. (Physicists use the words ‘body’ and ‘object’ interchangeably.) Some simple examples are shown in Figure A2.35.

◆ **Free-body diagram** Diagram showing all the forces acting on a single object, and no other forces.


a A box on the ground
b The Moon orbiting the Earth
c A swinging pendulum
d A box pulled along the ground (at constant speed)

■ **Figure A2.35** Free-body diagrams; the object has a solid outline and the forces are shown in red

The diagrams are often further simplified by representing the object as a small square, or circle, and considering it to be a **point particle / mass**.

◆ **Point particle, mass or charge** Theoretical concept used to simplify the discussion of forces acting on objects (especially in gravitational and electric fields).

Nature of science: Models

Point objects, particles and masses

A point particle is an idealized, simplified representation of any object, whatever its actual size and shape. As the name suggests, a point particle does not have any dimensions, or occupy any space. Typically, the ‘point’ will be located at the centre of mass of the object.

When the concept is used, we do not need to consider the complications and variations that are involved with extended objects. For example, if we consider an object as a point particle, all forces act through the same point and analysis can ignore any possible rotational effects caused by the forces acting on it.

Resultant forces and components

SYLLABUS CONTENT

- Free-body diagrams can be analysed to find the resultant force on a system.

Tool 3: Mathematics

Add and subtract vectors in the same plane

Vector addition is an important mathematical skill that occurs in several places in the IB Physics course, but the addition of forces is the most common application. Figure A2.36 shows an example of how to find the resultant of two force vectors.

A **resultant force** is represented in size and direction by the diagonal of the parallelogram (or rectangle) which has the two original force vectors as adjacent sides.

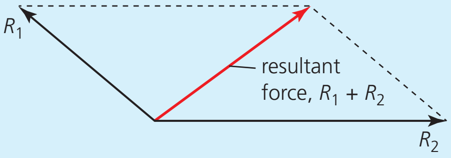


Figure A2.36 Adding two forces to determine a resultant

◆ **Resultant force** The vector sum of the forces acting on an object, sometimes called the unbalanced or net force.

◆ **Resultant** The single vector that has the same effect as the combination of two or more separate vectors.

◆ **Components (of a vector)** Any single vector can be considered as having the same effect as two parts (components) perpendicular to each other.

◆ **Inclined plane** Flat surface at an angle to the horizontal (but not perpendicular). A simple device that can be used to reduce the force needed to raise a load; sometimes called a ramp.

Tool 3: Mathematics

Resolve vectors

As we have seen, two forces can be combined to determine a single **resultant**. The 'opposite' process is very useful: a single force, *F*, can be considered as being equivalent to two smaller forces at right angles to each other. The two separate forces are called **components**.

This process is called resolving a force into two components. It can be used when the original force is not acting in a direction which is convenient for analysis. Because the two components are perpendicular to each other their effects can be considered separately. Figure A2.37 shows how a force can be resolved into two perpendicular components.

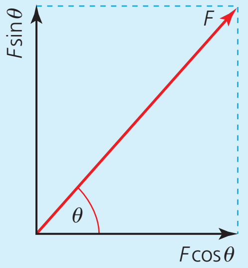


Figure A2.37 Force, *F*, resolved into two components

Any force, *F*, can be resolved into two independent components which are perpendicular to each other:

$F\sin\theta$ and $F\cos\theta$

WORKED EXAMPLE A2.6

a Draw a free-body diagram for an object which is stationary on a slope (**inclined plane**) which makes an angle of 35° with the horizontal.

b The object has a mass of 12.7 kg and just begins to slide down the slope if the angle is 35°. Using $g = 9.81\,\text{N}\,\text{kg}^{-1}$, calculate the component of the weight for this angle:

- **i** down the slope
- **ii** perpendicular into the slope.

c State values of the frictional force and the normal force acting on the object.

d Determine the coefficient of static friction in this situation.

Answer

a See Figure A2.38, which represents the object as a point. The resultant contact force from the slope on the object must be equal and opposite to the weight, F_g. The contact force can be considered as the combination of two perpendicular components: F_N perpendicular to the slope, and F_f, the frictional force stopping the object from sliding down the slope.

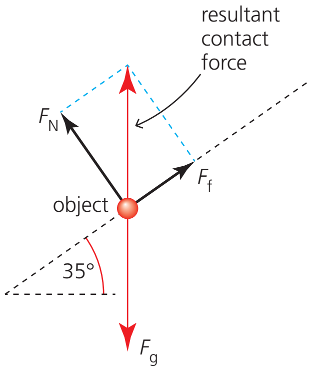


Figure A2.38 Free-body diagram for an object on a slope

Sometimes it is preferred to represent the object as more than just a point. See Figure A2.39 for an example. However, this may cause confusion about exactly where the forces act.

b See Figure A2.39.

Component down slope $mg \sin 35° = 12.7 \times 9.81 \times 0.574 = 71.5\,\text{N}$

Component into slope $= mg \cos 35° = 12.7 \times 9.81 \times 0.819 = 102\,\text{N}$

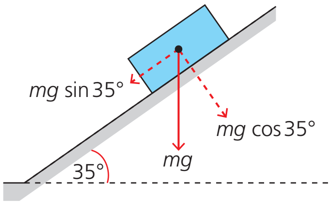


Figure A2.39 Components of weight

c Frictional force equals component down the slope, but in the opposite direction = 71.5 N up the slope.

Normal force equals component into the slope, but in the opposite direction = 102 N upwards.

d $F_f = \mu_s F_N$

$$\mu_s = \frac{F_f}{F_N} = \frac{71.5}{102} = 0.70 \text{ (which is equal to } \tan\theta\text{)}$$

25 Draw fully labelled free-body diagrams for:

- **a** a car moving horizontally with a constant velocity
- **b** an aircraft moving horizontally at constant velocity
- **c** a boat decelerating after the engine has been switched off
- **d** a car accelerating up a hill.

26 A wooden block of mass 2.7 kg rests on a slope which is inclined at 22° to the horizontal.

- **a** Make calculations which will enable you to draw a free-body diagram, similar to Figure A2.38, but giving numerical values for the forces.
- **b** If the angle is increased, the block will slide down the slope. Calculate the coefficient of friction.
- **c** State whether your answer to part **b** is for static or kinetic friction.

27 A pendulum on the end of a string has a mass of 158 g.

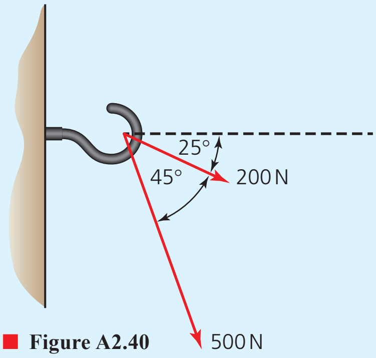
- **a** Draw a free-body diagram representing the situation when the string is making an angle of 20° to the vertical.
- **b** By adding components of weight to your diagram, show that the tension in the string is 1.5 N.
- **c** What effect does the other component ($mg \sin\theta$) have on the pendulum?
- **d** Discuss how the tension in the string changes while the pendulum is swinging from side to side.

28 Parallel forces of 1 N, 2 N and 3 N can act on an object at the same time. State the values of all the possible resultant forces.

29 Calculate the resultant force (size and direction) of 4.7 N and 5.9 N which are perpendicular to each other and acting away from a point mass.

30 Show that a mass on an inclined plane will just begin to slip down the slope when the tangent of the angle to the horizontal equals the coefficient of static friction.

31 Determine by scale drawing or calculation the size and direction of the resultant force acting on the hook shown in Figure A2.40.



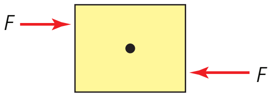
■ **Figure A2.40**

◆ **Newton's laws of motion** **First law**: an object will remain at rest, or continue to move in a straight line at a constant speed, unless a resultant force acts on it; **Second law**: acceleration is proportional to resultant force; **Third law**: whenever one body exerts a force on another body, the second body exerts exactly same force on the first body, but in the opposite direction.

◆ **Balanced forces** If an object is in mechanical equilibrium, we describe the forces acting on it as 'balanced'.

◆ **Equilibrium** An object is in equilibrium if it is unchanging under the action of two or more influences (e.g. forces). Different types of equilibrium include **translational**, **rotational** and **thermal**.

◆ **Translational** Changing position.



■ **Figure A2.41** The object is in translational equilibrium, but not in rotational equilibrium

Newton's laws of motion

SYLLABUS CONTENT

- Newton's three laws of motion.

Newton's three laws of motion are among the most famous in classical physics. They describe the relationships between force and motion. Although they were first stated more than three hundred years ago, they are equally important today and are essential for an understanding of all motion (except when a speed of motion is close to the speed of light, as discussed in Topic A.5).

Newton's first law of motion

Newton's first law of motion states that an object will remain at rest or continue to move in a straight line at a constant speed, unless a resultant force acts on it.

In other words, a resultant force will produce an acceleration (change in velocity).

When the influences on any system are **balanced**, so that the system does not change, we describe it as being in **equilibrium**. (As another example, if an object stays at the same temperature, we say that it is in **thermal equilibrium**.)

When there is no resultant force on an object, we say that it is in **translational equilibrium**.

The term **translational** refers to movement from place to place. An object is in translational equilibrium if it remains at rest or continues to move with a constant velocity (in a straight line at a constant speed), as described by Newton's first law.

In passing, it should be noted that, if equal forces act in opposite directions, an object will be in translational equilibrium, but if the forces are not aligned (see Figure A2.41) then the object may start to rotate, so it will not be in **rotational equilibrium**. The subject of rotational dynamics is covered in Topic A.4.

Nature of science: Observations

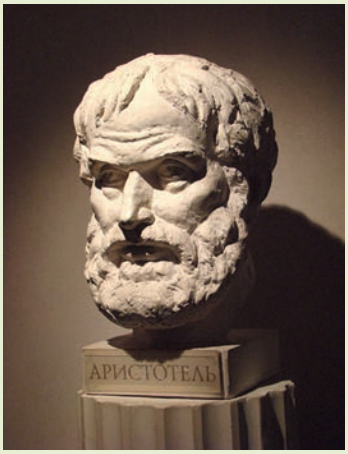
Natural philosophy

‘Forces are needed to keep an object moving and, without those forces, movement will stop.’ This accepted ‘fact’ is not true, but it is still widely believed. It was the basis of theories of motion from the time of Aristotle (about 2350 years ago) until the seventeenth century, when scientists began to understand that the forces of friction were responsible for stopping movement.

Aristotle is one of the most respected figures in the early development of human thought. He appreciated the need for wide-ranging explanations of natural phenomena but the ‘science’ of that time – called natural philosophy – did not involve careful observations, measurements, mathematics or experiments.

Aristotle believed that everything in the world was made of a combination of the four elements of earth, fire, air and water. The Earth was the centre of everything and each of the four Earthly elements had its natural place. When something was not in its natural place, then it would tend to return – in this way he explained why rain falls, and why flames and bubbles rise, for example.

Modern science (characterized by experimentation and the development of unbiased, testable theories) began in the seventeenth century. It includes the work of famous physicists mentioned in this topic: Hooke, Galileo and Newton.

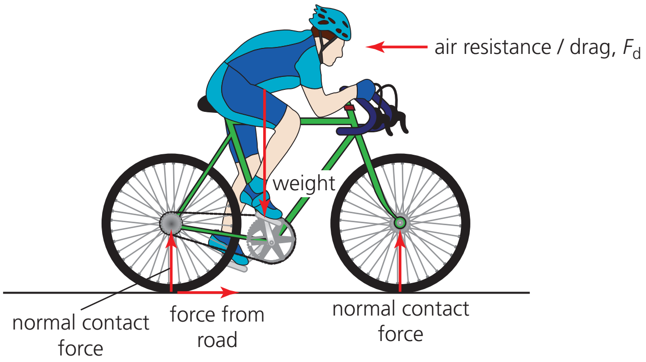


■ **Figure A2.42** A representation of Aristotle

◆ **Natural philosophy** The name used to describe the (philosophical) study of nature and the universe before modern science.

Examples of translational equilibrium

Because all objects on Earth have weight, it is not possible for an object to be in equilibrium because there are no forces acting on it. So, all translational equilibrium arises when two or more forces are balanced.

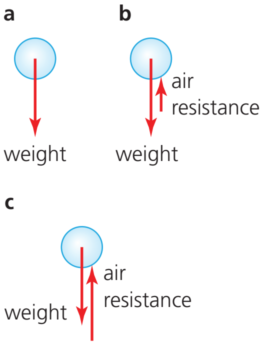


■ **Figure A2.43** A cyclist moving at constant speed in translational equilibrium

- A book on a horizontal table (Figure A2.8) is in equilibrium because its downwards weight is balanced by the upwards normal contact force.
- A stationary block on a slope (Figures A2.38 and A2.39) is in equilibrium because the component of its weight down the slope is balanced by surface friction up the slope and the component of its weight into the slope is balanced by the normal component of the contact force.
- A cyclist moving with constant speed (Figure A2.43) is in equilibrium because their weight is balanced by the sum of the two normal contact forces and the frictional force from the road is balanced by the drag.

Falling through the air at terminal speed

Figure A2.44 shows three positions of a falling ball. In part **a** the ball is just starting to move and there is no air resistance / drag. In part **b** the ball has accelerated and has some air resistance acting against its motion, but there is still a resultant force and an acceleration downwards. In part **c** the speed of the falling ball has increased to the point where the increasing air resistance has become equal and opposite to the weight. There is then no resultant force and the ball is in translational equilibrium, falling with a constant velocity called its terminal velocity or terminal speed. (Any buoyancy forces are considered to be negligible under these circumstances.) Terminal speed was introduced in Topic A.1.

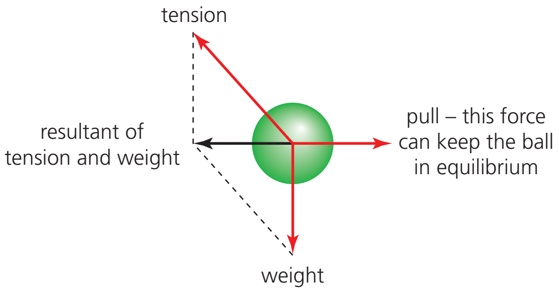


■ **Figure A2.44** The resultant force on a falling object changes as it gains speed

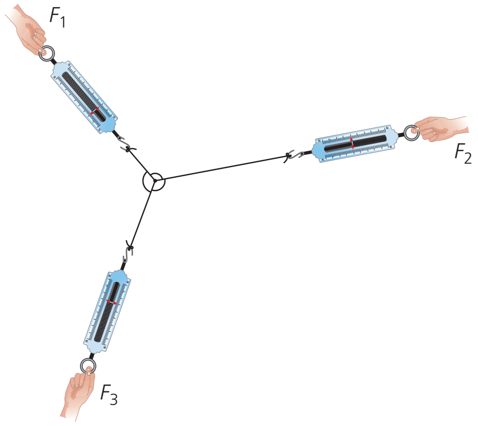
Three forces in equilibrium

If two forces are acting on an object such that it is not in equilibrium, then to produce equilibrium a third force can be added that is equal in size to the resultant of the other two, but in the opposite direction. All three forces must act through the same point. For example, Figure A2.45 shows a free-body diagram of a ball on the end of a piece of string kept in equilibrium by a sideways pull that is equal in magnitude to the resultant of the weight and the tension in the string.

The translational equilibrium of three forces can be investigated in the laboratory simply by connecting three force meters together with string just above a horizontal surface, as shown in Figure A2.46. The three forces and the angles between them can be measured for a wide variety of different values, each of which maintains the system stationary.



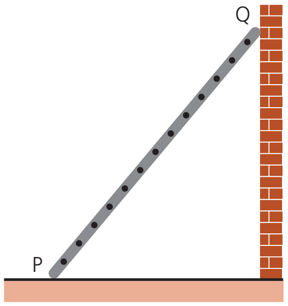
■ **Figure A2.45** Three forces keeping a suspended ball in equilibrium



■ **Figure A2.46** Investigating three forces in equilibrium

WORKED EXAMPLE A2.7

A ladder is leaning against a wall, as shown in Figure A2.47. Friction at point P is stopping the ladder from slipping, but there is no need for any friction acting at point Q.



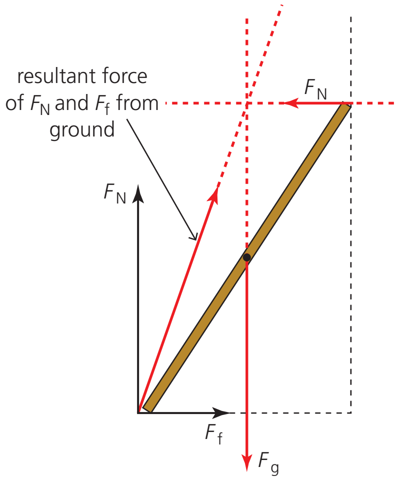
■ **Figure A2.47** A ladder leaning against a wall

a Draw a free-body diagram of the ladder, including its weight and the normal force from the wall.

b The resultant force on the ladder from the ground must be directed at the point where the lines of action of the other two forces intersect. Add this line to your diagram.

c Complete the diagram by adding the two perpendicular components of the force from the ground on the ladder.

Answer



■ **Figure A2.48**

32 Under what circumstances will a moving car be in translational equilibrium?

33 If you are in an elevator (lift) without windows discuss whether it is possible to know if you are moving up, moving down or stationary.

34 Figure A2.49 shows a mountain climber who, at that moment, is stationary.
  a Draw a free-body diagram that shows that he is in equilibrium.
  b Outline the features of your diagram which show that the climber is in equilibrium.

35 Can the Moon be described as being in translational equilibrium? Explain your answer.

■ **Figure A2.49**

Newton's second law of motion

We have seen that Newton's first law establishes that there is a connection between resultant force and acceleration. Newton's second law takes this further and states the mathematical connection: when a resultant force acts on a (constant) mass, the acceleration is proportional to the resultant force: $a \propto F$.

Both force and acceleration are vector quantities and the acceleration is in the same direction as the force.

Investigating the effects of different forces and different masses on the accelerations that they produce is an important part of most physics courses, although reducing the effects of friction is essential for consistent results.

Inquiry 1: Exploring and designing

Exploring

Aristotle's understanding of motion was formed through making observations of the behaviour of objects in motion, but without any deep understanding of the concept of force he was unable to account for the effects of friction or air resistance. What methods are available for reducing friction in investigations into the effects of different forces and masses on an object's acceleration?

In groups, brainstorm how experiments can be designed to reduce or to cancel the effects of frictional forces. Decide on a selection of search terms or phrases that can be used by individual students for internet research. Use your research to formulate a research question and hypothesis.

Such experiments also show that when the same resultant force is applied to different masses, the acceleration produced is inversely proportional to the mass, m: $a \propto 1/m$

Combining these results, we see that acceleration, $a \propto \frac{F}{m}$.

Newton's second law can be written as: $F \propto ma$

If we define the SI unit of force, the newton, to be the force that accelerates 1 kg by 1 m s^{-2}, then we can write: force (N) = mass (kg) × acceleration (m s^{-2})

Newton's second law of motion: resultant force, $F = ma$

◆ **Proportional relationship** Two variables are (directly) proportional to each other if they always have the same ratio.

◆ **Uncertainty bars** Vertical and horizontal lines drawn through data points on a graph to represent the uncertainties in the two values.

This version of Newton's second law assumes that the mass of the object is constant. We will see later in this topic that there is an alternative version which allows for changing mass.

When discussing a gravitational force, weight, we have used the symbol F_g and the acceleration involved is g, the acceleration of free fall.

So, the equation $F = ma$ becomes the familiar:

$$F_g = mg$$

Tool 3: Mathematics

On a best-fit linear graph, construct lines of maximum and minimum gradients with relative accuracy (by eye) considering all uncertainty bars

Many basic physics experiments are aimed at investigating if there is a **proportional relationship** between two variables, and this is usually best checked by drawing a graph.

If two variables are (directly) proportional, then their graph will be a straight line passing through the origin

Figure A2.50 represents a proportional relationship. It is important to stress that a linear graph that does not pass through the origin does *not* represent proportionality (Figure A2.51).

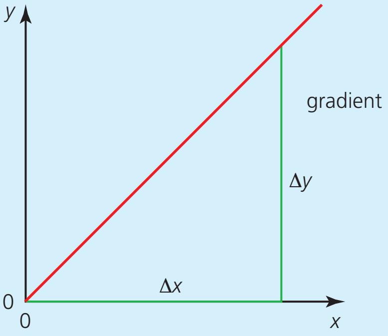


Figure A2.50 A proportional relationship

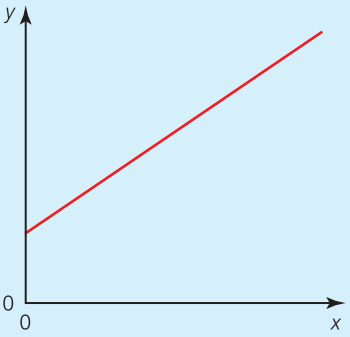


Figure A2.51 A linear relationship that is not proportional does not pass through the origin. **See also Tool 3: Mathematics (Understand direct and inverse proportionality) on page 129.**

Uncertainty in gradients and intercepts

It is often possible to draw a range of different straight lines, all of which pass through the **uncertainty bars** representing experimental data.

We usually assume that the line of best fit is midway between the lines of maximum possible gradient and minimum possible gradient. Figure A2.52 shows an example (for simplicity, only the first and last error bars are shown, but in practice all the error bars need to be considered when drawing the lines).

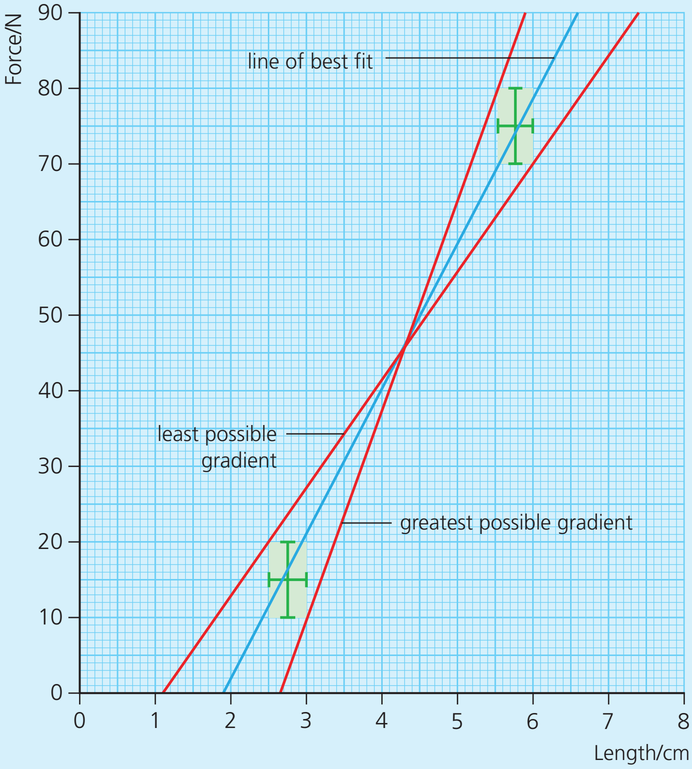


Figure A2.52 Finding maximum and minimum gradients for a spring-stretching experiment

Figure A2.52 shows how the length of a metal spring changed as the force applied was increased. We know that the measurements were not very precise because the uncertainty bars are large. The line of best fit has been drawn midway between the other two. This is a linear graph (a straight line) and it is known that the gradient of the graph represents the force constant (stiffness) of the spring and the horizontal intercept represents the original length of the spring. Taking measurements from the line of best fit, we can make the following calculations:

$$\text{force constant} = \text{gradient} = \frac{(90 - 0)}{(6.6 - 1.9)} = 19\,\text{N}\,\text{cm}^{-1}$$

$$\text{original length} = \text{horizontal intercept} = 1.9\,\text{cm}$$

To determine the uncertainty in the calculations of gradient and intercept, we need only consider the range of straight lines that could be drawn through the first and last error bars. The uncertainty will be the maximum difference between these extreme values obtained from graphs of maximum and minimum possible gradients and the value calculated from the line of best fit. In this example it can be shown that: force constant is between 14 Ncm^{-1} and 28 Ncm^{-1}, original length is between 1.1 cm and 2.6 cm.

The final result can be quoted as:
force constant = 19 ± 9 Ncm^{-1}, original length = 1.9 ± 0.8 cm. Clearly, the large uncertainties in these results confirm that the experiment lacked precision.

Table A2.3 shows the results that a student obtained when investigating the effects of a resultant force on a constant mass. Plot a graph of these readings, including uncertainty bars. Then draw lines of maximum and minimum gradients through the error bars. Finally, use your graph to determine the mass that the student used in the experiment and the uncertainty in your answer.

■ **Table A2.3**

Resultant force, N, ± 0.5 N	**Acceleration, m s^{-2}, ± 0.2 m s^{-2}**
1.0	0.7
2.0	1.3
3.0	2.0
4.0	2.8
5.0	3.3
6.0	4.1

Common mistake

Many students believe that the force involved when an object hits the ground is its weight. In reality, the force will depend on the nature of the impact. The longer the duration of the impact, the smaller the force, as explained below.

Non-mathematical applications of Newton's second law

We can use Newton's second law to explain why, for example, a glass will break when dropped on the floor, but may survive being dropped onto a sofa. A collision with the floor will be for a much shorter duration, which means the deceleration will be greater and (using $F = ma$) the force will be greater, and probably more destructive. Similar arguments can be used to explain how forces can be reduced in road accidents.

WORKED EXAMPLE A2.8

A car of mass 1450 kg is accelerated from rest by an initial resultant force of 3800 N.

a Calculate the acceleration of the car.

b If the force and acceleration are constant, what will its speed be after 4.0 s?

c Determine how far it will have travelled in this time.

d After 4.0 s the resistive forces acting on the car are 1800 N. Show that the new force required to maintain the same acceleration is approximately 5.5 kN.

Answer

a $a = \frac{F}{m} = \frac{3800}{1450} = 2.62\,\text{m}\,\text{s}^{-2}$

b $v = u + at = 0 + (2.62 \times 4.0) = 10.5\,\text{m}\,\text{s}^{-1}$

c $s = \frac{(u + v)}{2} \times t = \frac{(0 + 10.5)}{2} \times 4.0 = 21.0\,\text{m}$

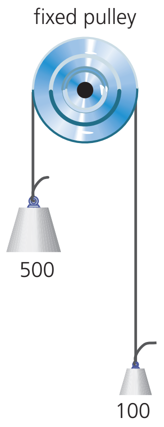
d $3800 + 1800 = 5600\,\text{N} \approx 5500\,\text{N} = 5.5\,\text{kN}$

WORKED EXAMPLE A2.9

Figure A2.53 shows two masses attached by a string which passes over a fixed pulley. Assuming that there is no friction in the system and that the string has negligible mass, determine:

a the acceleration of the system

b the tension in the string.



■ **Figure A2.53** Two masses attached by a string which passes over a fixed pulley

Answer

a The resultant force on the system of two masses = weight of the 500 g mass − weight of 100 g mass = (0.500 − 0.100) × 9.8 = 3.9 N

$$a = \frac{F}{m} = \frac{3.9}{(0.500 + 0.100)} = 6.5\,\mathrm{m\,s^{-2}}$$

The 500 g mass will accelerate down while the 100 g mass accelerates up at the same rate.

b Consider the 100 g mass: the resultant force acting = tension, T, in the string upwards – weight acting downwards = T – (0.100 × 9.8) = T – 0.98

$F = ma$

$(T - 0.98) = 0.100 \times 6.5$

$T = 1.6\,\mathrm{N}$

Equally, we could consider the 500 g mass:
the resultant force acting =
weight acting downwards − tension, T, in the string upwards =
$(0.500 \times 9.8) - T = 4.9 - T$

$F = ma$

$(4.9 - T) = 0.500 \times 6.5$

$T = 1.6\,\mathrm{N}$

36 A laboratory trolley accelerated at 80 cm s^{-2} when a resultant force of 1.7 N was applied to it. What was its mass?

37 When a force of 6.4 N was applied to a mass of 2.1 kg on a horizontal surface, it accelerated by 1.9 m s^{-2}. Determine the average frictional force acting on the mass.

38 When a hollow rubber ball of mass 120 g was dropped on a concrete floor the velocity of impact was 8.0 m s^{-1} and it reduced to zero in 0.44 s (before bouncing back).

a Calculate:

i the ball's average deceleration

ii the average force exerted on the ball.

b Repeat the calculations for a solid steel ball of the same size, 10 times the mass, but with the same impact velocity. Assume that its speed reduced to zero in 0.080 s.

c Outline why the steel ball can do more damage to a floor than the rubber ball.

39 A small aircraft of mass 520 kg needs to take off with a speed of 30 m s^{-1} from a runway in a distance of 200 m.

a Show that the aircraft needs to have an average acceleration of 2.3 m s^{-2}.

b What average resultant force is needed during the take off?

40 Discuss why the forces on the long-jumper shown in Figure A2.54 are reduced because he is landing in sand.

■ **Figure A2.54** Impact in a sand-pit reduces force

41 **a** What resultant force is needed to accelerate a train of total mass 2.78×10^6 kg from rest to 20 m s^{-1} in 60 s?

b If the same train was on a sloping track which had an angle of 5.0° to the horizontal, what is the component of its weight parallel to the track?

c Suggest why railway designers try to avoid hills.

42 Calculate the average force needed to bring a 2160 kg car travelling at 21 m s^{-1} to rest in 68 m.

43 Use Newton's second law to explain why it will hurt you more if you are struck by a hard ball than by a soft ball of the same mass and speed.

44 A trolley containing sand is pulled across a frictionless horizontal surface with a small but constant resultant force. Describe and explain the motion of the trolley if sand can fall through a hole in the bottom of the trolley.

45 A man of mass 82.5 kg is standing still in an elevator that is accelerating upwards at 1.50 m s^{-2}.

a What is the resultant force acting on the man?

b What is the normal contact force acting upwards on him from the floor?

46 Figure A2.55 shows two masses connected by a light string passing over a pulley.

a Assuming there is no friction, calculate the acceleration of the two blocks.

b What resultant force is needed to accelerate the 2.0 kg mass by this amount?

c Draw a fully labelled free-body diagram for the 2 kg mass, showing the size and direction of all forces.

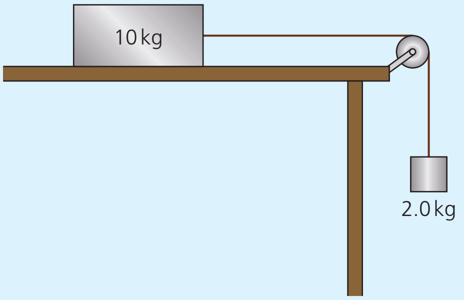


Figure A2.55 Two masses connected by a light string passing over a pulley

47 Outline how air bags (and/or seat belts) reduce the injuries to drivers and passengers in car accidents.

◆ **Inertia** Resistance to a change of motion. Depends on the mass of the object.

Newton's second law offers us a different way of understanding mass: larger masses accelerate less than smaller masses under the action of the same resultant force. So, mass can be considered as a measure of an object's resistance to acceleration. Physicists use the term **inertia** to describe an object's resistance to a change of motion.

Mass is a measure of inertia.

Newton's third law of motion

Whenever any two objects come in contact with each other, or otherwise interact, they exert forces on each other (Figure A2.56). Newton's third law compares these two forces.

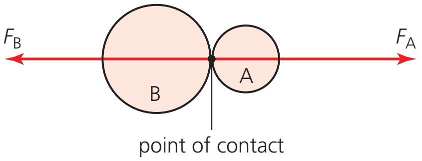


Figure A2.56 When two bodies interact, $F_A = -F_B$

Newton's third law of motion states that whenever one body exerts a force on another body, the second body exerts a force of the same magnitude on the first body, but in the opposite direction.

Essentially this law means that forces must always occur in equal pairs, although it is important to realize that the two forces must act on different bodies and in opposite directions, so that only one of each force pair can be seen in any free-body diagram. The two forces are always of the same type, for example gravity/gravity or friction/friction. Sometimes the law is quoted in the form used by Newton: 'to every action there is an equal and opposite reaction'. In everyday terms, it is simply not possible to push something that does not push back on you. Here are some examples:

- If you pull a rope, the rope pulls you.
- If the Earth pulls a person, the person pulls the Earth (Figure A2.57).
- If a fist hits a cheek, the cheek hits the fist (Figure A2.58).

- If you push on the ground, the ground pushes on you.
- If a boat pushes down on the water, the water pushes up on the boat.
- If the Sun attracts the Earth, the Earth attracts the Sun.
- If an aircraft pushes down on the air, the air pushes up on the aircraft.

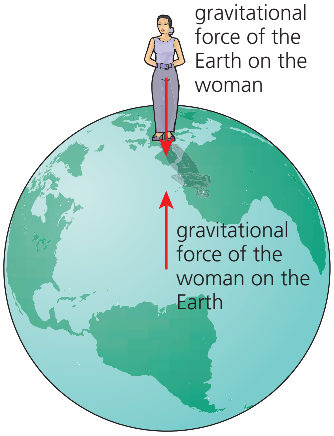


Figure A2.57 The force on the woman is equal and opposite to the force on the Earth

Figure A2.58 The force on the glove is equal and opposite to the force on the cheek

48 A book has a weight of 2 N and is at rest on a table. The table exerts a normal contact force of 2 N upwards on the book.
Explain why these two forces are *not* an example of Newton's third law.

49 Seven examples of pairs of Newton's third law forces are provided above. Give three more examples. Try to use different types of force.

50 Consider Figure A2.58. Outline reasons why forces of equal magnitude, for example on a face and on a fist, can have very different effects.

51 Discuss why the person shown in Figure A2.59 could end up in the water.

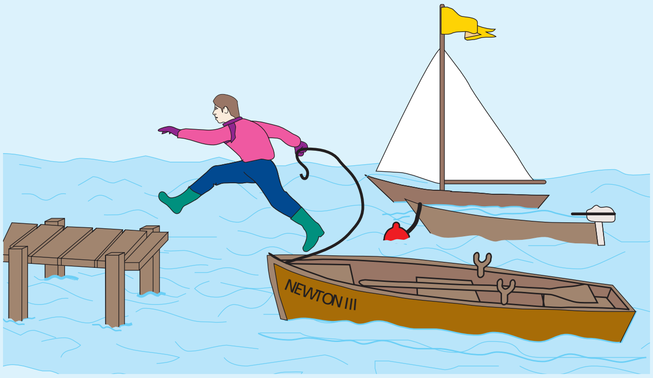


Figure A2.59

52 Figure A2.60 shows a suggestion to make a sailing boat move when there is no wind.
Discuss how effective this method could be.

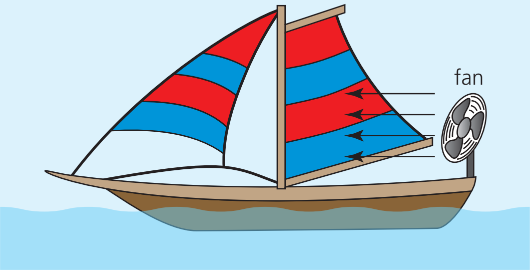


Figure A2.60 Sailing boat

53 A large cage with a small bird sitting on a perch is placed on weighing scales.
Discuss what happens to the weight shown on the scales when the bird is flying around the cage (compared with when it is sitting still).

We have seen (Figure A2.56) that, when any two objects, A and B, interact, $F_A = -F_B$.

Using Newton's second law $F = ma$ we can write: $[ma]_A = -[ma]_B$.

Remembering that:

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

we can write:

$$\left[\frac{m(v-u)}{t}\right]_A = \left[\frac{m(v-u)}{t}\right]_B$$

The time of the interaction is the same for both, so: $[m(v - u)]_A = [m(v - u)]_B$.

Putting this into words: (mass × change of velocity) for A = −(mass × change of velocity) for B.

◆ Momentum (linear), *p* Mass times velocity, a vector quantity.

(mass × velocity) is an important concept in physics. It is called **momentum**. The momentum gained by object A = momentum lost by object B. Always. This assumes that there are no external forces. This is covered in more detail in the next section.

TOK

The natural sciences

- What kinds of explanations do natural scientists offer?

A clockwork universe?

Everything is made of particles and it has been suggested that, if we could know everything about the present state of all the particles in a system (their positions, energies, movements and so on), then maybe we could use the laws of classical physics to predict what will happen to them in the future. The Universe would then behave like a mechanical clock. If these ideas could be expanded to include everything, then the future of the Universe would already be decided and predetermined, and the many apparently unpredictable events of everyday life and human behaviour (like you reading these words at this moment) would just be the laws of physics in disguise.

However, we now know that the laws of physics (as imagined by humans) are not always so precisely defined, nor as fully understood as physicists of earlier years may have believed. The principles of quantum physics and relativity in particular contrast with the laws of classical physics. Furthermore, in a practical sense, it is totally inconceivable that we could ever know enough about the present state of everything in the Universe in order to use that data to make detailed future predictions.

Momentum

SYLLABUS CONTENT

- Linear momentum as given by: $p = mv$ remains constant unless the system is acted upon by a resultant external force.
- Newton's second law in the form $F = ma$ assumes mass is constant whereas $F = \frac{\Delta p}{\Delta t}$ allows for situations where mass is changing.

DB linear momentum (SI unit: kg m s^{-1}) = mass × velocity, $\boldsymbol{p = mv}$

Momentum is a vector quantity and its direction is always important.

◆ **System** The object(s) being considered (and nothing else). An **isolated system** describes a system into which matter and energy cannot flow in, or out.

The explanation at the end of the last section shows that when two objects interact, with forces between them, the change of momentum for one object is equal and opposite to the change in momentum of the other: one object gains momentum while the other object loses an equal amount of momentum. This means that the total amount of momentum is unchanged, although this is only true if no resultant external force is acting on the objects. (We describe this as an **isolated system**.). This very important principle, which is a consequence of Newton's third law, can be stated as follows:

The law of conservation of momentum: the total momentum of any system is constant, provided that there is no resultant external force acting on it.

This law of physics is always true. There are no exceptions. It is very useful in helping to predict the results of interactions like collisions. See below.

Nature of science: Models

Systems and the environment

We use the term *system* to describe and limit the collection of objects we are considering. You may think of this as 'drawing a line around' an object together with all of the surrounding objects with which there are significant interactions. This is especially important when using conservation laws. Objects outside of the 'system' are usually referred to as the environment, or the **surroundings**.

In practice, any situation can be complicated and we often have to decide which objects we can ignore (assume to be outside of the system) because their effect is minimal.

Take a collision between two cars as an example. Commonly, we calculate an outcome by considering the system to be just the two cars. This will give us a quick, reasonably accurate and useful prediction for what happens immediately after impact. Such a calculation has chosen not to include the air and the road in the system. If they were included, the situation would be much more complex, but the immediate consequences of any collision may be similar.

◆ **Surroundings** Everything apart from the system that is being considered; similar to the 'environment'.

We know that, for uniform acceleration:

$$F = ma = \frac{m(v - u)}{t} = \frac{mv - mu}{t}$$

This demonstrates an alternative, more generalized, interpretation of Newton's second law ($F = ma$) in terms of a *change* of momentum, Δp ($= mv - mu$) that occurs in time Δt.

force = rate of change of momentum: $\boldsymbol{F} = \dfrac{\boldsymbol{\Delta p}}{\boldsymbol{\Delta t}}$

This equation allows for the possibility of a changing mass, whereas the use of $F = ma$ assumes a constant mass. An application of this is given later in the section on explosions and propulsion.

Inquiry 2: Collecting and processing data

Interpreting results

Significant figures

An answer should not have more significant figures than the least precise of the data used in the calculation.

The more precise a measurement is, the greater the number of **significant figures** (digits) that can be used to represent it. For example, a mass stated to be 4.20 g (as distinct from 4.19 g or 4.21 g) suggests a greater precision than a mass stated to be 4.2 g.

Significant figures are all the digits used in data to carry meaning, whether they are before or after a decimal point, and this includes zeros.

But sometimes zeros are used without thought or meaning, and this can lead to confusion. For example, if you are told that it is 100 km to the nearest airport, you might be unsure whether it is approximately 100 km, or 'exactly' 100 km. This is a good example of why **scientific notation** is useful. Using 1.00×10^2 km makes it clear that there are 3 significant figures. 1×10^2 km represents much less precision. When making calculations, the result cannot be more precise than the data used to produce it. As a general and simplified rule, when answering questions or processing experimental data, the result should have the same number of significant figures as the data used. If the number of significant figures is not the same for all pieces of data, then the number of significant figures in the answer should be the same as the least precise of the data (which has the fewest significant figures).

You may assume that all the digits seen in the data shown in this book are significant. For example 100 km should be interpreted as three significant figures.

For example, if a mass of 583 g changed velocity by 15 m s^{-1} in two seconds, then the resultant force acting was:

$$F = \frac{\Delta p}{\Delta t} = \frac{(0.583 \times 15)}{2} = 4.3725\,\text{N}$$

(showing all figures seen on calculator display).

But, since the time was only given to one significant figure, then the answer should have the same: $F = 4$ N.

Maybe this can seem unsatisfactory, but remember that when the time is quoted as 2 s, it simply means that it was more than 1.5 s and less than 2.5 s. (A time of 1.5 s would give an answer of 5.8 N, a time of 2.5 s would give an answer of 3.5 N.) If the time had been 2.0 s, then the quoted answer for the force should be 4.4 N. If the time had been 2.00 s, then the quoted answer for the force should still be 4.4 N, because the velocity was only given to 2 significant figures.

Conservation of momentum in collisions and explosions

SYLLABUS CONTENT

- Elastic and inelastic collisions of two bodies.
- Energy considerations in elastic collisions, inelastic collisions, and explosions.

We will use the word '**collision**' to describe any event in which two, or more, objects move towards each other and exert forces on each other. In physics this term is not limited to its typical everyday use to describe accidental events, often involving large forces. The term '**explosion**' will be used to describe any event in which internal forces within a stationary system result in separate parts of the system moving apart. Everyday usage of the term is much more dramatic.

◆ **Scientific notation** Every number is expressed in the following form: $a \times 10^b$, where a is a decimal number larger than 1 and less than 10 and b is an exponent (integer).

◆ **Significant figures (digits)** All the digits used in data to carry meaning, whether they are before or after a decimal point.

◆ **Collision** Two (or more) objects coming together and exerting forces on each other for a relatively short time.

◆ **Explosion** Term used in physics to describe when two or more masses, which were initially at rest, are propelled apart from each other.

Figure A2.61 Newton's cradle is a demonstration of collisions

Collisions

We can *always* use the law of conservation of momentum to help predict what will happen immediately after a collision, but we must also have some other information. For example, the results of two tennis balls colliding will be very different from two cushions colliding and neither can be predicted by using *only* the law of conservation of momentum. Consider the following worked example.

WORKED EXAMPLE A2.10

An object A of mass of 2.1 kg was moving to the left with a velocity of 0.76 m s^{-1}. At the same time, object B of mass 1.2 kg was moving in the opposite direction with a velocity of 3.3 m s^{-1}. Discuss what happens after the collision.

Momentum of A = 2.1 × 0.76 = 1.60 kg m s^{-1}. But this is a magnitude only. We have not considered direction:

If we choose that velocity to the left is positive, momentum of A = 2.1 × (+ 0.76) = +1.60 kg m s^{-1} (to the left).

(Alternatively, if we prefer to say velocity to the right is positive, then we get: momentum of A = 2.1 × (−0.76) = −1.60 kg m s^{-1} (to the left)

Using velocity and momentum to the left to be positive:

momentum of B = 1.2 × (−3.3) = −3.96 kg m s^{-1} (to the right)

The combined momentum of A and B before the collision = 1.60 + (−3.96) = −2.36 kg m s^{-1} (to the right)

The law of conservation of momentum tells us that after the collision, this momentum will be the same (assuming there is no resultant external force). But we need further information to determine exactly what happened. That information may come in the form of identifying the type of collision (see below), or telling us what happened to one of the objects, so that we can calculate what happened to the other, as follows:

If, after the collision, object A moved to the left with a velocity of 0.87 m s^{-1}, what happened to object B?

After the collision, momentum of A + momentum of B = −2.36 kg m s^{-1}

$[2.1 \times (-0.87)] + (1.2 \times v_B) = -2.36\,\text{kg m s}^{-1}$

$v_B = -0.44\,\text{m s}^{-1}$ (to the left)

All of this has been explained in detail to help understanding. More directly it can be represented by: momentum before collision = momentum after collision

$[2.1 \times (+0.76)] + [1.2 \times (-3.3)] = [2.1 \times (-0.87)] + (1.2 \times v_B)$

$v_B = -0.44\,\text{m s}^{-1}$ (to the left)

Top tip!

In Topic A.3, we will introduce the law of conservation of energy and the concept of kinetic energy, which is the energy of moving masses, calculated by $E_k = \frac{1}{2}mv^2$. That knowledge is needed in order to understand the rest of this section on collisions.

You may prefer to delay the rest of this topic on collisions and explosions until Topic A.3 has been covered in detail.

Kinetic energy in collisions

We need to consider the transfer of energy in a collision. Any moving object has kinetic energy and during a collision some, or all, of this energy will be transferred between the colliding objects. Typically, some energy, perhaps most of the energy, will be transferred to the surroundings as thermal energy and maybe some sound. We can identify the extreme cases:

◆ **Collisions** In an **elastic collision** the total kinetic energy before and after the collision is the same. In any **inelastic collision** the total kinetic energy is reduced after the collision. If the objects stick together it is described as a **totally inelastic collision**.

◆ **Macroscopic** Can be observed without the need for a microscope.

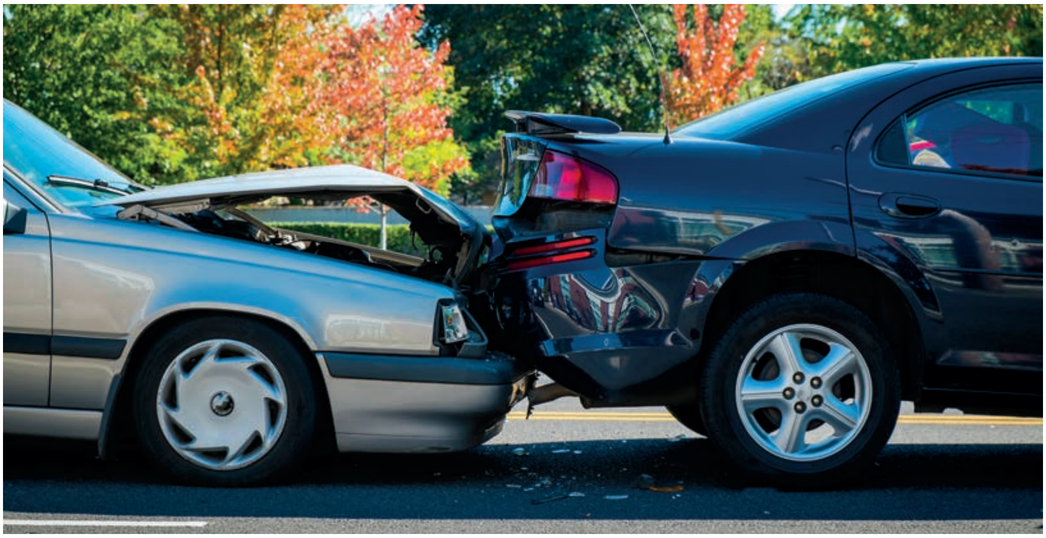
◆ **Microscopic** Describes anything that is too small to be seen with the unaided eye.

A collision in which the *total* kinetic energy before and after the collision is the same is called an **elastic collision.**

All other collisions can be described as **inelastic collisions**, meaning that kinetic energy has not been conserved. In everyday, **macroscopic** events, elastic collisions are a theoretical ideal and they do not happen perfectly. However, elastic collisions are common for **microscopic** particle collisions.

A collision after which the colliding objects stick together is called a **totally inelastic collision**.

In a totally inelastic collision, the maximum possible amount of kinetic energy is transferred from the moving objects to the environment.



■ **Figure A2.62** An inelastic collision

Nature of science: Models

Macroscopic and microscopic

In general, physicists use the terms:

- **macroscopic** to describe events that can be observed with the unaided eye
- **microscopic** to describe events on the molecular, atomic, or subatomic scale.

It was not until scientists began to realize that matter consisted of atoms and molecules (that could not be seen), that many observations of the world around us could be explained.

Perhaps the best example of an (almost) elastic macroscopic collision is that between steel spheres. If a stationary sphere is struck by an identical moving sphere, the moving sphere stops and the other sphere continues with the velocity of the first. 'Newton's cradle', as seen in Figure A2.61 is a famous demonstration of this.

It is easy to find examples where *all* kinetic energy appears to have been lost, for example, when a student jumps down to the floor. The student has had a totally inelastic collision with the Earth, but the change to the motion of the Earth is insignificant and unobservable.

WORKED EXAMPLE A2.11

A 2.1 kg trolley moving at 0.82 m s^{-1} collides with a 1.7 kg trolley moving at 0.98 m s^{-1} in the opposite direction. After the collision, the 1.7 kg trolley reverses direction and moves at 0.43 m s^{-1}.

a Discuss what happened to the other trolley.

b Without making any calculations, comment on the difference in total kinetic energy before and after the collision.

c If the collision had been totally inelastic, what would have happened after the collision?

Answer

a Total momentum before = total momentum after

$(2.1 \times 0.82) + (1.7 \times -0.98) = (2.1 \times v) + (1.7 \times 0.43)$

Velocities in the original direction have been given a + sign and velocities in the opposite direction are given a − sign (or it could be the other way around).

$0.056 = 2.1v + 0.731$

$v = -0.32$ m s^{-1}. The – sign shows us that the trolley reverses its direction of motion.

b Both velocities have been reduced, so the total kinetic energy is significantly less.

c The trolleys will stick together if the collision is totally inelastic.
Total momentum before = total momentum after

$(2.1 \times 0.82) + (1.7 \times -0.98) = (2.1 + 1.7) \times v$

$0.056 = 3.8v$

$v = +0.015$ m s^{-1}. The + sign shows us that they move in the original direction of the 2.1 kg trolley.

The combined trolleys are moving slowly, so there has been a considerable loss of kinetic energy.

LINKING QUESTION

- In which way is conservation of momentum relevant to the workings of a nuclear power station?

This question links to understandings in Topic E.4

54 An object of mass 4.1 kg travelling to the right with a velocity of 1.9 m s^{-1} has a totally inelastic collision with a stationary object of mass 5.6 kg. Determine how they move immediately after the collision.

55 A bus of mass 4900 kg travelling at 22 m s^{-1} collides with the back of a 1300 kg car travelling at 16 m s^{-1}. If the car is pushed forward with a velocity of 20 m s^{-1}, calculate the velocity of the bus immediately after the collision.

56 In an experiment to find the speed of a 2.40 g bullet, it was fired into a 650 g block of wood at rest on a friction-free surface. If the block (and bullet) moved off with an initial speed of 96.0 cm s^{-1}. Calculate the speed of the bullet.

57 A ball thrown vertically upwards decelerates and its momentum decreases, although the law of conservation of momentum states that total momentum cannot change. Explain this observation.

58 Figure A2.63 shows two trolleys on a friction-free surface joined together by a thin rubber cord under tension. When the trolleys are released, they accelerate towards each other and the cord quickly becomes loose.

a Show that the two trolleys collide at the 20 cm mark.

b Predict what happens after they collide if they stick together.

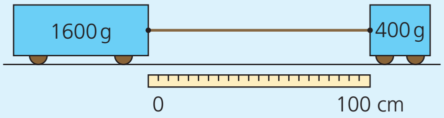


Figure A2.63 Two trolleys on a friction-free surface joined together by a thin rubber cord

59 Two toy cars travel in straight lines towards each other on a friction-free track. Car A has a mass of 432 g and a speed of 83.2 cm s^{-1}. Car B has a mass of 287 g and speed of 68.2 cm s^{-1}. If they stick together after impact, predict their combined velocity.

60 A steel ball of mass 1.2 kg moving at 2.7 m s^{-1} collides head-on with another steel ball of mass 0.54 kg moving in the opposite direction at 3.9 m s^{-1}. The balls bounce off each other, each returning back in the direction it came from on a horizontal surface.

a If the smaller ball had a speed after the collision of 6.0 m s^{-1}, use the law of conservation of momentum to predict the speed of the larger ball.

b In fact, the situation described in part **a** is not possible. Discuss possible explanations of why not.

ATL A2C: Thinking skills

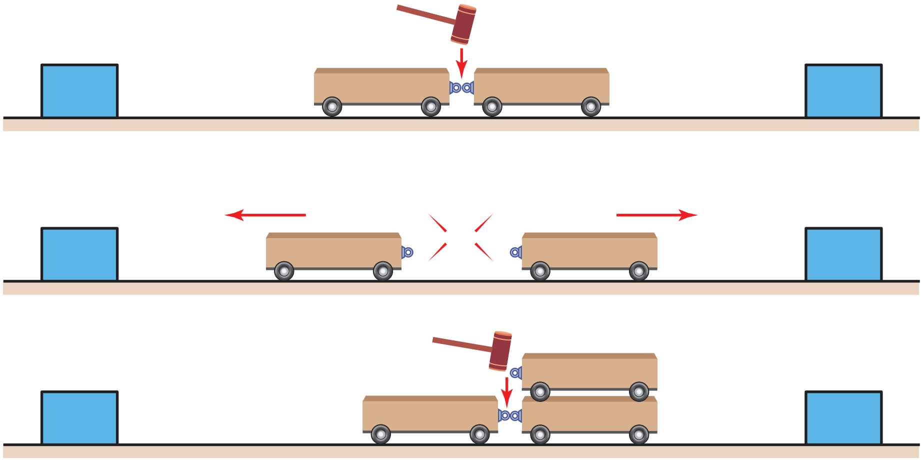
Reflecting on the credibility of results

A student carried out an experiment into the momentum of colliding trolleys on a horizontal runway. A trolley of mass 2.0 kg and speed 80 cm s^{-1} collided with a trolley of mass 1.0 kg and speed 220 cm s^{-1} travelling in the opposite direction. After the collision, both the trolleys reversed their directions and the student measured the speeds of both trolleys to be 60 cm s^{-1}.

Explain why the student must have made a mistake.

Explosions

Figure A2.64 shows a possible laboratory investigation into a one-dimensional ‘explosion’. A blow from the hammer releases springs which push the previously stationary trolleys apart. If the trolleys are identical, they will move apart with equal speeds. If the mass on one side is doubled, as shown in the third drawing, the speeds will be in the ratio 2:1, the more massive trolley will move more slowly



■ **Figure A2.64** A simple ‘explosion’

■ **Figure A2.65** Firing a cannon

◆ **Recoil** When a bullet is fired from a gun (or similar), the gun must gain equal momentum in the opposite direction.

Firing a gun, or a cannon, is a more dramatic example. See Figure A2.65. Since there is zero momentum to begin with, the momentum of the bullet / cannon ball must be equal and opposite to the momentum of the gun itself (or cannon), so that the total momentum after firing is also zero. The word **recoil** is used to describe this ‘backwards’ motion.

WORKED EXAMPLE A2.12

A rifle of mass 1.54 kg fires a bullet of mass 22 g at a speed of 250 m s^{-1}. Calculate the recoil speed of the rifle.

Answer

Total momentum before = total momentum after

$0 = (1.54 \times v) + (0.022 \times 250)$

$v = -3.6\,\text{m s}^{-1}$.

The '–' sign shows us that the gun moves in the opposite direction to the bullet's velocity.

61 In an experiment similar to that shown in the first drawing of Figure A2.64, after being released one trolley with a mass of 980 g recoiled with a velocity of 0.27 m s^{-1}. What was the speed of the other trolley, which had a mass of 645 g?

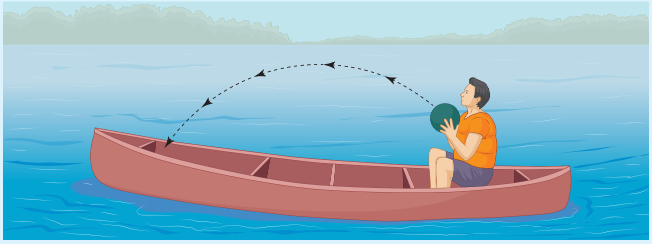
62 An isolated and 'stationary' astronaut of mass 65 kg accidentally pushes a 2.3 kg hammer away from her body with a speed of 80 cm s^{-1}.

- **a** Outline a reason why the word 'stationary' has been put in quotation marks.
- **b** Predict what happened to the astronaut.
- **c** Suggest how she can stop moving.

63 Cannons have been used extensively in wars for hundreds of years. A large cannon from 200 years ago could fire a 25 kg cannon ball with a speed of over 150 km h^{-1} (42 m s^{-1}). The recoil momentum of these cannons could be dangerous, although the recoil speed was limited by the very large mass of the cannon. If the speed of recoil was 0.30 m s^{-1}, calculate the mass of the cannon.

64 Figure A2.66 shows a heavy ball being thrown from one end of a canoe to the other. Describe what will happen to the canoe (and the passenger) when:

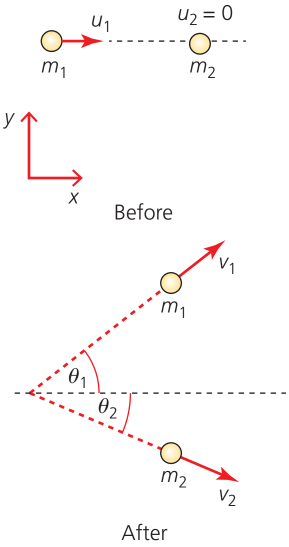
- **a** the ball is being thrown
- **b** the ball is in the air
- **c** the ball lands back in the canoe. Assume that the water does not resist any possible movement of the canoe.



■ **Figure A2.66** A heavy ball being thrown from one end of a canoe to the other

HL ONLY

Collisions and explosions in two dimensions



■ **Figure A2.67** Two objects colliding in two dimensions

So far, we have only considered interactions in one direction. This section extends the study to two dimensions. *It is aimed at Higher Level students only.* We will consider that the interacting objects behave as point particles.

Figure A2.67 shows two objects, m_1 and m_2, before and after a collision. In this example m_2 was stationary before the collision.

For collisions or explosions in two dimensions the law of conservation of momentum can be applied in two perpendicular directions.

For the example shown in Figure A2.67 we need to know the components of v_1 and v_2 in the x and y directions:

$$v_{1x} = v_1 \cos\theta_1 \qquad v_{1y} = v_1 \sin\theta_1$$

$$v_{2x} = v_2 \cos\theta_2 \qquad v_{2y} = v_2 \sin\theta_2$$

WORKED EXAMPLE A2.13

Considering Figure A2.67, let $m_1 = 0.50\,\text{kg}$, $m_2 = 0.30\,\text{kg}$ and $u_1 = 4.0\,\text{m s}^{-1}$.

If the angles were $\theta_1 = 36.9°$ and $\theta_2 = 26.6°$, determine the value of v_2 if v_1 was $2.0\,\text{m s}^{-1}$.

Answer

Applying conservation of momentum in the y-direction:

$0 = m_1v_1 \sin\theta_1 + m_2v_2 \sin\theta_2$

$0 = 0.600 + 0.134v_2$

$v_2 = -4.5\,\text{m s}^{-1}$, the negative sign shows that it is moving in the negative y-direction.

Alternatively, and less simply, we can apply conservation of momentum in the x-direction:

$m_1u_1 = m_1v_1 \cos\theta_1 + m_2v_2 \cos\theta_2$

$(0.50 \times 4.0) = (0.50 \times 2.0 \times \cos 36.9°) + (0.30 \times v_2 \times \cos 26.6°)$

$2.0 = 0.800 + (0.268 \times v_2)$

$v_2 = 4.5\,\text{m s}^{-1}$, as before.

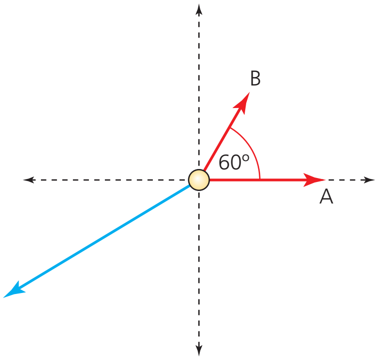
WORKED EXAMPLE A2.14

A stationary small mass of mass 5.0 g explodes into three particles.

One particle, A, of mass 1.5 g moves with a velocity of $23\,\text{m s}^{-1}$ in a direction that we will call the x-direction.

The second particle, B, of mass 2.0 g moves with a velocity of $18\,\text{m s}^{-1}$ in a direction of 60° to the first.

Determine what happened to the third particle, C.



■ **Figure A2.68** A stationary small mass explodes into three particles

Answer

Before the explosion there is zero momentum. After the explosion, in any chosen direction momentum must be conserved: $p_A + p_B + p_C = 0$.

Resolving the velocity of the B particle into two components:

In x-direction, $v_x = 18 \cos 60° = 9.0\,\text{m s}^{-1}$

In y direction, $v_y = 18 \sin 60° = 15.6\,\text{m s}^{-1}$

Consider momentum in the x-direction: $p_A + p_B + p_C = 0$.

$(1.5 \times 23) + (2.0 \times 9.0) = 52.5\,\text{g m s}^{-1} = -p_C$

Consider momentum in the y-direction: $0 + (2.0 \times 15.6) = 31.2\,\text{g m s}^{-1} = -p_C$

Dividing momentum by mass ($m = 5.0 - 1.5 - 2.0 = 1.5$) gives us the two components of the velocity of C: $-35\,\text{m s}^{-1}$ and $-21\,\text{m s}^{-1}$.

These two components can be added (using a scale drawing or trigonometry) to determine the actual velocity of C: $41\,\text{m s}^{-1}$ at an angle of 31° to the $-x$-direction.

65 Masses of 200 g and 500 g are travelling directly towards each other with speeds of $1.2\,\text{m s}^{-1}$ and $0.30\,\text{m s}^{-1}$, respectively. After they collide, the speed of the 200 g mass reduces to $0.10\,\text{m s}^{-1}$ as it continues in a direction at 30° to its original motion.
Determine what happened to the other mass.

66 A mass of mass 1.0 kg moving at $2.0\,\text{m s}^{-1}$ explodes into three parts. One part, which has a mass of 250 g, has a velocity of $8.5\,\text{m s}^{-1}$ in the original direction of motion. The second part has a mass of 450 g and moves with a velocity of $5.6\,\text{m s}^{-1}$ at an angle of 90° to the first part. Show that the third part has a speed of $8.4\,\text{m s}^{-1}$.

Propulsion

◆ **Propel** Provide a force for an intended motion.

◆ **Jet engine** An engine that achieves propulsion by emitting a fast-moving stream of gas or liquid in the opposite direction from the intended motion.

◆ **Rocket engine** Similar to a jet engine, but there is no air intake. Instead, an oxidant is carried on the vehicle, together with the fuel.

If the ball shown in Figure A2.66 had been thrown over the end of the canoe, the canoe would keep moving to the right (until resistive forces stopped it). An unusual example perhaps, but this shows us a very useful concept: to start, or maintain motion (**propel**), we can create momentum in the opposite direction. This can be restated using Newton's third law: if we want a force to move us to the right (for example), we exert a force on the surroundings to the left. The person in the boat pushes the ball to the left and the ball pushes the person (and the boat) to the right. Using friction for walking and car movement has already been discussed.

A boat can be pushed forward by pushing water backwards, using an oar, or a propeller. See Figure A2.69. The momentum of the boat forwards is equal and opposite to the momentum of the water backwards.

A propeller can also be used for a small airplane, but typically the propeller needs to be much larger and rotate faster, because the density of air is much less than water. Larger aircraft use the same conservation of momentum principle, but in a different way:

There are many designs of **jet engines**, but the basic principle is that they take in the surrounding air and use it to burn vaporized fuel. The resulting hot exhaust gases are ejected at the back of the engine with considerable momentum (much greater than the momentum of the air input), the difference is equal and opposite to the forward momentum given to the aircraft.

Rocket engines use the same principle, but they travel where there is little or no air, so they use oxygen that has been stored on the vehicle.

■ **Figure A2.69** Boat propeller

■ **Figure A2.70** A Chinese rocket launching a spacecraft to Mars

WORKED EXAMPLE A2.15

A rocket is ejecting exhaust gases at a rate of $1.5 \times 10^4\,\text{kg s}^{-1}$. If the speed of the exhaust gases (relative to the rocket) is $2.3 \times 10^3\,\text{m s}^{-1}$, what is the forward force acting on the rocket?

Answer

$$F = \frac{\Delta p}{\Delta t} \Rightarrow \left(\frac{\Delta m}{\Delta t}\right) \times v = (1.5 \times 10^4) \times (2.3 \times 10^3) = 3.5 \times 10^7\,\text{N}$$

67 A rocket's mass at lift-off was $2.7 \times 10^6\,\text{kg}$. If gases were ejected at a rate of $1.9 \times 10^4\,\text{kg s}^{-1}$ with a speed of $2.0 \times 10^3\,\text{m s}^{-1}$:

a determine the initial acceleration of the rocket

b explain why the acceleration will increase as the rocket rises, while the engines provide the same force.

Forces acting for short times: impulses

SYLLABUS CONTENT

- A resultant force applied to a system constitutes an impulse, J, as given by: $J = F\Delta t$, where F is the average resultant force and Δt is the time of contact.
- The applied external impulse equals the change in momentum of the system.

◆ **Impulse** The product of force and the time for which the force acts.

Many forces only act for a short time, Δt. Clearly the longer the time for which a force acts, the greater its possible effect, so the concept of **impulse**, J, becomes useful:

impulse, $\boldsymbol{J = F\Delta t}$ SI unit: N s

If a force varies during an interaction, we can use an average value to determine the impulse.

We have seen that $F = \dfrac{\Delta p}{\Delta t}$ which can be rearranged to give $F\Delta t = \Delta p$, showing us that

impulse, $J = \Delta p$ (change of momentum)

An impulse on an isolated object results in a change of momentum, which is numerically equal to the impulse. The units N s and kg m s^{-1} are equivalent to each other.

WORKED EXAMPLE A2.16

A constant force of 12.0 N acts on a stationary mass of 0.620 kg for 0.580 s.

a Calculate the impulse applied to the mass.

b Calculate the change of momentum of the mass.

c Calculate the final velocity of the mass.

Answer

a $J = F\Delta t = 12.0 \times 0.580 = 6.96\,\text{Ns}$

b $6.96\,\text{kg m s}^{-1}$ (or Ns)

c $\Delta p = m\Delta v$

$6.96 = 0.620 \times \Delta v$

$\Delta v = 11.2\,\text{m s}^{-1}$

The final velocity could also be determined by use of $F = ma$ and $v = u + at$.

Force–time graphs

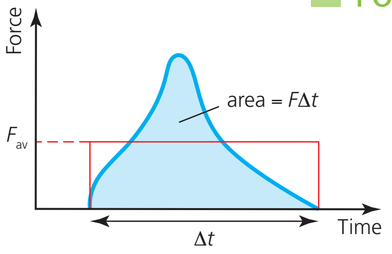


Figure A2.71 Graph showing how a force varies with time

In many simple impulse calculations, we may assume that the forces involved are constant, or that the average force is half of the maximum force. For more accurate work this is not good enough, and we need to know in detail how a force varies during an interaction. Such details are commonly represented by force–time graphs. The curved line in Figure A2.71 shows an example of a force varying over a time Δt.

The area under any force–time graph for an interaction equals force × time, which equals the impulse (change of momentum).

This is true whatever the shape of the graph. The area under the curve in Figure A2.71 can be estimated by drawing a rectangle of the same area (as judged by eye), as shown in red. F_{av} is then the average force during the interaction.

Inquiry 2: Collecting and processing data

Applying technology to collect data

Force sensors that can measure the magnitude of forces over short intervals of time can be used with data loggers to gather data and draw force–time graphs for a variety of interactions, both inside and outside a laboratory. Stop-motion replay of video recordings of collisions can also be very interesting and instructive.

Force–time graphs can be helpful when analysing any interaction, but especially impacts involved in road accidents and sports.

WORKED EXAMPLE A2.17

Figure A2.72 shows how the force on a 57 g tennis ball moving at 24 m s^{-1} to the right varied when it was struck by a racket moving in the opposite direction.

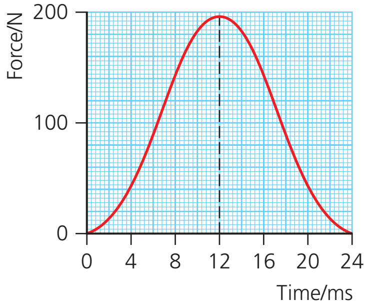


Figure A2.72 Force–time graph for striking a tennis ball

- **a** Estimate the impulse given to the ball.
- **b** Calculate the velocity of the ball after being struck by the racket.
- **c** The ball is struck with the same force with different rackets. Explain why a racket with looser strings could return the ball with greater speed.
- **d** Suggest a disadvantage of playing tennis with a racket with looser strings.

Answer

a Impulse = area under graph $\approx 200 \times (12 \times 10^{-3}) =$ 2.4 Ns to the left

b $m\Delta v = 2.4$

$$\Delta v = \frac{2.4}{0.057} = 42 \text{ m s}^{-1} \text{ to the left}$$

$v_{final} - v_{initial} = -42$ (velocity to the left chosen to be negative)

$v_{final} - (+24) = -42$

$v_{final} = -18 \text{ m s}^{-1}$ (to the left)

c The time of contact with the ball, Δt, will be longer with looser strings, so that the same force will produce a greater impulse (change of momentum).

d There is less control over the direction of the ball.

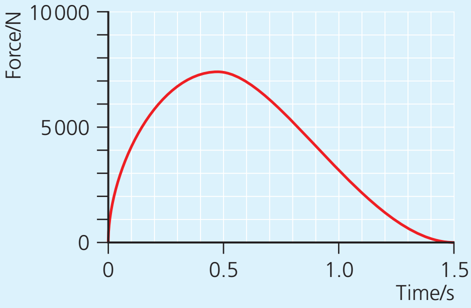
68 A ball of mass 260 g falls vertically downwards and hits the ground with a speed of 7.3 m s^{-1}.

- **a** What was its greatest momentum?
- **b** If it rebounded with a speed of 5.5 m s^{-1}, calculate the change of momentum.
- **c** Determine the impulse on the ground.
- **d** If the duration of the impact was 0.38 s, calculate the average force on the ball during the collision.
- **e** Estimate the maximum force on the ball.

69 A baseball bat hits a ball with an average force of 970 N that acts for 0.0088 s.

- **a** What impulse is given to the ball?
- **b** What is the change of momentum of the ball?
- **c** The ball was hit back in the same direction that it came from. If its speed before being hit was 32 m s^{-1}, calculate its speed afterwards. (Mass of baseball is 145 g.)

70 Figure A2.73 shows how the force between two colliding cars changed with time. Both cars were driving in the same direction and after the collision they did not stick together.



■ **Figure A2.73** How the force between two colliding cars changed with time

- **a** Show that the impulse was approximately 6500 Ns.
- **b** Before the collision the faster car (mass 1200 kg) was travelling at 18 m s^{-1}. Estimate its speed immediately after the collision.

71 Consider Figure A2.74.

- **a** Discuss how the movement of the karate expert can maximize the force exerted on the boards.
- **b** What features of the boards will help to make this an impressive demonstration?

■ **Figure A2.74** Karate expert

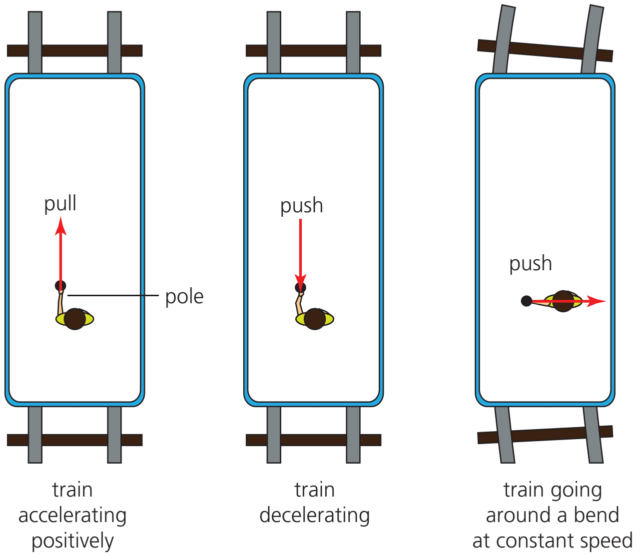
72 A soft ball, A, of mass 500 g is moving to the right with a speed of 3.0 m s^{-1} when it collides with another soft ball, B, moving to the left. The time of impact is 0.34 s, after which ball A rebounds with a speed of 2.0 m s^{-1}.

- **a** What was the change of velocity of ball A?
- **b** What was the change of momentum of ball A?
- **c** Calculate the average force exerted on ball A.
- **d** Sketch a force–time graph for the impact.
- **e** Add to your sketch a possible force–time graph for the collision of hard balls of similar masses and velocities.
- **f** Suggest how a force–time graph for ball B would be different (or the same) as for ball A.

Circular motion and centripetal forces

SYLLABUS CONTENT

- ▶ Circular motion is caused by a centripetal force acting perpendicularly to the velocity.
- ▶ A centripetal force causes the body to change direction even if the magnitude of its velocity may remain constant.



■ **Figure A2.75** Forces which make a passenger accelerate in a train

An object moving along a circular path with a constant speed has a continuously changing velocity because its *direction* of motion is changing all the time. From Newton's first law, we know that any object that is not moving in a straight line must be accelerating and, therefore, it must have a resultant force acting on it, even if it is moving with a constant speed.

Perfectly uniform motion in complete circles may not be a common everyday observation, but the theory for circular motion can also be used with objects, such as people or vehicles, moving along arcs of circles and around curves and bends. Circular motion theory is also very useful when discussing the orbits of planets, moons and satellites. It is also needed to explain the motion of subatomic particles in magnetic fields, as discussed in Topic D.3.

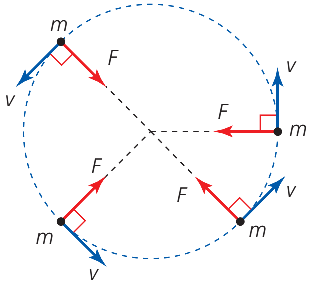
◆ **Centripetal force** The name given to any force which results in motion along a circular path.

Imagine yourself to be standing in a train on a slippery floor, holding on to a post (Figure A2.75). While you and the train are travelling in a straight line with a constant speed (constant velocity) there is no resultant force acting on you and you do not need to hold on to the post, but as soon as the train changes its motion (accelerates in some way) there needs to be a resultant force on you to keep you in the same place in the train. If there is little or no friction with the floor, the post is the only thing that can exert a force on you to change your motion. The directions of these forces (from the post) are shown in the diagram for different types of acceleration. If the post pushes or pulls on you, then by Newton's third law you must be pushing or pulling on the post, and that is the force you would be most aware of.

In particular, note that the direction of the force needed to produce a curved, circular path is *perpendicular* to the motion.

The term **centripetal force** is used to describe any type of force which results in motion in a circle (or part of a circle). See Figure A2.76.

A centripetal force continuously changes direction so that it is always acting perpendicularly to the instantaneous velocity.



■ **Figure A2.76** Velocity and centripetal force vectors during circular motion

Identifying different types of centripetal force

Gravity provides the centripetal force for planets moving around the Sun, and for satellites moving around the Earth (including the Moon).

Tension provides the centripetal force for an object being spun around on a string in an (almost) horizontal circle.

Friction provides the centripetal force for a vehicle, cyclist or a person, moving in a curved path on a horizontal surface. As an example, consider the cyclist shown in Figure A2.77. To move in a curved path there needs to be a centripetal force perpendicular to his motion. This is provided by friction: The cyclist leans 'into the bend' so that the tyre pushes outwards on the ground and the ground pushes inwards on the tyre (another example of Newton's third law).

◆ **Banked track** A sloping surface to enable faster motion around curves.

If a greater speed is desired for movement around a curved track, friction may not be enough. By having a **banked track** greater speeds are possible (and safer). See Figure A2.78. A component of the contact force can then act in the necessary direction.

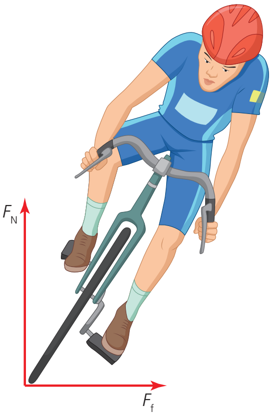


Figure A2.77 Cycling around a bend

Figure A2.78 Banked track at Daytona 500 race

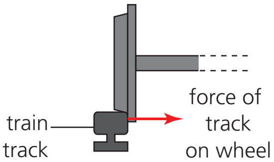


Figure A2.79 The contact force of the train track pushes inwards on the wheel of a train moving in a circular path

A train can travel around a curved track because of the contact force acting on the rim of the wheel. See Figure A2.79.

A driver or passenger in a car can move in a curved path because of the forces between them and the seat.

An aircraft can change direction by tilting, so that the air pushes the airplane perpendicular to its motion (Figure A2.80).

Figure A2.80 Aircraft changing direction

*Electrical force*s provide the centripetal force for electrons moving around the nuclei of atoms.

Common mistake

Centripetal force is *not* a different type of force, like for example, tension or gravity. It is simply a way of describing the results of a force. Centripetal force should not be labelled as such in a free-body diagram.

It is common for people to refer to *centrifugal* forces, but this will only lead to confusion and the term is best avoided in this course. It is a matter of point of view (frame of reference): if a system is seen from 'outside', a centripetal force is needed for circular motion, but 'inside' the system an object seems to experience a force moving it outwards from its circular path.

LINKING QUESTION

- Why is no work done on a body moving along a circular trajectory?

This question links to understandings in Topic A.3.

73 Draw a free-body diagram for an aircraft changing direction ('banking') at constant altitude. Ignore air resistance for this question.

74 Consider a cyclist on a horizontal curved track. State three factors which will result in the need for a greater centripetal force.

75 If you were a passenger in a car going 'too fast' around a bend, outline what you would do to exert more centripetal force on yourself.

76 Draw a free-body diagram for a car on a banked curved surface.

Centripetal acceleration

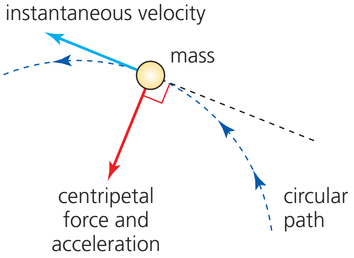


Figure A2.81 Centripetal force and acceleration

◆ **Centripetal acceleration** The constantly changing velocity of any object moving along a circular path is equivalent to an acceleration towards the centre of the circle.

We know that a resultant force causes an acceleration, *a*. Therefore, a centripetal force towards the centre of any circular motion must result in a **centripetal acceleration**, also towards the centre of the circle. This is shown more clearly in Figure A2.81. Although there is an acceleration directed towards the centre, there is no movement in that direction, or change in the magnitude of the velocity of the mass. Instead, the action of the force continually changes the direction of the motion of the mass.

Remember that acceleration means a change of velocity, and the velocity of a mass can change by going faster, going slower, or *changing direction*.

Any body moving in a circular path has a centripetal acceleration towards the centre of the circle.

The mathematics of uniform circular motion

SYLLABUS CONTENT

- Motion along a circular trajectory can be described in terms of the angular velocity, ω, which is related to the linear speed, v, by the equation as given by: $v = \frac{2\pi r}{T} = \omega r$.
- Bodies moving along a circular trajectory at a constant speed experience an acceleration that is directed radially towards the centre of the circle – known as a centripetal acceleration as given by: $a = \frac{v^2}{r} = \omega^2 r = \frac{4\pi^2 r}{T^2}$.

Tool 3: Mathematics

Use of units whenever appropriate: radians

◆ **Radians** Unit of measurement of angle. There are 2π radians in 360°.

In physics, it is usually much easier in calculations to use angles measured in **radians**, rather than degrees (which are based on the historical and arbitrary choice of 360 degrees for a complete circle). If you are studying Mathematics: Applications and Interpretations Standard Level, this may be a new concept for you.

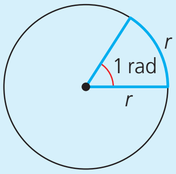


Figure A2.82 One radian

One radian (rad) is defined as the angle which has a length of arc equal to the radius of the circle. (See Figure A2.82.) Rotation through a complete circle passes through an angle of $\frac{2\pi r}{r} = 2\pi$ rad, so that:

$$1 \text{ rad} = \frac{180°}{\pi} (= 57.3°)$$

Consider an object of mass *m* moving with a constant linear speed, *v*, in a circle of radius *r*, as shown in Figure A2.83. The linear speed, *v*, of the mass can be calculated from:

$$v = \frac{2\pi r}{T}$$

where *T* is called the **time period** of the repeating motion, the time taken for one complete rotation. SI unit: s.

◆ **Time period, *T*** The duration of an event which occurs regularly. $T = \frac{1}{f}$.

◆ **Frequency, *f*** The number of repeating events per unit time.

◆ **Hertz, Hz** Derived SI unit of measurement of frequency. 1 Hz = one event per second.

◆ **Angular velocity, *ω*** Change of angle / change of time. Sometimes called angular speed.

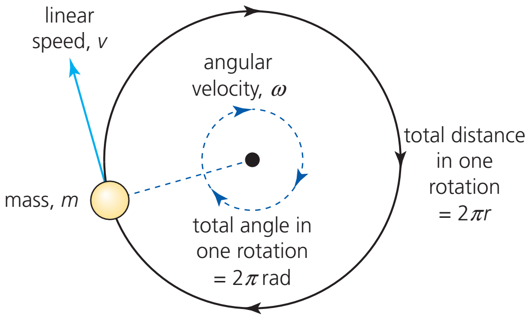


Figure A2.83 Relating linear speed to angular velocity

The **frequency**, ***f***, of the motion is the number of rotations in unit time (per second). SI unit: **hertz**, **Hz**. A frequency of 1 Hz means one rotation per second.

$$f = \frac{1}{T} \quad \text{SI unit: hertz, Hz}$$

We also commonly refer to **angular velocity**, ***ω***, the rate at which an object rotates. In the context of uniform circular motion, the vector nature of a constant angular velocity is not important. It is also sometimes called angular speed.

Angular velocity = angle moved through / time taken. It can be measured in degrees per second, so that a constant angular velocity in degrees per second would be 360 / *T*. However, the use of radians per second is considered more convenient.

$$\omega = \frac{2\pi}{T} = 2\pi f \quad \text{SI unit: rad s}^{-1}$$

Comparing the last equation with $v = \frac{2\pi r}{T}$, it should be clear that:

$$v = \omega r$$

Top tip!

Period, frequency and angular velocity represent exactly the same information about a constant circular motion. (Given any one, we can calculate the other two.) In a question, we are most likely to be told the period, or the frequency of a rotation, but in calculations the angular velocity is often needed.

WORKED EXAMPLE A2.18

A car is travelling at a constant speed of 12 m s^{-1}. Its wheels each have a radius (including tyres) of 26 cm.

a What is the linear speed of a point on the surface of the tyre?

b Calculate the frequency and time period of the wheel's rotation.

c Determine the angular velocity of the wheels.

d Through what total angle does the wheel rotate in 10 seconds in:

  i radians

  ii degrees?

Answer

a 12 m s^{-1}

b $f = \frac{12}{2\pi r} = \frac{12}{(2 \times \pi \times 0.26)} = 7.3\,\text{Hz}$ (7.3456... seen on calculator display)

$T = \frac{1}{f} = \frac{1}{7.3456} = 0.14\,\text{s}$

c $\omega = 2\pi f = 2 \times \pi \times 7.3456 = 46\,\text{rad s}^{-1}$ (46.1538... seen on calculator display)

d i $46.1538 \times 10 = 4.6 \times 10^2$ rad

ii $461.538 \times 57.3 = 2.6 \times 10^{4\circ}$

77 a Convert an angle of 157° to radians.

b i How many degrees does a rotating object pass through in five complete rotations?

ii How many radians does a rotating object pass through in five complete rotations?

78 The diameter of the clock face seen in Figure A2.84 is 43 m.

Figure A2.84 The clock face on the Abraj Al-Bait Tower in Mecca is the largest in the world

a Determine the linear speed of the tip of the minute hand.

b What is the angular velocity of the minute hand?

79 If a rotating object completes 30.0 rotations in 47.4 s, calculate:

a its time period

b its frequency

c its angular velocity.

80 a Calculate the angular velocity of the Earth's motion around the Sun.

b What is your angular velocity as you rotate on the Earth's surface?

c Determine the linear speed of someone on the equator spinning on the Earth's surface.
(radius of Earth = 6.4×10^6 m)

81 A bicycle wheel which has a radius of 31 cm is rotating with an angular velocity of 41.9 rad s^{-1}.

a Calculate the linear speed of a point on the circumference of the wheel.

b What is the speed of the bicycle along the road?

Equations for centripetal acceleration and force

Even if an object moving in a circle of radius r has a constant linear speed, v, its centripetal acceleration, a, will have a numerical value, which represents how quickly the object's direction of motion is changing. The equation for calculating centripetal acceleration is shown below, and its derivation is included below.

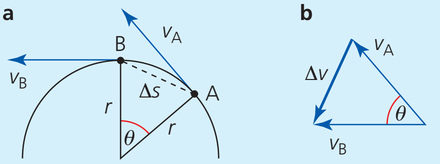
$$\boldsymbol{a = \frac{v^2}{r} = \omega^2 r = \frac{4\pi^2 r}{T^2}}$$

Tool 3: Mathematics

Derive relationships algebraically

When basic principles of physics are used to explain the origin of an equation it is called **deriving** the equation.

Consider a mass moving in a circular path of radius, r, as shown in Figure A2.85a. It moves through an angle, θ, and a distance, Δs, along the circumference as it moves from A to B, while its velocity changes from v_A to v_B.



■ **Figure A2.85** Deriving an equation for centripetal acceleration

To calculate acceleration, we need to know the change of velocity, Δv. This is done using the vector diagram shown in Figure A2.85b. Note that the direction of the change of velocity (and therefore the acceleration) is towards the centre of motion. The two triangles are similar and, if the angle is small enough that Δs can be approximated to a straight line, we can write:

$$\theta = \frac{\Delta v}{v} = \frac{\Delta s}{r}$$

(The magnitudes of v_A and v_B are equal and represented by the speed, v.)

Dividing both sides of the equation by Δt we get:

$$\frac{\Delta v}{(\Delta t \times v)} = \frac{\Delta s}{(\Delta t \times r)}$$

Then, because $a = \frac{\Delta v}{\Delta t}$ and $\frac{\Delta s}{\Delta t} = v$:

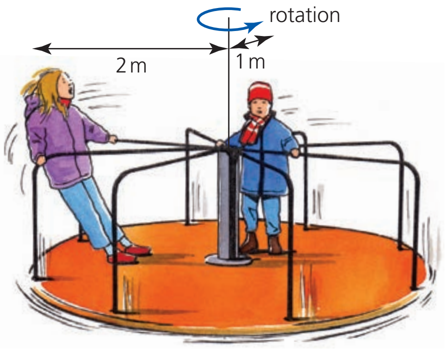
$$a = \frac{v^2}{r}$$

◆ **Derive** Explain in detail the origin of an equation.

We know $F = ma$ from Newton's second law of motion, so, the equation for the centripetal force acting on a mass m moving in a circle is:

$$F = \frac{mv^2}{r} = m\omega^2 r$$

In Figure A2.86, although both children have the same angular velocity, the bigger child needs a much greater centripetal force, so she should hold on tighter. This is because she has greater mass and is travelling with a greater linear speed.



■ **Figure A2.86** Children on a playground ride

WORKED EXAMPLE A2.19

Consider a ball of mass 72 g whirled with a constant speed of $3.4\,\mathrm{m\,s^{-1}}$ around in a (nearly) horizontal circle of radius 65 cm on the end of a thin piece of string, as shown in Figure A2.87.

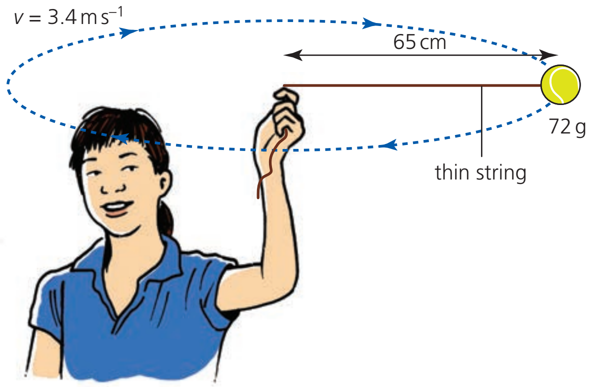


Figure A2.87 Ball whirled with a constant speed in a (nearly) horizontal circle

- **a** Calculate the centripetal acceleration and force.
- **b** Explain why the force provided by the string cannot act horizontally.
- **c** Suggest a probable reason why the string breaks when the speed is increased to $5.0\,\mathrm{m\,s^{-1}}$.
- **d** Predict in which direction the ball moves immediately after the string breaks.

Answer

a $a = \frac{v^2}{r} = \frac{3.4^2}{0.65} = 18\,\mathrm{m\,s^{-2}}$

$F = ma = 0.072 \times 18 = 1.3\,\mathrm{N}$

b If the force is horizontal, it cannot have a vertical component with which to support the weight of the ball (see Figure A2.88).

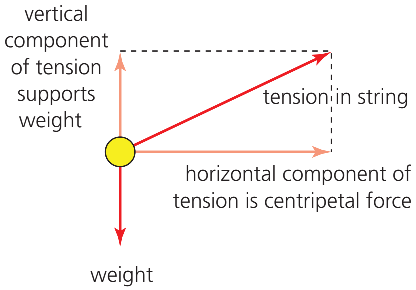


Figure A2.88 Free-body diagram for a ball whirled in a circle

c As the speed of the ball is increased, a greater centripetal force is needed for the same radius. If this force is greater than can be provided by the string, the string will break. This occurs when the speed reaches $5\,\mathrm{m\,s^{-1}}$.

d The ball will continue its instantaneous velocity in a straight line after the string breaks. It will move at a tangent to the circle, but gravity will also affect its motion.

82 Estimate how much greater is the size of the centripetal force acting on the larger child (than the smaller child) in Figure A2.86. Explain your answer.

Figure A2.89 Throwing the hammer

83 The hammer being thrown in Figure A2.89 completed two full circles of radius 2.60 m at a constant speed in 1.38 s just before it was released. Assuming that the motion was horizontal:

- **a** Calculate its centripetal acceleration.
- **b** What force did the thrower need to exert on the hammer if its mass was 4.00 kg?
- **c** The thrower will aim to release the hammer when it is moving at an angle of 45° to the horizontal. Explain why.

84 What is the centripetal acceleration of an object moving in a circular path of radius 84 cm if there are exactly two revolutions every second?

85 The Moon's distance from the Earth varies but averages about 380 000 km. The Moon orbits the Earth in an approximately circular path every 27.3 days.

- **a** Show that the Moon's orbital speed is about $1\,\mathrm{km\,s^{-1}}$.
- **b** Calculate the centripetal acceleration of the Moon towards the Earth.

86 A car of mass 1240 kg moved around a bend of radius 37 m at a speed of 16 m s^{-1} (see Figure A2.90). If the car was to be driven any faster, there would not have been enough friction and it would have begun to skid off the road.

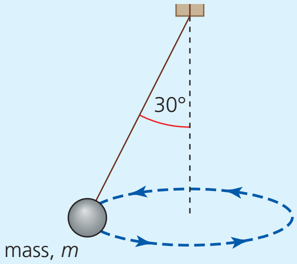
■ **Figure A2.90** Car driving around a tight mountain bend

- **a** Calculate the magnitude of the centripetal force assuming that the road is horizontal.
- **b** Determine a value for the coefficient of friction between the road and the tyre.
- **c** State whether this is a coefficient of static friction or dynamic friction.
- **d** Discuss whether a heavier car would be able to move faster around this bend.

87 A girl of mass 42 kg living in Sydney is moving (like everyone else) in a circle because of the rotation of the Earth. Sydney is 5.31 × 10^6 m from the Earth's axis of rotation.

- **a** Calculate her linear speed of rotation.
- **b** What is her centripetal acceleration?
- **c** Determine the resultant force needed on her to maintain her circular motion.
- **d** What provides this centripetal force?
- **e** Your answer to part **c** should be much less than 1% of the girl's weight. It is so small that this force is usually considered to be insignificant. However, draw a free-body diagram of her standing on the Earth's surface that includes numerical values of the forces involved.

88 Figure A2.91 shows a pendulum of mass 120 g being swung in a horizontal circle.



■ **Figure A2.91** Pendulum

- **a** Draw a free-body diagram of the mass, *m*.
- **b** Calculate the centripetal force acting on the mass.
- **c** If the radius of the circle is 28.5 cm, **i** what is the speed of the pendulum and **ii** how long does it take to complete one circle?

TOK

Knowledge and the knower

- How do our expectations and assumptions have an impact on how we perceive things?
- What constitutes a 'good reason' for us to accept a claim?

Most people accept that we live on a spherical rotating planet, but they have no 'direct' evidence of that. And as the Earth spins, we are told that the invisible force of gravity provides the necessary centripetal force that keeps us attracted to the Earth's surface. Our own observations are more likely to suggest that we live on a mostly flat Earth, and that the Sun and stars move around us.

Foucault's pendulum (Figure A2.92) provides evidence of the Earth's rotation, but not in an obvious way, and it needs to be explained to non-scientists.

■ **Figure A2.92** Foucault's pendulum

Are we sensible to believe what we are told by 'experts' and teachers, rather than to trust our own senses? How can we decide whether or not to accept knowledge claims made by our predecessors?

Non-uniform circular motion

Vertical motion

As an example, consider a ball of mass, *m*, on the end of string which is being spun in a vertical circle of radius *r*, as shown in Figure A2.93. There are two forces which can act on the ball: the weight of the ball, $F_g = mg$, and the tension in the string, F_T.

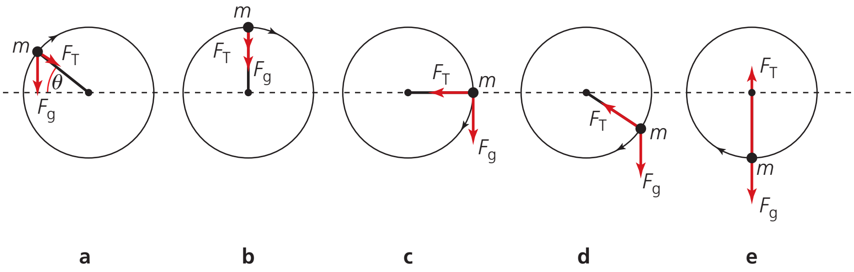


Figure A2.93 Forces on a mass moving in a vertical circle

Vertical motion is more complicated than horizontal motion because the centripetal force is affected by the combination of the tension in the string and the component of the ball's weight, both of which vary continuously during the motion.

If the ball was moving at a constant speed in a circle of constant radius:

- In position **b**, centripetal force $(mv^2/r) = F_T + F_g$, so that the required tension will have its minimum value. If the weight is greater than the necessary centripetal force, the string will lose tension and the ball will move inwards from its circular path. See further example below.
- In position **e**, centripetal force $= F_T - F_g$, so that the required tension will have its maximum value. It is at this position that the string is most likely to break.
- In position **c**, centripetal force $= F_T$, because there is no component of weight acting to, or from, the centre.
- In position **a** there will be a component of weight acting towards the centre.
- In position **d** there will be a component of weight acting away the centre.

In practice, it is unlikely that the tension can be continuously adjusted to keep the centripetal force constant. This means that the speed of the ball will change during its rotation. It will not be uniform circular motion.

The situation shown in **b** is the most widely discussed, and whirling a bucket of water in a vertical circle always makes for an interesting demonstration. See Figure A2.94.

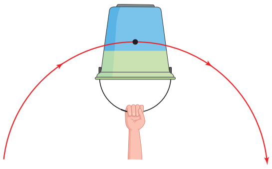


Figure A2.94 Bucket of water spinning in a circle

The hand pulling on the handle provides a force on the bucket. The normal contact force from the bottom of the bucket provides a force on the water. In order for the water to stay in contact with the bucket, it must be spinning with a speed which requires a centripetal force which is equal to, or greater than, its weight (so that a force from the bucket is also needed):

$$F = mv^2 \geq mg$$

$$v_{\text{min}}{}^2 = gr$$

$$v_{\text{min}} = \sqrt{gr}$$

WORKED EXAMPLE A2.20

a What is the minimum linear speed needed to keep water in a bucket rotating in a circle of radius of 0.95 m?

b What is the minimum angular velocity needed?

c How long will each rotation take if the bucket could be kept moving at the same speed?

d What maximum pulling force would be needed to maintain that speed if the bucket and water had a combined mass of 2.1 kg?

Answer

a $v_{\text{min}} = \sqrt{gr} = \sqrt{(9.8 \times 0.95)} = 3.1\,\text{m}\,\text{s}^{-1}$ (3.0512... seen on calculator display)

b $\omega = \frac{v}{r} = \frac{3.0512}{0.95} = 3.2\,\text{rad}\,\text{s}^{-1}$

c $T = \frac{2\pi r}{v} = \frac{(2\pi \times 0.95)}{3.0512} = 2.0\,\text{s}$

d Maximum pull (tension) is needed when the bucket is at its lowest point:

$$\frac{mv^2}{r} = F_{\text{T}} - F_{\text{g}}$$

$$F_{\text{T}} = F_{\text{g}} + \frac{mv^2}{r} = (2.1 \times 9.8) + \frac{(2.1 \times 3.0512^2)}{0.95} = 20.6 + 20.6 = 41\,\text{N}$$

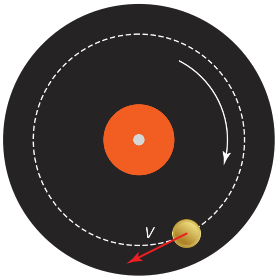


Figure A2.95 Coin on rotating turntable

Horizontal motion

We have seen that any object moving along the arc of a circle requires a centripetal force of magnitude: $F = \frac{mv^2}{r}$.

But what happens if the circumstances change? As an example, consider a coin on a rotating horizontal turntable, as shown from above in Figure A2.95. Friction provides the centripetal force on the coin. As the rotational speed increases, a greater frictional force is needed to keep the coin in the same place on the turntable. If the speed continues to increase, eventually there will not be enough friction and the coin will be thrown off the turntable (approximately along a tangent). A similar coin (to the first) placed closer to the centre will be able to stay on the turntable at greater speeds because less centripetal force (friction) is needed.

Consider again the car shown in Figure A2.90. If the radius, r, of a bend changes, the centripetal force needed changes. For example, if the radius reduces (the bend gets 'tighter'), a greater frictional force is needed to maintain the same speed. This may not be possible, so the driver should reduce speed. Similarly, a slower speed is advisable if water or ice on the road reduces frictional forces.

89 a Outline how the passengers seen in Figure A2.96 remain in their seats even though they are upside down (and even if they were not secured by safety harnesses!).

b What is the minimum speed needed if the carriage moves in a vertical arc of radius 15 m?

c In another part of the track the passenger carriage is upside down in a vertical arc of radius 20 m. Predict if the carriage needs to move faster, slower or the same speed. Explain your answer.

Figure A2.96 Upside down on a fairground ride

90 Consider Figure A2.93a. If the mass was 240 g, the radius of the circle was 52 cm, $\theta = 40°$, and the mass was moving with a linear speed of $2.12\,\text{m}\,\text{s}^{-1}$:

a Write down an expression for the component of weight acting towards the centre of motion.

b What was the necessary centripetal force in this position?

c Hence, determine the tension in the string.

91 A coin of mass 10 g was rotating on a turntable turning at 34 revolutions/minute. The coin was 17 cm from the centre.

a Calculate the magnitude of the centripetal force acting on the coin.

b Friction provides this force. The turntable's speed is increased so that more friction is required to keep the coin in place. If the coefficient of static friction is 0.43, what is the greatest possible value for the frictional force between the turntable and the coin?

c Determine the maximum angular velocity of the coin which will enable it to stay in the same place.

d How would your answers change if an identical coin had been fixed on top of the first coin?

A.3 Work, energy and power

Guiding questions

- How are concepts of work, energy and power used to predict changes within a system?
- How can a consideration of energetics be used as a method to solve problems in kinematics?
- How can transfer of energy be used to do work?

◆ **Work, *W*** The energy transfer that occurs when an object is moved with a force. More precisely, work done = force × displacement in the direction of the force.

◆ **Energy** Ability to do work.

In the last two topics, we have discussed movement (A.1), and how forces can change the motion of objects (A.2). In this topic (A.3) we will move on to introduce two very important, closely related, numerical concepts: **work** and **energy**. Together these provide the 'accounting system' for science, enabling explanations and useful predictions to be made.

Of course, 'work' and 'energy' are words in common use in everyday language, but in physics they have much more precise definitions.

Nature of science: Science as a shared endeavour

Same words, different meanings

The terms used in physics have very precise meanings – but the same words are often used differently in everyday life. There is a long list of such words, such as 'work', 'energy' and 'power', as well as the various meanings of 'conservation', 'law', 'momentum', 'pressure', 'stress', 'efficiency', 'heat', 'interference', 'temperature' and so on.

This ambiguity is a problem that all students must overcome when learning physics. For example, what is the connection, if any, between work = force × displacement and 'I have a lot of homework to do tonight'?

Work

SYLLABUS CONTENT

- Work, W, done on a body by a constant force depends on the component of the force along the line of displacement as given by: $W = Fs\cos\theta$.

Work done by constant forces

We say that work is done when any force moves an object: work is done *on* the object *by* the force. The work done, W, can be calculated by multiplying the displacement, s, by the component of the force acting in that direction, $F\cos\theta$.

See Figure A3.1 and note that θ is the angle between the force and the direction of motion.

work done, $W = Fs\cos\theta$ SI unit: Joule, J

One **joule** is defined to be the work done when a force of 1 N moves through a distance of 1 m.

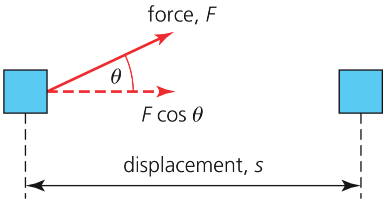


Figure A3.1 Work done by a force

◆ **Joule, J** Derived SI unit of work and energy. 1 J = 1 N m.

Commonly, a force acts in the same direction as the motion, in which case, the equation reduces to $W = Fs$.

WORKED EXAMPLE A3.1

Calculate how much work is done when a 1.5 kg mass is raised 80 cm vertically upwards.

Answer

The force needed to raise an object (at constant velocity) is equal to its weight (mg). The symbol h is widely used for vertical distances. (To avoid confusion, W will normally be used to represent work, and not weight.)

$W = Fs = mg \times h = 1.5 \times 9.8 \times 0.80 = 12\,\text{J}$

WORKED EXAMPLE A3.2

The 150 kg box in Figure A3.2 was pulled 2.27 m across horizontal ground by a force of 248 N, as shown.

- **a** Determine how much work was done by the force.
- **b** Suggest why it may make it easier to move the box if it is pulled in the direction shown by the dashed line.
- **c** When the box was pulled at an angle of 20.0° to the horizontal, the force used to slide the box was 248 N. Calculate the work done by this force in moving the box horizontally the same distance.

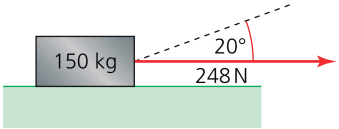


Figure A3.2 Box being pulled across the ground

Answer

- **a** $W = Fs = 248 \times 2.27 = 563\,\text{J}$
- **b** When the box is pulled in this direction, the force has a vertical component that helps reduce the normal contact force between the box and the ground. This will reduce the friction opposing horizontal movement.
- **c** The force is not acting in the same direction as the movement. To calculate the work done we need to use the horizontal component of the 248 N force.

  $W = F\cos 20° \times s = 248 \times 0.940 \times 2.27 = 529\,\text{J}$

It is important to realize that there are some surprising examples involving forces where *no* work is being done, as shown in Figures A3.3 and A3.4.

Figure A3.3 No work is being done on the weights at this moment

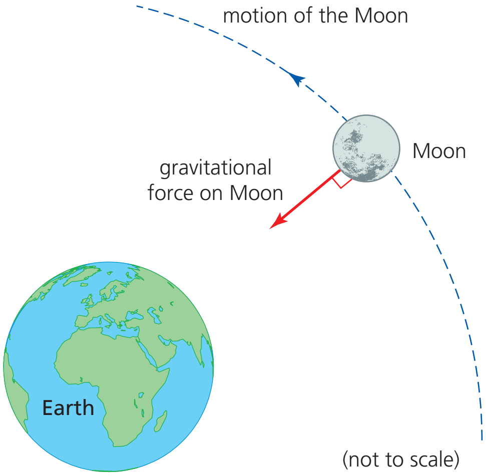


Figure A3.4 No work is done as the Moon orbits the Earth

In Figure A3.3 a large upwards force is being exerted on the stationary weights, but since there is no movement at that moment, no mechanical work is being done on the weights. However, work *is* being done in the weightlifter's muscles. In Figure A3.4, the Moon is moving *perpendicularly* to the force of gravity (cos 90° = 0), so there is no component of force in the direction of motion. A satellite in a similar circular path would not need to do any work, so that it would not need an engine to maintain the motion.

Work done by varying forces

When making calculations with the equation $W = Fs\cos\theta$ we are assuming a single, constant value for the force, but in reality, forces are rarely constant. In order to calculate the work done by a *varying* force, we have to make an estimate of the *average* force involved. This may be best done with the help of a force–time graph, or a force–distance graph, as shown in Figure A3.5, which could represent, for example, the resultant force used to decelerate a car.

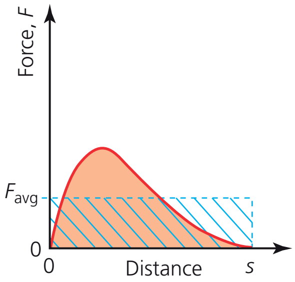


Figure A3.5 A force varying with distance

The horizontal dotted blue line represents the average force, as judged by eye.

Then, work done = $F_{avg} \times s$, which is the same as the rectangular area, and it is also equal to the area under the original curved line.

Work done is equal to the area under a force–distance graph.

WORKED EXAMPLE A3.3

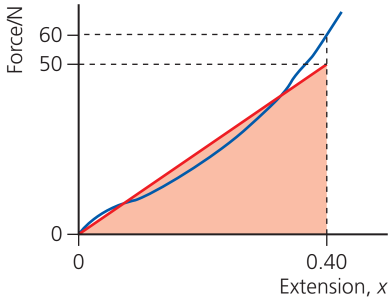


Figure A3.6 Force–extension (displacement) graph for rubber under tension

The blue line in Figure A3.6 represents the variations in force and extension as a rubber band was stretched. Estimate the work done as the force increased from zero to 60 N.

The work done is equal to the area under the graph. The red line has been drawn so that the area under it is the same as the area under the blue line (as judged by eye):

$$W = \frac{1}{2} \times 50 \times 0.40 = 10\text{ J}$$

1 Calculate the work done to:

  a lift an 18 kg suitcase a vertical height of 1.05 m

  b push the same suitcase the same distance horizontally against an average frictional force of 37 N.

2 What was the magnitude of the average resistive force opposing the forward motion of a car if 2.3×10^6 J of work were done while maintaining a constant speed over a distance of 1.4 km?

3 In Figure A3.7 a gardener is pushing a lawnmower at a constant speed of 0.85 m s^{-1} with a force, F, of 70 N at an angle of 40° to the ground.

  a Calculate the component of F in the direction of movement.

  b What is the magnitude of the frictional force?

  c Determine the work done in moving the lawnmower for 3.0 s.

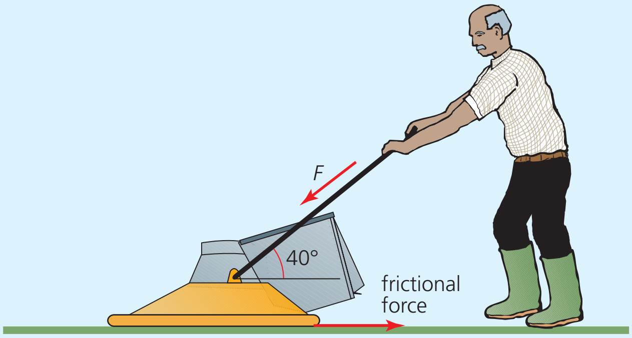


Figure A3.7 Gardener pushing a lawnmower

4 A spring which obeyed Hooke's law was stretched with a force which increased from zero to 24 N. The spring extended by 12 cm.

  a What was the average force used during the extension of the spring?

  b Calculate the work done on the spring.

Nature of science: Models

Macroscopic and microscopic work

The term ‘work’ is normally used to describe large-scale (macroscopic) movements and forces – such as kicking a ball – in which countless billions of particles move as a whole object. We do not concern ourselves with the individual particles.

But the concept of ‘doing work’ can also be applied to microscopic processes involving individual particles. For example, work is done when two particles collide with each other.

Work done during the random and unknown motions of individual particles can result in a different kind of energy transfer, called thermal energy. This important distinction is discussed in Topic B.4.

Energy

SYLLABUS CONTENT

- Work done by a force is equivalent to a transfer of energy.

Energy is probably the most widely used concept in the whole of science. However, the idea of energy is not easy to fully explain or truly understand.

When a battery is placed in a child’s toy dog (Figure A3.8), it moves, jumps up in the air and barks. After a certain length of time, the toy stops working. In order to try to explain these observations we will almost certainly need to use the concept of energy: chemical energy in the battery is transferred to electrical energy, which produces motion energy in a small electric motor. Some energy is also transferred from electricity to sound in a loudspeaker. Eventually, all the useful energy in the battery will be transferred to the surroundings and the toy will stop activity. Without the concept of *energy* all this is very difficult to explain.

■ **Figure A3.8** The toy dogs get their energy from batteries

We can talk about the energy *in* the gasoline (petrol) we put in the tanks in our cars (for example) and go on to describe that energy being transferred to the movement of the car. But nothing has actually flowed out of the gasoline into the car, and this is just a convenient way of expressing the idea that the controlled combustion of gasoline with oxygen in the air can do something that we consider to be useful.

Perhaps the easiest way to understand the concept of energy is this: energy is needed to make things happen. Whenever anything changes, energy is transferred from one form to another. *Most importantly, energy transfers can be calculated*, and this provides the basic ‘accounting system’ for science. Any event will require a certain amount of energy for it to happen and, if there is not enough energy available, it cannot happen. For example, if you do not get enough energy (originally from your food), you will not be able to climb a 500 m hill; if your phone battery is not charged, you cannot call your friends; if you do not put enough gasoline in your car, you will not get to where you want to go; if energy is not transferred quickly enough from an electrical heater, your shower will not be hot enough.

A person, device or machine which provides a force to do work must be *able* to do the work. We say that they must have enough *energy* to do the work. Energy is often described as the capacity to do work. To do work, there must be a ‘source’ of energy.

When work is done *energy is transferred* from a source to the object.

In Figure A3.9, the archer uses her store of energy (originally from food) to do work on the bow, which then stores energy because it is stretched out of shape. The bow then does work when energy is transferred to the movement of the arrow. When the arrow hits the target, work is done on it by the arrow as energy is transferred causing a change of shape and a small rise in temperature.

■ **Figure A3.9** Archer and target

TOK

Knowledge and the knower

- What criteria can we use to distinguish between knowledge, belief and opinion?

Abstract concepts

Energy is one of many **abstract concepts** in physics. Everyday abstract concepts include ‘hope’, ‘justice’ and ‘freedom’. They are all very useful ideas that can be explained (often with difficulty!) and people can understand them at various levels, and in differing, often personal and **subjective** ways, but they have no actual physical form.

Should we believe that a non-abstract physics concept, like force, for example, is more ‘real’ than the abstract concept of energy?

◆ **Abstract concept** An idea which has no physical form.

◆ **Subjective** Describes an opinion based on personal experiences and emotions. Compare with **objective**, meaning free from emotion and bias.

◆ **Potential energy** Energy that arises because of forces between different parts of the system. Sometimes described as stored energy.

Different forms of energy

The capacity of something to do work can exist in many different *forms* (of energy). These forms can be difficult to classify, and no two sources of information ever seem to agree on a simple, definitive list or even on how to use the term ‘forms’! The different forms of energy are a constant background to the study of physics and need to be well understood. The following is a broad initial summary.

Potential energy sources store energy because of their position or arrangement, and the forces between different parts of the system.

- **Gravitational potential energy**: The energy stored due to the position of a mass in a gravitational field. For example, a weight raised above the ground.
- **Electric potential energy**: The energy stored due to the position of a charge in an electric field (see Theme D).
- **Magnetic potential energy**: The energy stored due to position in a magnetic field (see Theme D).
- **Elastic potential energy**: The energy stored in a deformed elastic material, or a spring.
- **Chemical potential energy**: The energy stored in the bonding of chemical compounds, released in chemical reactions.
- **Nuclear potential energy**: The energy stored in the arrangement of particles in the nuclei of atoms.

Figure A3.10 Water behind the Three Gorges Dam (China) stores enormous amounts of gravitational potential energy

Figure A3.11 Kori nuclear power station (S. Korea) is the world's largest

A macroscopic object that is able to do work is said to possess *mechanical energy*.

Mechanical energy can come in one of three forms:

- **Kinetic energy**: The energy of all moving masses which could do work on anything they collide with. (Includes wind and mechanical waves, including sound).
- Elastic potential energy: As described above.
- Gravitational potential energy: As described above.

All matter contains large amounts of energy inside it.

Internal energy is the name we give to the enormous amount of energy which exists within all matter because of the motions and positions of the particles it contains.

The following types of energy transfer will be discussed in later topics:

- **Thermal energy**: Energy transferred because of a temperature difference.
- **Electrical energy**: Energy carried along metal wires because of a potential difference (voltage – see Topic B.5).
- **Radiation energy**: Light, for example: energy transferred as electromagnetic waves.

In Theme E we will discuss the equivalence of energy and mass.

◆ **Gravitational potential energy** Energy that masses have because of the gravitational forces between them.

◆ **Electric potential energy** Energy that charges have because of the electric forces between them.

◆ **Elastic potential energy** Energy that is stored in a material that has been deformed elastically.

◆ **Chemical potential energy** Energy related to the arrangement of electrons within the structure of atoms and molecules.

◆ **Nuclear potential energy** Energy related to the forces between particles in the centres (nuclei) of atoms.

◆ **Kinetic energy** Energy of moving masses.

◆ **Thermal energy (heat)** The (non-mechanical) transfer of energy between two or more bodies at different temperatures (from hotter to colder).

◆ **Electrical energy** Energy delivered in a circuit by an electrical current.

◆ **Radiation energy** Energy transferred by electromagnetic waves.

◆ **Internal energy** Total potential energies and random kinetic energies of the particles in a substance.

Energy transfers

Energy transfer (between the forms listed above) is the central, recurring theme of all of physics. Energy can be quantified and measured, and the total amount within a system remains the same, although some useful energy is always 'wasted' in every macroscopic energy transfer. Some examples of useful energy transfers:

- Our bodies transfer chemical energy from food to internal energy and kinetic energy.
- A rocket transfers chemical energy to gravitational potential energy.
- A 'clockwork' toy transfers elastic potential energy to kinetic energy.
- An electric light transfers electrical energy to radiation energy.
- A nuclear power station transfers nuclear potential energy to electrical energy.

These are just a few random examples. Any list like this can be very long!

Principle of conservation of energy

SYLLABUS CONTENT

- ▶ The principle of the conservation of energy.

The total energy of an isolated system remains constant.

An alternative way to state the same law is 'energy cannot be created or destroyed'. We can move energy around and transfer it from one form to another, but the total amount remains the same.

This is one of the most important principles in the whole of science, not only because it is one of very few principles of science which is always true, but also because it is highly relevant to every event that occurs, helping us to predict what can, and what cannot, happen.

A financial analogy may help: if you leave home with \$10 cash in your pocket, spend \$5 and arrive back home with \$2, then you will probably assume that you lost \$3 somewhere. And you would not expect to have, say, \$7 left in your pocket. You believe in the 'conservation' of cash, even if you are not sure of where it has gone.

Similarly, we know that if, for example, 20 000 J of energy is transferred into a system (when charging a mobile phone for example), we cannot take 25 000 J out, and if only 4000 J remain, then we know that 16 000 J has been transferred somewhere else.

Common mistake

'Energy conservation' has evolved to have a slightly different meaning in everyday use. We are often advised to 'conserve' energy, meaning that we should not use too much now, because there is a limited supply and we may not have enough for when we want it in the future.

LINKING QUESTION

- Where do the laws of conservation apply in other areas of physics? (NOS – see page 92.)

This question links to understandings in Topics A.2, B.4 and E.3.

◆ **Conservation laws** In isolated systems, some physical quantities remain constant under all circumstances: energy/mass, charge, momentum.

Nature of science: Theories

Conservation laws

A **conservation law** tells us that a measurable physical quantity does not change after any event, or with the passage of time. The following quantities, which occur in this course, are always conserved in any well-defined system.

- Energy
- Mass (the equivalence of mass and energy will be outlined in Topics E.4 and E.5)
- Linear momentum (see Topic A.2)
- Angular momentum (see Topic A.4)
- Electric charge (see Topic B.5)

Other quantities, such as force, velocity and temperature, are not conserved.

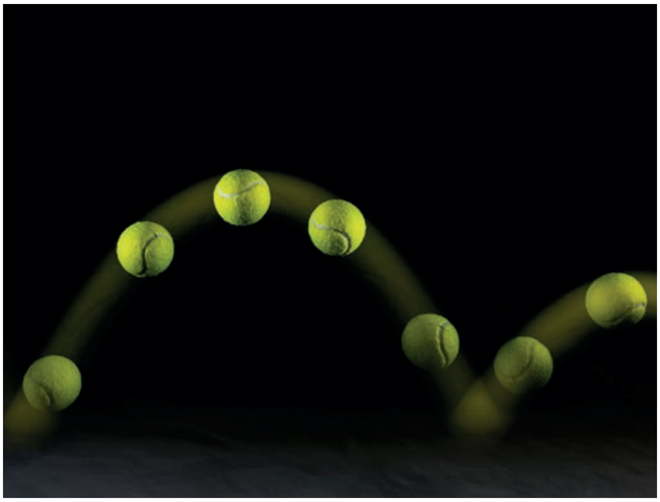
Conservation laws are important in predicting the outcome of events.

Dissipation of energy

So, we cannot create energy: the total energy after any event cannot be more than we started with, but in the macroscopic world in which we live, it seems that *every* event seems to 'lose' energy. Take a bouncing ball as an example, as shown in Figure A3.12. The height of each bounce is less than the one before, which shows us that the ball is losing (gravitational potential) energy. Also, just before each time the ball hits the ground it will also have less kinetic energy.

◆ **Dissipated** or **degraded energy** Energy that has spread into the surroundings and cannot be recovered to do useful work.

But if we think of the ball and the ground together as the 'system', the mechanical energy 'lost' by the ball has been 'gained' by the ground, keeping the total energy constant. The energy gained by the ground is in the form of *internal energy* and a very accurate temperature measurement would show that it had become a little warmer at the points of contact. (A little sound energy will also be present.) Some of the mechanical energy of the ball will also have been transferred to internal energy in the ball, which will also be slightly warmer.



■ **Figure A3.12** Bouncing ball

It is very important to appreciate that gravitational potential energy and kinetic energy can be considered as 'useful' mechanical energies, but the energy transferred to the ground spreads out into the surroundings and can never be recovered to do any useful work. It is sometimes called 'wasted' energy, but it may be better described as **dissipated energy**. Sometimes it is called *degraded energy*.

All macroscopic processes dissipate energy into the surroundings.

The law of conservation of energy is a constant theme throughout science, but in this chapter, we will concentrate on energy transfers in *mechanical systems*. Numerical examples will be given later.

Using Sankey diagrams to represent energy transfers

SYLLABUS CONTENT

- Energy transfers can be represented on a Sankey diagram.

Energy transfers can be usefully shown in flow diagrams, such as shown in Figure A3.13, which represents the energy transfers in a small LED lamp in a specified time.

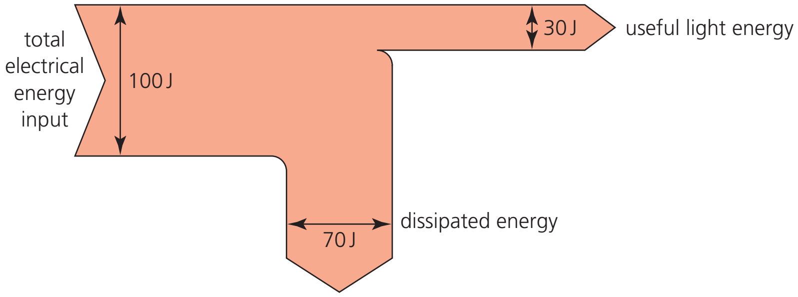


Figure A3.13 A simple Sankey diagram

◆ Sankey diagram
Diagram representing the flow of energy in a system.

The width of each section is proportional to the amount of energy (or power), starting with the energy input shown at the left of the diagram. Dissipated energy is shown with downwards arrows and useful energy flows to the right. Diagrams like these are known as **Sankey diagrams** and they can be used to help represent many energy transformations.

Sankey energy flow diagrams can be used to visualize quantified transfers of energy.

As a more complicated example, Figure A3.14 represents the useful energy transformations in a fossil-fuel power station and Figure A3.15 shows a Sankey diagram representing the energy flow through the same system, including the dissipated energy.

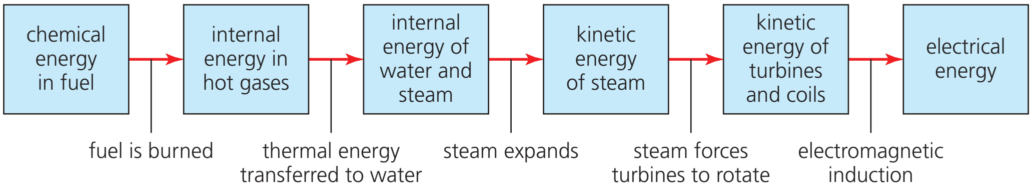


Figure A3.14 Energy transfers in a fossil-fuel power station

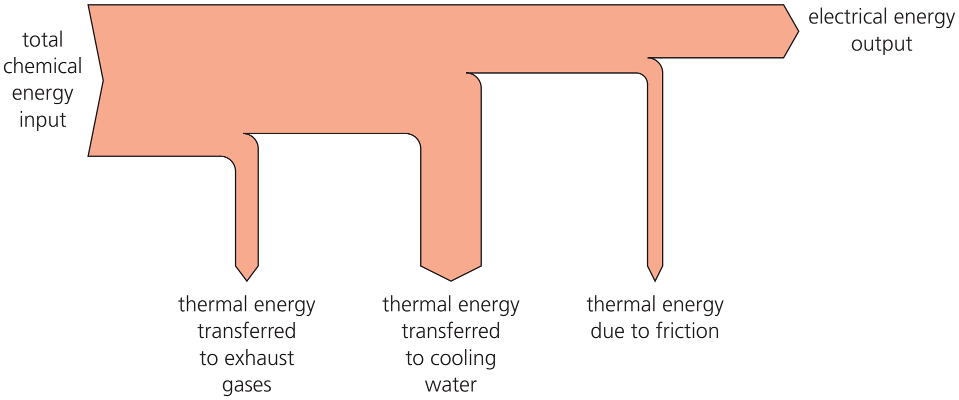


Figure A3.15 Sankey diagram for a fossil-fuel power station

Nature of science: Theories

Law or principle?

Many physics books refer to the 'law' of conservation of energy, others refer to the 'principle' of conservation of energy. Is there any difference? And why is it not called the 'theory' of conservation of energy?

Research to determine if there are any differences between 'theories', 'laws' and 'principles.'

5 Outline the energy transfers that occur when a mass hanging vertically on the end of a metal spring is displaced and allowed to move up and down (oscillate) freely until it stops moving.

6 State the main energy transfers that occur in the use of a mobile phone.

7 State the names of devices whose main uses are to perform the following energy transfers:

- **a** electricity to sound
- **b** chemical to electricity
- **c** sound to electricity
- **d** chemical to radiation
- **e** chemical to kinetic
- **f** elastic to kinetic
- **g** kinetic to electricity
- **h** chemical to internal
- **i** electromagnetic radiation to electricity.

8 Sketch and annotate a Sankey diagram to represent the following process. A car transfers 1500 kJ from its fuel as it gains 100 kJ of kinetic energy and 200 kJ of gravitational potential energy. The rest of the energy is dissipated into the environment.

9 **a** What are the two main types of energy of an aircraft flying at a height of 12 km?

b After the airplane has landed at an airport, what has happened to most of that energy?

10 An adult male body transfers about 10^7 J of energy every day.

a Name the source of this energy.

b **i** Outline the principle uses for this energy and

ii why does it have to be replaced?

11 A car slows down for traffic lights. State two causes of energy dissipation.

Common mistake

Many students believe (wrongly) that batteries store electrical energy. Batteries contain chemical compounds that react when a circuit containing the battery is connected. Chemical energy is then transferred to an electric current. Many batteries can be recharged if the direction of current is reversed. This reverses the chemical changes.

Calculating mechanical energies

SYLLABUS CONTENT

- Mechanical energy is the sum of kinetic energy, gravitational potential energy and elastic potential energy.
- Work done by the resultant force on a system is equal to the change in the energy of the system.
- Kinetic energy of translational motion as given by: $E_k = \frac{1}{2}mv^2 = \frac{p^2}{2m}$.
- Gravitational potential energy, when close to the surface of the Earth as given by: $\Delta E_p = mg\Delta h$.
- Elastic potential energy as given by: $E_H = \frac{1}{2}k\Delta x^2$.

◆ **Mechanical energy** Energy of a macroscopic object which can do useful work.

Work can be done on a macroscopic object / system to give it kinetic energy, gravitational potential energy, or elastic potential energy. The object then has the ability to transfer that energy to do useful work. We say that it has **mechanical energy**. We will now show how these three types of energy can be calculated.

The symbol E can be used to represent energy, usually with a subscript to represent the particular type of energy. (Note: E is also used for electric field.)

Kinetic energy

Tool 3: Mathematics

Derive relationships algebraically

Kinetic energy can be calculated from the equation $E_k = \frac{1}{2}mv^2$. This is explained / derived below. (Ideally, we would like to fully explain the origin of all of the equations used in this course, but that is not always possible.)

Consider a mass m accelerated horizontally from rest by a constant resultant force, F, acting in the direction of motion, as shown in Figure A3.16.

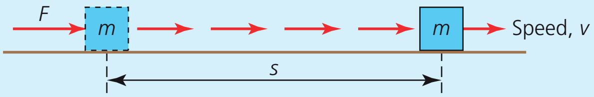


Figure A3.16 Doing work to increase movement

Using the equation of motion $v^2 = u^2 + 2as$ (from Topic A.1) and noting that, in this example, $u = 0$, we see that the distance travelled, s, can be determined from the equation: $s = \frac{v^2}{2a}$.

The work done W, in a distance s, can be calculated from: $W = Fs = ma \times \frac{v^2}{2a} = \frac{1}{2}mv^2$.

This amount of energy has been transferred from the origin of the force to the moving mass. We say that the mass has *gained* kinetic energy, E_k.

$\frac{1}{2}mv^2$ can also be written as: $\frac{[mv]^2}{2m}$.

Since momentum, mv, is given the symbol p, kinetic energy can also be determined from the equation: $E_k = \frac{p^2}{2m}$.

This version is most commonly used with the kinetic energy and momentum of atomic particles (See Topic E.5).

The kinetic energy of a moving mass can be calculated using: $\boldsymbol{E_k = \frac{1}{2}mv^2}$

Or from: $\boldsymbol{E_k = \frac{p^2}{2m}}$

◆ **Order of magnitude** An approximate value rounded to the nearest power of ten.

Some typical examples of kinetic energies (to an **order of magnitude**):

- a molecule in air at 20° C: 10^{-20} J
- a falling snowflake: 10^{-7} J
- a boy running in a 100 m race: 10^3 J
- a bullet from a rifle: 10^4 J
- a car moving along an open road: 10^5 J
- a large meteor moving towards Earth: 10^{17} J
- the Earth in its motion around the Sun: 10^{33} J.

Tool 3: Mathematics

Compare and quote values and approximations to the nearest order of magnitude

Physics is the fundamental science that tries to explain how and why everything in the Universe behaves in the way that it does. Physicists study everything from the smallest parts of atoms to distant objects in our galaxy and beyond (Figure A3.17).

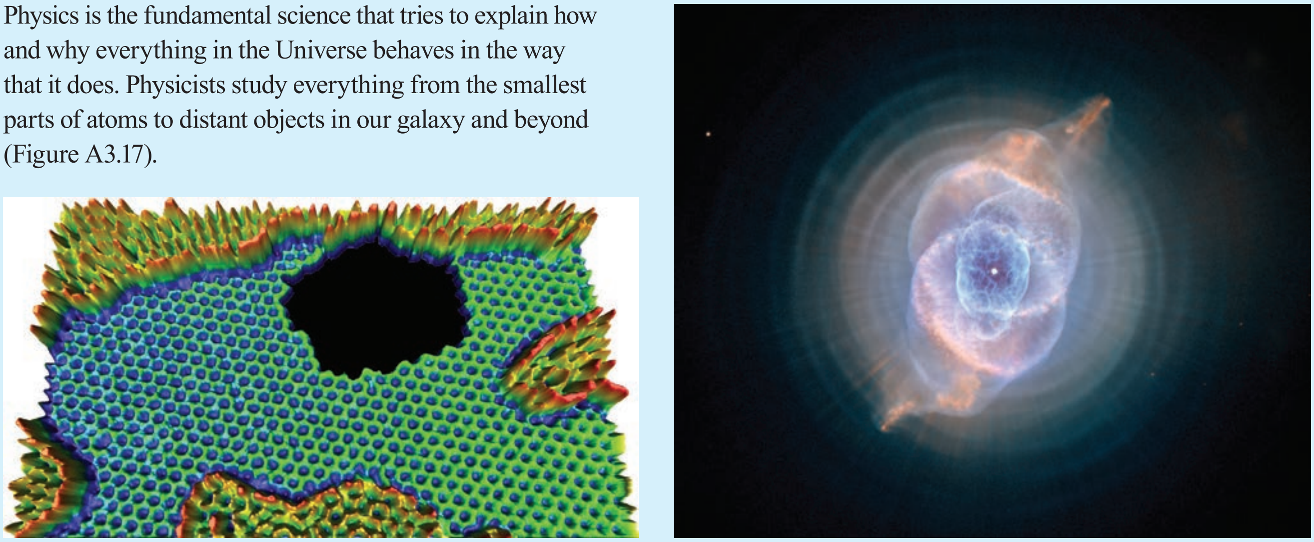


Figure A3.17 a The arrangement of individual atoms in graphene (a material made from a single layer of carbon atoms) can be seen using a special type of electron microscope; **b** complex gas and dust clouds in the Cat's Eye nebula, 3000 light-years away

Measurements and calculations commonly relate to the world that we can see around us (the macroscopic world), but our observations may require microscopic explanations, often including an understanding of molecules, atoms, ions and subatomic particles. Astronomy is a branch of physics that deals with the other extreme – quantities that are very much bigger than anything we experience in everyday life.

The study of physics therefore involves dealing with both very large and very small numbers. When numbers are so different from our everyday experiences, it can be difficult to appreciate their true size. For example, the age of the Universe is believed to be about 10^{18} s, but just how big is that number?

When comparing quantities of very different sizes (magnitudes), for simplicity we often make approximations to the nearest power of 10. When numbers are approximated and quoted to the nearest power of 10, it is called giving them an order of magnitude. For example, when comparing the lifetime of a human (the worldwide average is about 70 years) with the age of the Universe (1.4×10^{10} y), we can use the approximate ratio $10^{10} / 10^2$. That is, the age of the Universe is about 10^8 human lifetimes, or we could say that there are eight orders of magnitude between them.

Here are three further examples:

- The mass of a hydrogen atom is 1.67×10^{-27} kg. To an order of magnitude this is 10^{-27} kg.
- The distance to the nearest star (*Proxima Centauri*) is 4.01×10^{16} m. To an order of magnitude this is 10^{17} m. (Note: log of $4.01 \times 10^{16} = 16.60$, which is nearer to 17 than to 16.)
- There are 86 400 seconds in a day. To an order of magnitude this is 10^5 s.

Tables A3.1, A3.2 and A3.3 list the ranges of mass, distance and time that occur in the Universe.

Table A3.1 The range of masses in the Universe

Object mass / kg	**Object mass / kg**
the observable Universe	10^{53}
our galaxy (the Milky Way)	10^{42}
the Sun	10^{30}
the Earth	10^{24}
a large passenger airplane	10^{5}
a large adult human	10^{2}
a large book	1
a raindrop	10^{-6}
a virus	10^{-20}
a hydrogen atom	10^{-27}
an electron	10^{-30}

Table A3.2 The range of distances in the Universe

Distance size / m	Distance size / m
distance to the edge of the visible Universe	10^{27}
diameter of our galaxy (the Milky Way)	10^{21}
distance to the nearest star	10^{16}
distance to the Sun	10^{11}
distance to the Moon	10^{8}
radius of the Earth	10^{7}
altitude of a cruising airplane	10^{4}
height of a child	1
how much human hair grows by in one day	10^{-4}
diameter of an atom	10^{-10}
diameter of a nucleus	10^{-15}

Table A3.3 The range of times in the Universe

Time period time interval / s	Time period time interval / s
age of the Universe	10^{18}
time since dinosaurs became extinct	10^{15}
time since humans first appeared on Earth	10^{13}
time since the pyramids were built in Egypt	10^{11}
typical human lifetime	10^{9}
one day	10^{5}
time between human heartbeats	1
time period of high-frequency sound	10^{-4}
time for light to travel across a room	10^{-8}
time period of oscillation of a light wave	10^{-15}
time for light to travel across a nucleus	10^{-23}

Estimation

Sometimes we do not have the data needed for accurate calculations, or maybe calculations need to be made quickly. Sometimes a question is so vague that an accurate answer is simply not possible. The ability to make sensible estimates is a very useful skill that needs plenty of practice.

When making estimates, different people will produce different answers and it is usually sensible to use only 1 (maybe 2) significant figures. Sometimes only an order of magnitude is needed.

The numbers in the list of kinetic energies given above cannot be given with precision because the situations are vague and there are a wide range of possibilities (with the exception of the Earth's kinetic energy). For example, boys (of all ages) in 100 m races could have masses between 20 kg and 70 kg (or more), and they could run at speeds between 2 m s^{-1} and 9 m s^{-1}. These figures correspond to a kinetic energy range of 90–2800 J. To give a value with 2, or 3, significant figures would be misleading. It is more sensible to give a typical value to the nearest power of 10 (or sometimes, 1 significant figure). In the case of boys running 100 m, values of 10^2 J or 10^3 J may be considered typical (a matter of opinion).

WORKED EXAMPLE A3.4

A constant resultant horizontal force of 40 N accelerated a box over a distance of 50 cm.

a How much work was done on the box?

b State the assumption that you made answering part **a**.

c Calculate the kinetic energy that was gained by the box.

d If the box was initially at rest and then reached a speed of 2.9 m s^{-1}, what was its mass?

Answer

a $W = Fs = 40 \times 0.50 = 20\,\text{J}$

b The force was in the same direction as the motion of the box.

c 20 J

d $E_k = \frac{1}{2}mv^2$

$$20 = \frac{1}{2} \times m \times 2.9^2$$

$$m = \frac{(2 \times 20)}{2.9^2}$$

$$= 4.8\,\text{kg}$$

Understand the significance of uncertainties in raw and processed data

Although scientists are perceived as working towards finding 'exact' answers, an unavoidable **uncertainty** exists in every measurement. The results of all scientific investigations have uncertainties and errors, although good experimenters will try to keep these as small as possible.

When we receive numerical data of any kind (scientific or otherwise) we need to know how much belief we should place in the information that we are reading or hearing. The presentation of the results of serious scientific research should always have an assessment of the uncertainties in the findings, because this is an integral part of the scientific process. Unfortunately, the same is not true of much of the information we receive through the media, where data is too often presented uncritically and unscientifically, without any reference to its source, uncertainties or reliability.

No matter how hard we try, even with the very best of measuring instruments, it is simply not possible to measure anything exactly. For one reason, the things that we can measure do not exist as perfectly exact quantities; there is no reason why they should.

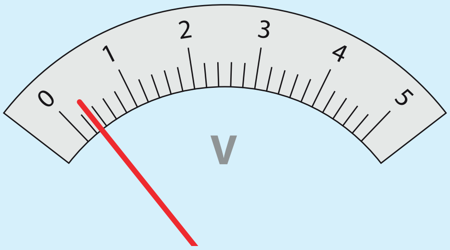
This means that every measurement is an approximation. A measurement could be the most accurate ever made, for example the width of a ruler might be stated as 2.283 891 03 cm, but that is still not perfect, and even if it was, we would not know because we would always need a more accurate instrument to check it. In this example we also have the added complication of the fact that when measurements of length become very small, we have to deal with the atomic nature of the objects that we are measuring.

A measurement is **accurate** if is close to its true value. For example, if you weigh a mass of 90.1 g and the result is 90.1 g, then the measurement can be described as accurate. However, be aware that in scientific research, true values are usually not known.

If the same mass (90.1 g) was measured by a different method, or a different person and the result was 89.8 g, then there was a clear **error** in the measurement. An error has occurred if the measurement is different from its true value (using an appropriate number of significant figures for the comparison).

Significant errors are often due to faulty apparatus or poor experimental skills. It is often possible to correct the source of such errors.

A **systematic error** occurs because there is something consistently wrong with the measuring instrument or the method used. A reading with a systematic error is always either bigger or smaller than the correct value by the same amount. Common causes are instruments that have an incorrect scale (wrongly calibrated), or instruments that have an incorrect value to begin with, such as a meter that displays a reading when it should read zero. This is called a **zero-offset error** – an example is shown in Figure A3.18. A thermometer that incorrectly records room temperature will produce systematic errors when used to measure other temperatures.



■ **Figure A3.18** This voltmeter has a zero-offset error of 0.3 V, so that all readings will be too large by this amount.

Uncertainties in measurements are an indication of the amount of variation seen in the readings taken (without considering their accuracy). For example, if you weigh the same (unknown) mass five times you might get the following results: 53.2 g, 53.4 g, 52.9 g. 53.0 g, and 53.1 g. The results are not all the same, so there is clearly some random uncertainty in the results. The uncertainty in the use of the measuring instrument itself is usually assumed to be equal to the smallest division on its scale / display. In this example this is ± 0.1 g.

All experimental data has uncertainties. Sometimes uncertainties will arise because of difficulties in taking measurements. For example, human reaction times affect measurements made when using a stopwatch, or measuring the distance moved by something moving quickly can be difficult.

It is usually not possible to reduce uncertainties using the same apparatus and techniques.

In science the word **precise** means that measurements have low uncertainty. If, and when, measurements are repeated they will produce similar results. Consider Figure A3.19, which shows where arrows fired at a target (not shown) landed on eight separate occasions. Because we do not know where the target is, we cannot tell if the arrows were fired accurately, or not. But we can describe **b** and **d** as more precise than **a** and **c**.

◆ **Uncertainty (random)** The range, above and below a stated value, over which we would expect any repeated measurements to occur. Uncertainty can be expressed in absolute, fractional or percentage terms.

◆ **Accuracy** A single measurement is described as accurate if it is close to the correct result. A series of measurements of the same quantity can be described as accurate if their mean is close to the correct result.

◆ **Error** When a measurement is not the same as the correct value.

◆ **Systematic error** An error which is always either bigger or smaller than the correct value by the same amount, for example a zero-offset error.

◆ **Zero offset error** A measuring instrument has a zero offset error if it records a non-zero reading when it should be zero.

◆ **Precision** A measurement is described as precise if a similar result would be obtained if the measurement was repeated.

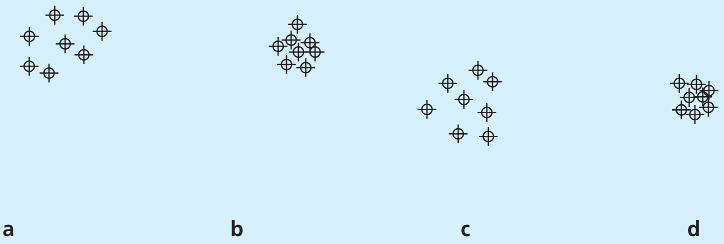


Figure A3.19 Precision of arrows hitting unseen target

In Figure A3.20 the positions of the targets have been included and we can see that the accuracy was good in **d**. But note that the accuracy seen in **c** is also good, because the average position of the arrows is close to the centre of the target.

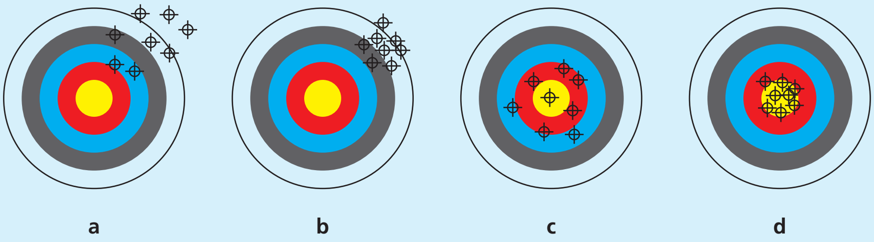


Figure A3.20 Accuracy of arrows hitting target

Common mistake

Uncertainties and errors are often confused, and different sources may define them slightly differently. **See Tool 3: Mathematics (Propagating uncertainties) on page 131.**

WORKED EXAMPLE A3.5

A student wanted to determine the increase of a trolley's kinetic energy as it accelerated down a slope. The trolley had a mass of 576 g ± 5 g and its length was 28.0 cm ± 0.5 cm. Using a stopwatch the student measured the time for the trolley to pass a point near the top of the slope to be 1.26 s. Near the bottom of the slope the trolley took 0.73 s to pass a particular point. Because it was difficult to start and stop the stopwatch at exactly the right time, it was estimated that the uncertainty in each time measurement was 0.10 s.

Calculate a value for the increase in kinetic energy of the trolley. Determine the absolute and percentage uncertainties in the answer. Comment on your answer

Answer

Near the top:

$$E_k = \frac{1}{2}mv^2 = \frac{1}{2} \times 0.576 \times \left(\frac{0.28}{1.26}\right)^2 = 1.42 \times 10^{-2}\,\text{J}$$

- Percentage uncertainty in mass = 100 × (5/576) = 0.87%
- Percentage uncertainty in length = 100 × (0.5/28) = 1.79%
- Percentage uncertainty in time = 100 × (0.10/1.26) = 7.94%
- Percentage uncertainty in kinetic energy = 0.87 + 1.79 + 1.79 + 7.94 + 7.94 = 20.33%
- Absolute uncertainty in kinetic energy = $(20.33/100) \times (1.42 \times 10^{-2}) = 0.29 \times 10^{-2}$ J

Near the bottom:

$$E_k = \frac{1}{2}mv^2 = \frac{1}{2} \times 0.576 \times \left(\frac{0.28}{0.73}\right)^2 = 4.24 \times 10^{-2}\,\text{J}$$

- Percentage uncertainty in mass = 100 × (5/576) = 0.87%
- Percentage uncertainty in length = 100 × (0.5/28) = 1.79%
- Percentage uncertainty in time = 100 × (0.10/0.73) = 13.70%
- Percentage uncertainty in kinetic energy = 0.87 + 1.79 + 1.79 + 13.70 + 13.70 = 31.85%
- Absolute uncertainty in kinetic energy = (31.85/100) × $4.24 \times 10^{-2} = 1.35 \times 10^{-2}$ J
- Difference in kinetic energies = $(4.24 \times 10^{-2}) - (1.42 \times 10^{-2}) = 2.82 \times 10^{-2}$ J
- Absolute uncertainty in difference = $(0.29 \times 10^{-2}) + (1.35 \times 10^{-2}) = 1.64 \times 10^{-2}$ J
- Percentage uncertainty in difference = $\left(\frac{1.64}{2.82}\right) \times 100 = 58\%$

This high percentage uncertainty in the results of this experiment may be surprising. The experiment needs redesigning if accurate results are needed.

Other types of kinetic energy

We can only use $E_k = \frac{1}{2}mv^2$ to determine the translational kinetic energy of objects travelling from place to place. Objects which are vibrating or rotating also have kinetic energy, but we need different equations to calculate their values. This will be covered in Topics A.4 and C.1 (for HL students).

12 Calculate the kinetic energy of a 57 g tennis ball served with a speed of 50 m s^{-1}.

13 a Determine the work needed to be done on a 1800 kg car to accelerate it from rest to 20 m s^{-1}.

b What average resultant force is needed to do this in a horizontal distance of 100 m?

c If the car decelerates from the same speed to rest in 70 m, calculate the average force exerted on the car

d What are the locations of this force?

14 a Calculate the kinetic energy of an electron (mass = 9.110×10^{-31} kg) moving with a speed which is 5% of the speed of light (3.0×10^8 m s^{-1}).

b What is the momentum of this electron?

15 Suppose that the force on the box in Worked example A3.4 continued to act for another 50 cm. Determine the final speed of the box after it had moved a total distance of 1.0 m.

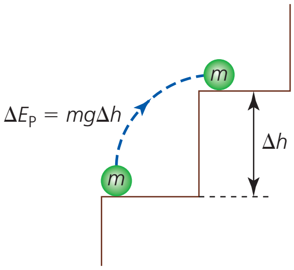
16 a Calculate the kinetic energy of a high-speed train (see Figure A3.21) which has a mass of 7.5×10^5 kg and is moving with a speed of 300 km h^{-1}.

b Compare your answer to the kinetic energy of a typical family car on a motorway.

c What average force is needed to reduce the speed of the train uniformly to zero in a distance of 5.0 km?

d Suggest two reasons why high-speed railway systems do not have many stations.

Figure A3.21 A high-speed train



■ **Figure A3.22** A mass gaining gravitational potential energy

◆ **Hydroelectric power** The generation of electrical power from falling water.

LINKING QUESTION

- Why is the equation for the change in gravitational potential energy only relevant close to the surface of the Earth, and what happens when moving further away from the surface?

This is discussed in Topic D.1 for HL students.

Gravitational potential energy

Consider raising a mass m a vertical height Δh, as shown in Figure A3.22. The minimum force required is equal to the weight of the mass, $F_g = mg$.

The work done by the force on the mass, $W = Fs = mg\Delta h$

This amount of energy has been transferred from the origin of the force to the raised mass. We say that the mass has gained gravitational potential energy, ΔE_p. If the mass is allowed to fall back down the same distance, the same amount of energy could be transferred to do useful work. Falling water in a **hydroelectric power** station uses this principle.

The gravitational potential energy of a mass raised a height Δh, close to the Earth's surface can be calculated using $\boldsymbol{\Delta E_p = mg\Delta h}$

It is important to realize that this equation can only be used accurately where a value of g can be considered as constant over the distance involved. For example, $g = 9.8\,\mathrm{N\,kg^{-1}}$ cannot be used accurately when moving masses large distances up from the Earth's surface.

Top tip!

A mass resting on a table, or on the ground, does not have *zero* gravitational potential energy. When we use $\Delta E_p = mg\Delta h$ we are calculating how much more, or less, gravitational potential energy the object has in its new position compared to the place from where it was moved.

In other words, we are calculating a *change* in gravitational potential energy.

In the detailed study of gravitational fields (Topic D.1) HL students will need to consider if there is any place where an object really does have zero gravitational energy.

WORKED EXAMPLE A3.6

Estimate the gravitational potential energy gained by a teenage girl who moves from ground level to the viewing platform on the 124th floor of the Burj Khalifa in Dubai (see Figure A3.23).

Assume that the girl has a mass of 40 kg and the height of each floor is 3.5 m.

■ **Figure A3.23** The Burj Khalifa in Dubai

Answer

$\Delta E_p = mg\Delta h = 40 \times 9.8 \times (124 \times 3.5) \approx 2 \times 10^5\,\mathrm{J}$

ATL A3A: Research skills

Use search engines and libraries effectively; evaluate information sources for accuracy, bias, credibility and relevance

When lifting a heavy object, the amount of gravitational potential energy that we need to transfer to it is decided only by its weight and the vertical height ($\Delta E_p = mg\Delta h$). For example, when a 50 kg mass is raised a height of 1 m it gains about 500 J of gravitational potential energy. Although 500 J may not be a lot of work to do, that does not mean that we can do this job easily.

There are two main reasons why this job could be difficult. Firstly, we may not be able to transfer that amount of energy in the time required to do the work. Another way of saying this is that we may not be powerful enough. (Power is discussed later in this topic.) Secondly, we may not be strong enough because we are not able to provide the required upwards force of 500 N. Power and strength are often confused with each other in everyday language.

Lifting (heavy) weights is a common human activity and many types of simple 'machine' were invented many years ago to make this type of work easier, by reducing the force needed. These include the ramp (inclined plane), the **lever** and the **pulley** (Figure A3.24).

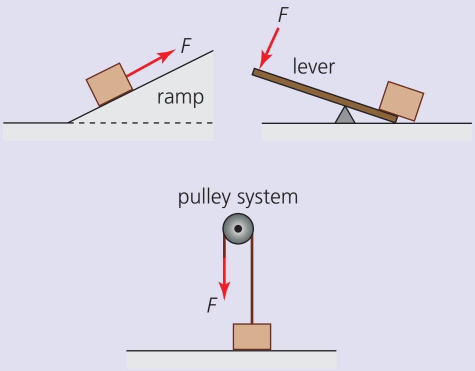


Figure A3.24 Simple machines which can be used to raise loads

In each of these simple machines the force needed to do the job is reduced, but the distance moved by the force is increased. If there was no energy dissipation (mainly due to friction), the work done by the force (Fs) would equal the useful energy transferred to the object being raised ($mg\Delta h$). In practice, because of energy dissipation, we will transfer more energy using a machine than if we lifted the load directly, without the machine. However, this is not a problem because we are usually much more concerned about how easy it is to do a job, rather than the total energy needed, or the efficiency of the process.

Figure A3.25 shows another example of a simple machine: a car jack being used to raise one side of a car.

Figure A3.25 Changing a car tyre using a simple machine (car jack)

By changing the design of a car jack, it is possible in theory to raise the car with any sized force that we choose. For example, a force of 50 N may raise the wheel off the ground if the handle is rotated 10 times, whereas a force of 10 N would require about 50 rotations of the handle to transfer the same amount of energy. The heavy weight of the car will produce a lot of friction in the car jack.

The construction of many ancient structures was only possible because of the simple machines seen in Figure A3.24. Research online to learn some details of how the pyramids (or any other building of a similar age) were made. Compare the information you retrieve from different sources. Do the sources agree? Which sources do you consider most credible? Why?

Elastic potential energy

When a spring, or an elastic material, is deformed from the shape it had when there was no force acting on it, as in Figure A3.26, it will become a store of elastic potential energy that can later be used to transfer useful energy when the spring / material returns to its original shape.

Remember Hooke's law for elastic stretching from Topic A.2: $F_H = -kx$, where k is the 'spring constant'. Note also that, if a force on a spring / elastic material increases from 0 to F_H, the average force used during the deformation is $\frac{1}{2}F_H$, assuming that the deformation is proportional to the force (as shown in Figure A3.26).

◆ **Lever** A simple machine consisting of a rigid bar and a pivot. Used to change the direction and magnitude of a force.

◆ **Pulley** A rotating wheel used to change the direction of a force. When two or more pulleys are combined, the system can reduce the force needed to do work.

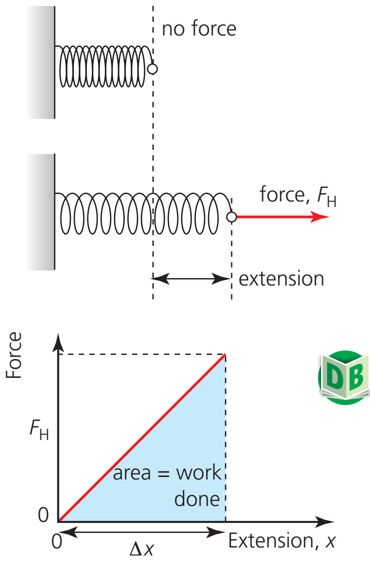


Figure A3.26 Force and extension when stretching a spring

For a spring / material which obeys Hooke's law, the work done, W, when it is deformed a distance Δx by a force F_H, is given by:

$$W = \text{average force} \times \text{distance} = \frac{1}{2}F_H \times \Delta x = \frac{1}{2}(k\Delta x)x = \frac{1}{2}k\Delta x^2$$

The work done is equal to the stored elastic potential energy, E_H. (There is no need to include a negative sign because work is not a vector quantity.)

The elastic potential energy of a deformed spring / material which obeys Hooke's law can be calculated by using:

$$\boldsymbol{E_H = \frac{1}{2}k\Delta x^2}.$$

As explained earlier in this topic, the work done is equal to the area under force–extension graph.

WORKED EXAMPLE A3.7

An open-coiled compression spring, like the one seen in Figure A3.27 has an overall length of 5.64 cm. A compressive force of 59.0 N reduced its length to 5.30 cm.

a Determine the spring constant, assuming that Hooke's law was obeyed.

b Calculate how much work was done to compress the spring.

c Use $E_H = \frac{1}{2}k\Delta x^2$ to confirm the amount of elastic potential energy stored in the spring.

Figure A3.27 An open-coiled compression spring

Answer

a $k = \frac{F_H}{x} = \frac{59.0}{(5.64 - 5.30)} = 174\,\text{N cm}^{-1}$ (or $1.74 \times 10^4\,\text{N m}^{-1}$)

b $W = \frac{1}{2}F_H\Delta x = 29.5 \times (5.64 - 5.30) \times 10^{-2} = 0.100\,\text{J}$

c $E_H = \frac{1}{2}k\Delta x^2 = 0.5 \times (1.74 \times 10^4) \times (3.4 \times 10^{-3})^2 = 0.100\,\text{J}$

The answers to **b** and **c** are the same, as we would expect.

See Figure A3.6 for an example of how to determine elastic potential energy from a force–extension graph.

17 How much potential energy is transferred when:

- **a** a 1.2 kg box is raised from the floor to a table top 0.85 m higher
- **b** a 670 g book falls to the floor from the same table?

18 A cable car rises a vertical height of 700 m in a total distance travelled of 6.0 km.

- **a** Show that approximately 15 MJ of gravitational potential energy must be transferred to a car of mass 1800 kg during the journey if it has six passengers with an average mass of 47 kg.
- **b** Suggest why considerably more energy (than your answer to part **a** has to be transferred in making this journey.

Figure A3.28 Ngong Ping cable car in Hong Kong

19 A rocket launches a 500 kg satellite to a height of 400 km above the Earth's surface.

- **a** Outline why the equation $\Delta E_p = mg\Delta h$, with $g = 9.8\,\text{N kg}^{-1}$ cannot be used to accurately determine the gravitational potential energy that has to be transferred to the satellite.
- **b** However, the actual value of g at that height has not reduced as much as many students expect: during the launch it reduced from $9.8\,\text{N kg}^{-1}$ to a value of $8.7\,\text{N kg}^{-1}$ at a height of 400 km. Predict the gravitational potential energy gained by the satellite.

20 A spring has a spring constant of $384\,\text{N m}^{-1}$ and is stretched by 2.0 cm.

- **a** Calculate the elastic potential energy stored in the spring. Assume that it obeys Hooke's law.
- **b** Predict a value for the extension that would be needed to store 1.0 J of energy.
- **c** Explain why the answer to part **b** is uncertain.

21
- **a** Calculate the work done in raising the centre of gravity of a trampolinist of mass 62 kg through a vertical height of 3.48 m (see Figure A3.29).
- **b** When he lands on the trampoline he is brought to rest for a moment before being pushed up in the air again. If the maximum displacement of the trampoline is 0.90 m, sketch a possible force–displacement graph for the surface of the trampoline.

Figure A3.29 The more the trampoline stretches, the higher the trampolinist can jump

Conservation of mechanical energy

SYLLABUS CONTENT

- In the absence of frictional resistive forces, the total mechanical energy of a system is conserved.
- If mechanical energy is conserved, work is the amount of energy transformed between different forms of mechanical energy in a system, such as: kinetic energy, gravitational potential energy and elastic potential energy.

Applying the law of conservation of energy to mechanical systems:

kinetic energy + gravitational potential energy + elastic potential energy = constant

◆ **Conservative force** A force, the action of which conserves mechanical energy. There is no energy dissipation.

But this is only true if there are no frictional (resistive) forces acting. Such forces can be described as 'non-conservative'. **Conservative forces** (such as gravitational forces, for example) conserve mechanical energy and do not involve the dissipation of energy. (More precisely: a *conservative force* is one for which the total work done in moving between two points is independent of the path taken.)

We have already explained that energy dissipation occurs in all mechanical systems to a greater or lesser extent, because of ever-present frictional / resistive forces (non-conservative forces). However, the equation above remains very useful for predicting 'ideal' outcomes, and for determining the amount of energy dissipated in other situations.

One of the most common examples of mechanical energy transfers is that between gravitational potential energy and kinetic energy. Assuming no resistive forces: change of gravitational potential energy = change of kinetic energy

$$mg\Delta h = \frac{1}{2}mv^2 - \frac{1}{2}mu^2$$

If the mass starts or ends with zero velocity during the time being considered: v^2(or u^2) = $2g\Delta h$. leading to: v (or u) = $\sqrt{2g\Delta h}$.

This equation can be used to relate height with speed for any mass falling from rest, or any mass projected upwards to its highest point (assuming that gravity is the only force acting).

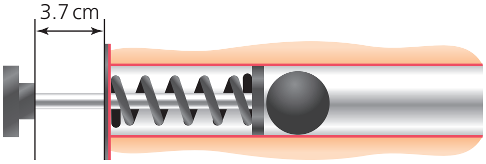
Note that the same equation can be obtained using the equation of motion (see Topic A.1): $v^2 = u^2 + 2as$, with u or v equal to zero, and using s instead of h.

WORKED EXAMPLE A3.8

A ball is thrown upwards with a speed of 23 m s^{-1}. Calculate the maximum height that it can reach.	**Answer** $u^2 = 2g\Delta h$ $23^2 = 2 \times 9.8 \times \Delta h$ $\Delta h = 27$ m

WORKED EXAMPLE A3.9

A ball of mass 7.8 g was fired horizontally by a spring, as shown in Figure A3.30. The spring has a spring constant of 620 N m^{-1} and was pulled back 3.7 cm from its uncompressed length.



■ **Figure A3.30** Ball being fired horizontally by a spring

a Calculate how much elastic potential energy was stored in the spring. Assume that it obeys Hooke's law.

b What average force was used to compress the spring?

c Determine the maximum possible speed of the ball after the spring was released.

d Explain why the actual speed will be less than your answer to part **c**.

Answer

a $E_H = \frac{1}{2}k\Delta x^2 = 0.5 \times 620 \times 0.037^2 = 0.42\text{ J}$

b $E_H = W = Fs$

$0.42 = F \times 0.037$

$F = 11\text{ N}$ (This was half of the maximum force.)

c $0.42 = \frac{1}{2}mv^2 = 0.5 \times 0.0078 \times v^2$

$v = 10\text{ m s}^{-1}$

d All of the elastic potential energy in the spring was not transferred to the kinetic energy of the ball. For example, there was some friction with the sides of the tube, the spring did not stop moving after propelling the ball, there was some sound produced.

WORKED EXAMPLE A3.10

A box of mass 4.7 kg slid down a slope of vertical height 80 cm.

a Calculate the gravitational potential energy of the box at the top of the slope (compared to the bottom of the slope).

b Assuming the conservation of mechanical energy, what would the speed of the box be at the bottom of the slope?

c But the actual speed of the box was measured to be 2.2 m s^{-1}. Explain why the speed was less than the answer to part **b**.

d Determine how much energy was dissipated into the surroundings.

e In what form(s) was this dissipated energy?

Answer

a $\Delta E_p = mg\Delta h = 4.7 \times 9.8 \times 0.80 = 37\text{ J}$

b $37 = \frac{1}{2}mv^2 = 0.5 \times 4.7 \times v^2$

$v = 4.0\text{ m s}^{-1}$

Note that this answer (which assumes no energy dissipation) would be the same for any mass on any slope, or a vertical fall.

c Friction with the slope (and a little air resistance) acted in the opposite direction to motion.

d $37 - \frac{1}{2}mv^2 = 37\text{–}\ (0.5 \times 4.7 \times 2.2^2) = 26\text{ J}$

e Internal energy in the surfaces of the slope and box (which then spreads out as thermal energy).

22 What is the maximum speed with which a mass can hit the ground after being dropped from a height of 1.80 m?

23 The ball shown in Figure A3.31 was released from rest in position A. It accelerated down the slope and had its highest speed at the lowest point. It then moved up the slope on the other side, reaching its highest point at B.

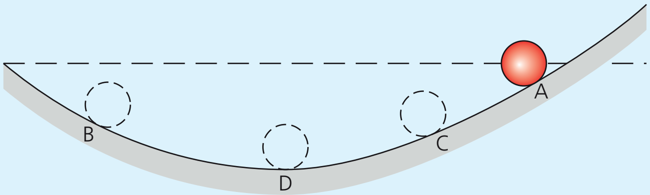


Figure A3.31 Ball rolling down slope

a Explain why B is lower than A.

b Describe the motion of the ball after leaving position B, explaining the energy transfers until it finally comes to rest at D.

24 The large angle swing of a pendulum is interrupted by a peg as shown in Figure A3.32. Sketch a copy of the diagram and indicate the position of the pendulum after it has momentarily stopped moving on the right-hand side of the peg.

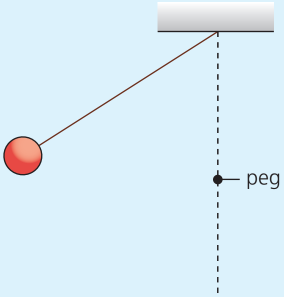


Figure A3.32 Pendulum

25 Figure A3.33 shows an experiment to measure the speed of a bullet which was fired into a block of wood. The bullet is embedded in the wood, so this was a totally inelastic collision. The block of wood and the bullet had kinetic energy and this was transferred to gravitational potential energy as they swung upwards.

a Outline what is meant by a totally inelastic collision.

b If the combined mass of the block and the bullet was 1.23 kg, determine their maximum gravitational potential energy.

c Use the law of conservation of energy to show that the initial velocity of the combined block and bullet was 1.5 m s^{-1}.

d If the bullet's mass was 15 g, use the law of conservation of momentum to determine its speed.

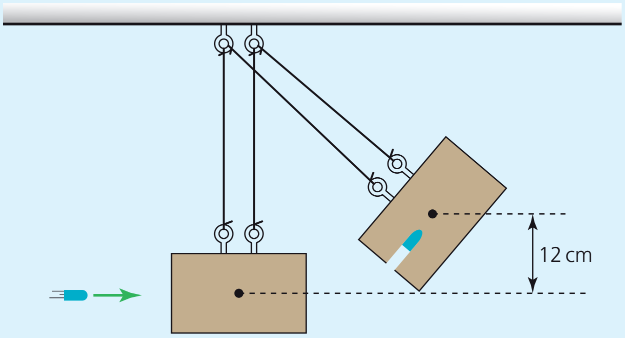


Figure A3.33 Experiment to measure the speed of a bullet

26 a A 10 g steel sphere moving to the left at 2.0 m s^{-1} collided with a similar sphere of mass 2.0 g moving in the opposite direction at 4.0 m s^{-1}. If after the collision the 10 g sphere remained stationary, determine what happened to the other sphere.

b Calculate the total kinetic energy:

- **i** before the collision
- **ii** after the collision.

c Was mechanical energy conserved in this collision?

d State the term we use to describe collisions like this.

27 In a laboratory experiment an 8.6 g wooden sphere moving at 0.39 m s^{-1} collided with a 5.7 g wooden sphere moving in the opposite direction with a speed of 0.72 m s^{-1}. After the collision they were both observed to move with a speed of 0.25 m s^{-1}, but in opposite directions.

a Show that these results are in good agreement with the law of conservation of momentum.

b Calculate the total kinetic energy:

- **i** before the collision
- **ii** after the collision.

b Was kinetic energy conserved in this collision?

c State the term we use to describe collisions like this.

28 A rubber band of mass 1.2 g was extended by 8.4 cm. The extension was proportional to the force and the band had a spring constant of 280 N m^{-1}

a If the force was released and the band fired vertically upwards, predict the maximum theoretical height that it could reach.

b Explain why, in practice, the height will be a lot less than your answer to part **a**.

29 A long steel wire of mass 150 g was extended by 3.2 cm by a force which had increased slowly from 0 N to 240 N.

- **a** Assuming that the extension was proportional to the force, how much elastic potential energy was stored in the wire?
- **b** The wire then snapped and the stored energy was transferred to the kinetic energy of the wire. Calculate an average value for the speed of the wire.
- **c** Discuss why breaking some metal wires can be dangerous.

30 'Crumple zones' are a design feature of most vehicles (Figure A3.34). They are designed to compress and deform permanently if they are in a collision. Use the equation $Fs = \frac{1}{2}mv^2$ to help explain why a vehicle should not be too stiff and rigid.

Figure A3.34 The front of the car is deformed but the passenger compartment is intact

31 A bungee jumper (Figure A3.35) of mass 61 kg is moving at 23 m s^{-1} when the rubber bungee cord begins to become stretched.

- **a** Calculate her kinetic energy at that moment.
- **b** Figure A3.36 shows how the extension of the cord varies with the applied force. State what quantity is represented by the area under this graph.
- **c** Describe the relationship between force and extension shown by this graph.
- **d** Use the graph to estimate how much the cord has extended by the time it has brought the jumper to a stop.

Figure A3.35 Bungee jumping in Taupo, New Zealand

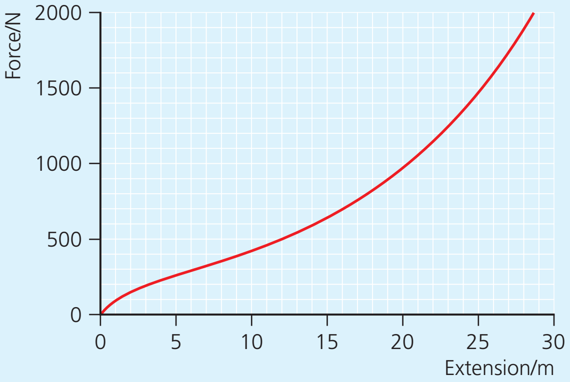


Figure A3.36 Force–extension graph for a bungee cord

32 A pole-vaulter of mass 59.7 kg falls from a height of 4.76 m onto foam.

- **a** Calculate the maximum possible kinetic energy on impact.
- **b** Will air resistance have had a significant effect in reducing the velocity of impact? Explain your answer.
- **c** If the foam deforms by 81 cm, estimate the average force exerted on the pole-vaulter.

ATL A3B: Thinking skills

Apply key ideas and facts in new contexts

Regenerative braking

Most ways of stopping moving vehicles involve braking systems in which the kinetic energy of the vehicle is transferred to internal energy because of friction in the braking system and with the ground / track. The internal energy is dissipated into the surroundings as thermal energy and cannot be recovered.

The kinetic energy of a long, fast-moving train is considerable. Values of 10^8 J, or more, are not unusual. When the train stops, all of that energy has to be transferred to other forms and, unless the energy can be recovered, the same amount of energy then has to be transferred to accelerate the train again. This is very wasteful, so the train and its operation should be designed to keep the energy wasted to a minimum. One way of doing this is to make sure that large, fast trains stop at as few stations as possible, perhaps only at their origin and final destination.

A lot of research has gone into designing efficient **regenerative braking** systems in recent years, usually involving the generation of an electric current, which can be used to transfer energy to stored chemical changes in batteries, or to operate on-board electrical equipment.

The process of converting the kinetic energy of the train into electrical energy decelerates the train, so that there is much less need for frictional braking. This also reduces the wear on the brakes and the thermal energy transferred to the environment.

Small electric trains, which are often operated underground or on overhead tracks, are a feature of most large cities around the world (Figure A3.37). Such trains usually have stations every few kilometres or less, so regenerative braking systems and other energy-saving policies are very important.

■ **Figure A3.37** The Light Rail Transit trains on the SBS network in Singapore have regenerative braking

When designing a new urban train system, it has been suggested that energy could be saved by having a track shaped as shown in Figure A3.38.

What might be the advantages of such a track profile? Consider the energy transfers taking place as the train moves between stations.

■ **Figure A3.38** Possible track profile

◆ **Regenerative braking** Decelerating a vehicle by transferring kinetic energy into a form that can be of later use (rather than dissipating the energy into the surroundings). For example, by generating an electric current that charges a battery.

Power

SYLLABUS CONTENT

▶ Power, P, is the rate of work done, or the rate of energy transfer, as given by: $P = \frac{E}{t} = Fv$.

Power is the rate of transferring energy. When energy is transferred by people, animals or machines to do something useful, we are often concerned about how much time it takes for the change to take place. If the same amount of useful work is done by two people (or machines), the one that does it faster is said to be more *powerful*. (In everyday use the word power is used more vaguely, often related to strength and without any connection to time.)

$$\text{power} = \frac{\text{energy transferred}}{\text{time taken}} \qquad P = \frac{E}{t} \qquad \text{SI unit: watt, W}$$

$$\text{Alternatively: power} = \frac{\text{work done}}{\text{time taken}} \qquad \boldsymbol{P = \frac{\Delta W}{\Delta t}}$$

◆ **Power, *P*** $\frac{\text{energy transferred}}{\text{time taken}}$ or, for mechanical energies, $\frac{\text{work done}}{\text{time taken}}$.

◆ **Watt, W** Derived SI unit of power. 1 W = 1 J s^{-1}.

1 W = 1 J s^{-1}. The units mW, kW, MW and GW are all in common use. (Avoid confusion between work done, *W* and the unit the watt, W.)

The following are some examples of values of power in everyday life.

Figure A3.39 Transferring energy at a rate of about 250 W

- A 0.0001 W calculator transfers energy at a rate of 0.0001 J every second.
- A 7 W light bulb transfers energy from electricity to light and thermal energy at a rate of 7 J every second.
- A girl walking upstairs may transfer chemical energy to gravitational energy at a rate of about 250 W.
- In many countries, homes use electrical energy at an average rate of about 1 kW.
- A 2 kW water heater transfers energy from electricity to internal energy at a rate of 2000 J every second.
- A typical family car might have a maximum output power of 100 kW.
- A 500 MW oil-fired power station transfers chemical energy to electrical energy at a rate of 500 000 000 J every second.

WORKED EXAMPLE A3.11

Calculate the average power of a 65 kg climber moving up a height of 40 m in 3 minutes.

Answer

$$P = \frac{\Delta W}{\Delta t} = \frac{(mg\Delta h)}{\Delta t} = \frac{(65 \times 9.8 \times 40)}{3 \times 60} = 1.4 \times 10^2\ \text{W}$$

Power needed to maintain a constant speed

◆ **Resistive force, *F*** Any force that opposes motion, for example friction, air resistance, drag.

It is common for a vehicle to maintain a constant velocity. Under those circumstances, the forward force, *F*, from the engines is equal in magnitude, but opposite in direction to the magnitude of the total **resistive forces**, ***F***. Then:

Power, *P*, needed to maintain a constant velocity, *v*, against the resistive forces, can be determined from:

$$P = \frac{\Delta W}{\Delta t} = \frac{F\Delta s}{\Delta t} = Fv$$

DB

Power needed to maintain a constant velocity, $\boldsymbol{P = Fv}$

WORKED EXAMPLE A3.12

a What average power is needed to accelerate a 1600 kg car from rest to 25 m s^{-1} in 12.0 s?

b What power is needed to maintain the same speed if the resultant resistive force is a constant 2300 N?

Answer

a $P = \frac{\Delta W}{\Delta t} = \frac{\text{kinetic energy gained}}{\text{time taken}} = \frac{\frac{1}{2}mv^2}{\Delta t} = \frac{(0.5 \times 1600 \times 25^2)}{12.0} = 4.2 \times 10^4\ \text{W}\ (= 42\ \text{kW})$

b $P = Fv = 2300 \times 25 = 5.8 \times 10^4\ \text{W}\ (= 58\ \text{kW})$

33 a How much useful energy must be transferred to lift twelve 1.7 kg bottles from the ground to a shelf that is 1.2 m higher?

b If this task takes 18 s, what was the average useful power involved?

34 Estimate the output power of an electric motor that can raise an elevator of mass 800 kg and six passengers 38 floors in 52 s. (Assume there is no counterweight.)

35 a Calculate the average power needed for a cyclist of mass 72.0 kg to accelerate from 8.00 m s^{-1} to 12.0 m s^{-1} in 22.0 s on a horizontal road. Assume that the resistive forces are negligible and the bicycle has a mass of 8.00 kg.

b Compare your answer to 6.5 W kg^{-1} (using body mass) for the best athletes.

36 A small boat is powered by an outboard motor with a maximum output power of 40 kW. The greatest speed of the boat is 27 knots (1 knot = 1.85 km h^{-1}). Determine the magnitude of the forward force provided by the motor at this speed.

37 a What is the constant speed of a car which has an output power of 22 kW when the resistive forces are 2.0 kN?

b What assumption did you make in answering part **a**?

38 What is the output power of a jet aircraft that has a forward thrust of 6.60 × 10^5 N when travelling at its top speed of 950 km h^{-1} (264 m s^{-1}) through still air?

39 a Find out which countries of the world have the highest average power consumption per person.

b Suggest reasons why they use so much energy.

Efficiency

SYLLABUS CONTENT

- Efficiency, η, in terms of energy transfer or power, as given by: $\frac{E_{\text{output}}}{E_{\text{input}}} = \frac{P_{\text{output}}}{P_{\text{input}}}$.

It is an ever-present theme of physics that, whatever we do, some of the energy transferred is ‘wasted’ (dissipated) because it is transferred to less ‘useful’ forms. In mechanics this is usually because friction, or air resistance, transfers kinetic energy to internal energy and thermal energy. The useful energy we get out of any energy transfer is *always* less than the total energy transferred.

When an electrical water heater is used, nearly all of the energy transferred makes the water hotter and it can therefore be described as ‘useful’, but when a mobile phone charger is used, for example, only some of the energy is transferred to the battery (most of the rest is transferred to thermal energy). Driving a car involves transferring chemical energy from the fuel and the useful energy is considered to be the kinetic energy of the vehicle, although at the end of the journey there is no kinetic energy remaining.

◆ **Efficiency (thermodynamic), η** Ratio of useful energy (or power) output to the total energy (or power) input.

A process that results in a greater useful energy output (for a given energy input) is described as being more efficient. In thermodynamics, **efficiency** is defined as follows:

$$\textbf{efficiency, } \eta = \frac{\textbf{useful energy output}}{\textbf{total energy input}} = \frac{E_{\text{output}}}{E_{\text{input}}}$$

Dividing energy by time to get power, we see that efficiency may also be defined as:

$$\textbf{efficiency, } \eta = \frac{\textbf{useful power output}}{\textbf{total power input}} = \frac{P_{\text{output}}}{P_{\text{input}}}$$

Because it is a ratio of two energies (or powers), efficiency has no units. It is often expressed as a percentage. It should be clear that, because of the principle of conservation of energy, efficiencies will always be less than 1 (or 100%).

It is possible to discuss the efficiency of any energy transfer, such as the efficiency with which our bodies transfer the chemical energy in our food to other forms. However, the concept of efficiency is most commonly used when referring to electrical devices and engines of various kinds, especially those in which the input energy or power is easily calculated. Sometimes we need to make it clear exactly what we are talking about. For example, when discussing the efficiency of a car, do we mean only the engine, or the whole car in motion along a road with all the energy dissipation due to resistive forces?

The efficiencies of machines and engines usually change with the operating conditions. For example, there will be a certain load at which an electric motor operates with maximum efficiency; if it is used to raise a very small or a very large mass it will be less efficient. Similarly, cars are designed to have their greatest efficiency at a certain speed, usually about 100 km h^{-1}. If a car is driven faster (or slower), then its efficiency decreases, which means that more fuel is used for every kilometre travelled.

Car engines, like all other engines that rely on burning fuels to transfer energy, are inefficient because of fundamental physics principles (see Topic B.4). There is nothing that we can do to change that, although better engine design and maintenance can make some improvements to efficiency.

In recent years we have all become very aware of the need to conserve the world's energy resources and limit the effects of burning fossil fuels in power stations and various modes of transport on global warming (see Topic B.2). Improving the efficiency of such 'heat engines' has an important role to play in this worldwide issue.

■ **Figure A3.40** Power stations which use natural gas have the greatest overall efficiency (≈ 60%)

WORKED EXAMPLE A3.13

Figure A3.41 shows an experiment designed to measure the efficiency of a small electric motor. Electrical power was supplied to the motor at a rate of 0.72 W. The hanging mass (25 g) went up a distance of 1.12 m in 2.46 s.

a Calculate how much gravitational potential energy was transferred to the mass.

b How much electrical energy was transferred in this time?

c Determine the efficiency of the motor.

d How much energy was dissipated into the surroundings?

e State the forms of this dissipated energy.

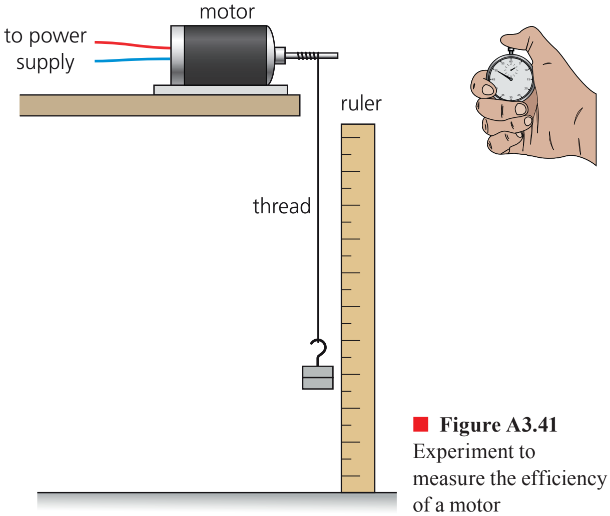


Figure A3.41 Experiment to measure the efficiency of a motor

Answer

a $\Delta E_p = mg\Delta h = 0.025 \times 9.8 \times 1.12 = 0.27\,\text{J}$ (seen on calculator display as 0.2744...)

b power, $P = \dfrac{\text{energy transferred}}{\text{time taken}}$

energy transferred $= P \times \Delta t = 0.72 \times 2.46 = 1.8\,\text{J}$ (seen on calculator display as 1.7712...)

c efficiency, $\eta = \dfrac{\text{useful energy output}}{\text{total energy input}} = \dfrac{0.2744}{1.7712} = 0.15$ (15%)

d $1.7712 - 0.2744 = 1.5\,\text{J}$

e internal energy, thermal energy (+ sound)

We will see in Topic B.5 that the power input to any electric device can be determined from:

power = voltage × current.

Inquiry 3: Concluding and evaluating

Concluding

Figure A3.42 represents the apparatus used by a student to investigate the efficiency of a pulley system.

The student measured the forces needed to lift loads of 0.5 kg, 1.0 kg, 4.0 kg and 5.0 kg: 0.6 N, 1.5 N, 7.5 N and 12.7 N. He assumed that the force always moves four times further than the load. His conclusion was that the efficiency of the system was 66%.

1 Is this a correct conclusion from the data collected?

2 Discuss the quantity and range of readings taken by the student.

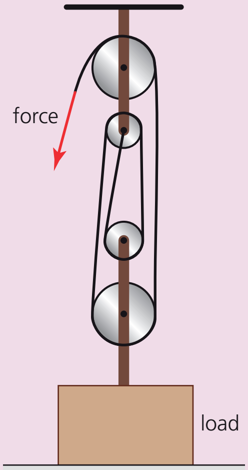


Figure A3.42 Pulley system

Energy density

SYLLABUS CONTENT

- Energy density of the fuel sources.

◆ **Energy density** The energy transferred from unit volume of fuel (SI unit: $J m^{-3}$).

◆ **Specific energy** Amount of energy that can be transferred from unit mass of an energy source (SI unit: $J kg^{-1}$).

One of the advantages of fossil fuels is the large amount of energy that can be transferred from each cubic metre of the fuels. Although, nuclear fuels are much more energy dense.

Energy density is the amount of energy that can be transferred from each cubic metre.

SI unit: $J m^{-3}$

Table A3.4 shows some typical energy densities of energy sources.

When discussing gaseous energy resources, it is more common to use **specific energy**, which is the amount of energy that can be transferred from each kilogram. For example, the specific energy of natural gas is $55 MJ kg^{-1}$.

Table A3.4 Some approximate energy densities

Source	Energy density / $MJ m^{-3}$
Reactor-grade uranium-235	66 000 000 000
coal	43 000
gasoline (petrol)	36 000
crude oil	37 000
ethanol	24 000
wood	15 000
electrical batteries	1000
hydroelectric	1

Tool 2: Technology

Represent data in a graphical form

Use data from the internet to determine specific energies for a variety of different energy sources.

Use a spreadsheet to present the information in the form of a bar chart.

ATL A3C: Research skills

How did you check the reliability of your online sources?

Common mistake

Energy density and specific energy are often confused. Some sources define energy/mass as energy density.

Inquiry 1: Exploring and designing

Designing

Controlling heat losses

Figure A3.43 shows how a student intended to investigate the energy density of ethanol. She knows that for every 4184 J of energy that are transferred to 1.0 kg of water, the temperature will rise by 1.0 °C (see Topic B.2).

1 Identify the student's choice of dependent, independent and control variables. How would the results be used to calculate energy density?

2 Suggest how the experiment (as shown) could be improved. Hint: how much of the energy stored in the fuel is transferred to the water?

3 What other energy transfers are taking place?

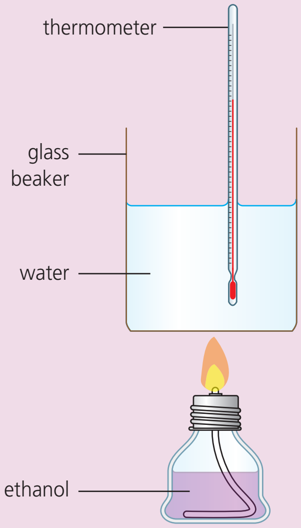


Figure A3.43 Estimating the energy density or specific energy of a fuel

40 The power output from a natural gas-fired electrical power station is 540 MW.

- **a** If its efficiency is 48%, what is the input power?
- **b** Calculate how much fuel the power station uses in one hour if 49 MJ can be obtained from each kilogramme.

41 An elevator (lift) which has a mass of 2400 kg when empty is connected as shown in Figure A3.44 to a counterweight of the same mass.

When the elevator goes up, the counterweight goes down. Five people of total mass 265 kg got in the elevator and went up four floors, each floor of height 3.2 m.

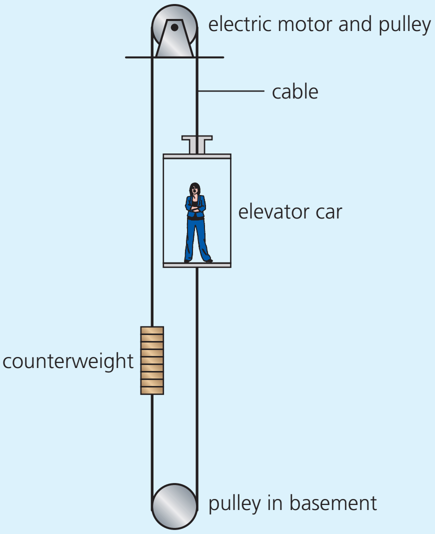


Figure A3.44 An elevator (lift) and its counterweight

- **a** Outline the reason for using a counterweight.
- **b** Calculate the useful work done by the motor.
- **c** If, in fact, 1.3×10^5 J of energy was transferred from the electrical supply to the motor, determine the efficiency of the process.
- **d** If the process took a total time of 18 s, what was the average input power to the motor?
- **e** How much energy was transferred in overcoming friction?
- **f** Discuss where friction would occur in this system.

42 Explain why it may be more useful to refer to the specific energy of natural gas, rather than its energy density.

43 A petrol-driven car was accelerated from rest in order to determine its overall efficiency.

- **a** Calculate the efficiency if it gained kinetic energy of 4.0×10^5 J while using 55 ml (5.5×10^{-5} m^3) of fuel.
- **b** Draw an annotated Sankey diagram to represent this process.
- **c** Discuss whether it is reasonable to state that the car has zero efficiency while travelling with constant speed.

44 An airplane has a mass of 200 tonnes (2.0×10^5 kg) and take-off speed of 265 $km\,h^{-1}$ (73 $m\,s^{-1}$) at the end of a distance of 2.24 km from where it began.

- **a** Calculate the kinetic energy of the airplane when it takes off.
- **b** Estimate the average power output from the airplane's engines while it is on the runway.
- **c** What average resultant forward force was acting on the airplane during its movement along the runway?
- **d** If 76 kg of fuel was used during take off, calculate the efficiency of the process if 1 kg of fuel can transfer 43 MJ.

45 An oil burning power station has an efficiency of 39% and an output of 770 MW.

- **a** Calculate the mass of oil burned:
  - **i** every second
  - **ii** every year.
- **b** Estimate how much reactor grade uranium-235 would be needed every year to produce the same output power (assume the same efficiency).

46 Show that the energy density of wind blowing at a speed of 5 $m\,s^{-1}$ (as could be used with a wind turbine to generate electricity) is about 15 $J\,m^{-3}$. The density of air is 1.3 $kg\,m^{-3}$.

ATL A3D: Thinking skills

Evaluating and defending ethical positions

After we have used any mode of transport, all of the energy used by the vehicle will have been dissipated into the environment. In a scientific sense, their efficiency is zero but, of course, they will have normally served a useful purpose.

For the sake of reducing pollution, conserving materials and limiting global warming, do individuals have any responsibility for limiting their travel? Or do people have the right to travel wherever and whenever they like, in whatever mode of transport they choose?

Should governments enact laws to limit our travel, and/or introduce or raise taxation on transportation and its fuels?

A.4 Rigid body mechanics

Guiding questions

- How can the understanding of linear motion be applied to rotational motion?
- How is the understanding of the torques acting on a system used to predict changes in rotational motion?
- How does the distribution of mass within a body affect its rotational motion?

Rotational dynamics

◆ **Dynamics** The science which explains the motion of objects.

◆ **Rotate** To move around a central point or axis (usually inside the rotating object).

◆ **Revolve** To move around a central point or axis (usually outside of the revolving object).

◆ **Rigid** Does not change shape.

◆ **Extended object** An object that has dimensions. Not a point.

◆ **Axis of rotation** Line about which an object can rotate.

Dynamics is the name we give to that branch of physics which is concerned with motion and its causes (forces). In Topics A.1, A.2 and A.3 we have mostly considered linear dynamics: bodies moving in straight lines. In this topic we will study the rotational motion of objects/bodies about a fixed axis. The term **rotation** usually refers to movement about an axis within the body, for example the rotation of the Earth every 24 hours. The rotation of a body does not affect its location, whereas **revolution** usually concerns movement of a body around an exterior point, for example the Earth revolves around the Sun every year. (Some situations may be described as a rotation or a revolution.)

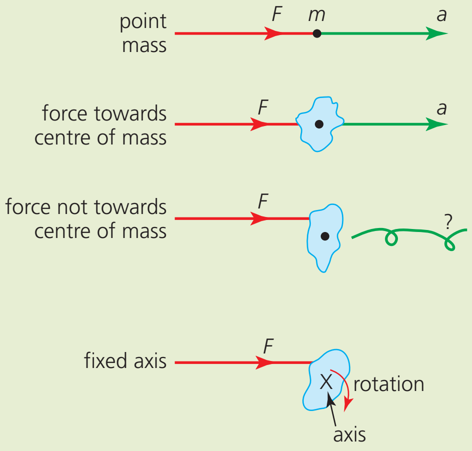
We will assume that the bodies are **rigid**, that is, their shape does not deform significantly under the action of forces. The simplest everyday examples of rotation include a wheel and a door handle.

Nature of science: Models

Point objects and extended objects

In Topics A.1–A.3 we have generally considered all bodies to be point objects. This was done in order to simplify the situations. This topic is different, as the focus of our attention will be on **extended objects** which are able to rotate.

We know that a resultant force acting on a point mass results in a linear translational acceleration ($F = ma$ from Topic A.2). This remains true if the force acts on an extended object, but only if the force is directed at the centre of mass of the object. However, when resultant forces are directed elsewhere on objects, the situations are complicated and the results will depend on how the magnitude and direction of the force vary after initial contact. See Figure A4.1.



■ **Figure A4.1** Resultant forces acting on point objects and extended objects

All of Topic A.4 is about rotations caused by forces on objects which are able to rotate in the same place because they have a fixed **axis of rotation** (wheels for example).

Comparing rotational motion to linear motion

We can greatly simplify our introduction to **rotational dynamics** by using our existing knowledge and understanding of linear dynamics. All the concepts of linear dynamics have rotational dynamics equivalents, with similar equations. They are summarized in Table A4.1.

Table A4.1 Comparing linear and rotational motion

Linear motion	Rotational motion
force	torque
mass (a measure of inertia)	moment of inertia
linear displacement	angular displacement
linear speed / velocity	angular speed / velocity
linear acceleration	angular acceleration
linear momentum	angular momentum
linear kinetic energy	rotational kinetic energy

◆ **Rotational dynamics** Branch of physics and engineering that deals with rotating objects.

◆ **Analogy** Applying knowledge of one subject to another because of some similarities.

TOK

Knowledge and the knower

- How do we acquire knowledge?

Analogies

An **analogy** is a useful comparison between two different things that have some features in common, with the intention that knowledge of one can be applied to the other. Making an analogy between linear and rotational motion is probably not surprising and, as we shall see, it is very useful. However, many other analogies may not be so obvious. For example, can the study of economics find useful analogies in the laws of physics?

Apart from assisting in the teaching and learning of a new situation, there may be two major purposes for using analogies.

1 An analogy may be used to make reliable predictions about the behaviour of the system to which it is applied and that, in itself, may be sufficient reason to justify the use of an analogy.

2 An analogy may help to provide a deeper understanding of the system to which it is applied. However, without further justification, analogies should not be assumed to be accurate descriptions of the systems to which they are applied.

Torque

SYLLABUS CONTENT

▶ The torque, τ, of a force about an axis, as given by: $\tau = Fr \sin\theta$

In situations where rotation may be possible, it is important to identify the place about which the rotation can occur. Most commonly this will involve an axis of rotation. (The terms **pivot**, **hinge** and **fulcrum** are widely used for various situations in which movement is not complete, nor continuous.)

A straight line showing the direction in which a force is applied is called its **line of action**. Any force applied to an object whose line of action is not through the axis of rotation will tend to start, or change, rotational motion, if that is possible.

Bigger forces will tend to produce larger rotational accelerations but the line of action (direction) of a force is also very important. See Figure A4.2.

◆ **Pivot** A fixed point supporting something which turns or balances.

◆ **Hinge** Device which connects two solid objects allowing one (or both) to rotate in one direction.

◆ **Fulcrum** See *pivot*.

◆ **Line of action (of a force)** A line through the point of action of a force, showing the direction in which the force is applied.

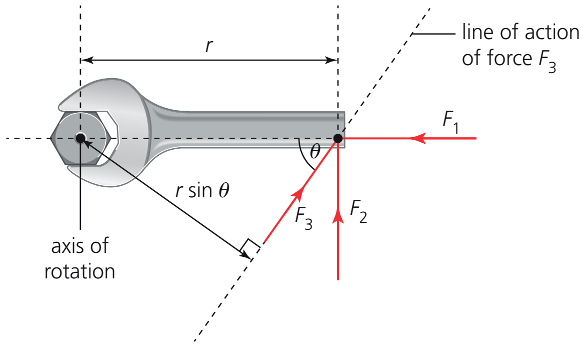


Figure A4.2 Forces producing rotation of a spanner (wrench) and bolt

In Figure A4.2, F_1 has no turning effect because its line of action is through the axis of rotation. F_2 has the biggest turning effect because its line of action is perpendicular to a line joining its point of application to the axis. F_3 has an effect between these two extremes. The turning effect also depends on the distance, r, from the axis of rotation to the line of action of the force.

The 'turning effect' of a force, F, is known as its **torque**, τ , and it depends on the magnitude of the force and the perpendicular distance from the axis of rotation to the line of action of the force. In Figure A4.2 this is shown for force F_3 as $r \sin\theta$, where θ is the angle between the line of action of the force and a line joining the point of application of the force to the axis of rotation; r is the distance from the axis of rotation to the point of application of the force.

DB

torque $\tau = Fr \sin\theta$

When there is no actual rotation, torque is sometimes called the **moment** of a force. (You may be familiar with the '**principle of moments**' for a body in equilibrium.)

Torque has the SI unit Nm (not Nm^{-1}) but note that it is not equivalent to the unit of energy, the joule, which is also Nm.

◆ **Torque** Product of a force and the perpendicular distance from the axis of rotation to its line of action.

◆ **Moment (of a force)** Term sometimes used as an alternative to torque, especially if rotation is incomplete.

◆ **Principle of moments** If an object is in rotational equilibrium, the sum of the clockwise moments (torques) equals the sum of the anticlockwise moments (torques).

Inquiry 1: Exploring and designing

Exploring

The right tool for the job

A torque wrench is a device which limits the torque that can applied when tightening a bolt. This is to prevent over-tightening and damage to the bolt and, for example, an engine. There are various designs. Figure A4.3 shows a modern digital type. Demonstrate insight by reflecting on the image to suggest how this type of torque wrench might be used. Check your ideas through research.

Figure A4.3 Torque wrench

WORKED EXAMPLE A4.1

Look again at Figure A4.2.

a If $r = 48$ cm, calculate the torque produced by a force of 35 N applied along the line of action of F_2.

b Determine the value of F_3 that would produce the same torque as in part **a**, if the angle $\theta = 55°$.

Answer

a $\tau = Fr\sin\theta = 35 \times 0.48 \times \sin 90° = 17\,\text{Nm}$

b $17 = F_3 \times 0.48 \times \sin 55°$

$$F_3 = \frac{17}{0.393} = 43\,\text{N}$$

Combining torques

Torque is a vector quantity, but generally we will only be concerned about its ‘sense’: whether it tends to produce clockwise or anticlockwise motion.

When more than one torque acts on a body the resultant (net) torque can be found by simple addition, but clockwise and anticlockwise torques will oppose each. For example, when an object is acted upon by a 12 Nm clockwise torque and a 15 Nm anticlockwise torque, the resultant torque is (15 − 12) = 3 Nm anticlockwise.

Couples

◆ **Couple (forces)** Pair of equal-sized forces that have different lines of action, but which are parallel to each other and act in opposite directions, tending to cause rotation.

A **couple** is the name we give to a pair of equal-sized forces that have different lines of action but which are parallel to each other and act in opposite directions, either side of the axis of rotation.

A couple produces no resultant force on an object, so there is no translational acceleration; the object will remain in the same location. Figure A4.4 shows a typical example, a couple used to turn a steering wheel.

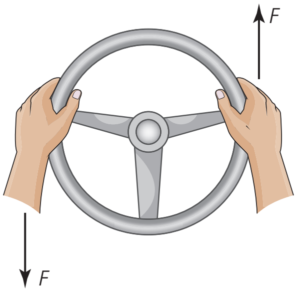


Figure A4.4 A couple used to turn a steering wheel

Other examples of using couples include the forces on a bar magnet placed in a uniform magnetic field, the forces on the handlebar of a bicycle and the forces on a spinning motor.

The magnitude of the torque provided by a couple is simply twice the magnitude of the torque provided by each of the two individual forces, $\tau = 2Fr \sin\theta$. See Figure A4.5.

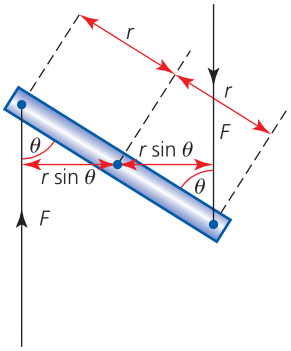


Figure A4.5 Calculating the torque provided by a couple

Rotational equilibrium

SYLLABUS CONTENT

- Bodies in rotational equilibrium have a resultant torque of zero.
- An unbalanced torque applied to an extended, rigid body, will cause rotational acceleration.

If an object remains at rest, or continues to move in exactly the same way, it is described as being in *equilibrium*. *Translational equilibrium* occurs when there is no resultant force acting on an object (Newton’s first law – Topic A.2), so that it remains stationary or continues to move with a constant velocity (that is, in a straight line at a constant speed). A similar definition applies to rotational motion:

Rotational equilibrium occurs when there is no resultant torque acting on an object, so that it remains stationary or continues to rotate with a constant angular speed (defined below).

If an object is in rotational equilibrium, there is no resultant torque, so that:
clockwise torque = anticlockwise torque.

If there is a resultant torque acting on a body, it will produce an angular acceleration. More details to follow.

LINKING QUESTION

- How does a torque lead to simple harmonic motion?

This question links to understandings in Topic C.1.

WORKED EXAMPLE A4.2

Figure A4.6 shows a view of a door from above. A person is trying to push the door open with a force of 74 N in the direction and position shown.

Calculate the minimum force, *F* (magnitude and direction) needed at the handle to stop this happening.

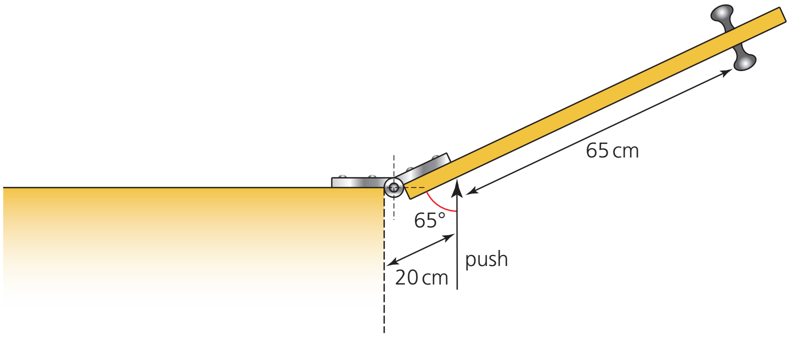


Figure A4.6 A view of a door from above

Answer

The maximum torque obtained with a given force at the door handle will be perpendicular to the door.

For the door to be in rotational equilibrium, the two torques must be equal in magnitude.

$Fr \sin\theta$ clockwise = $Fr \sin\theta$ anticlockwise

$F \times 0.65 \times \sin 90° = 74 \times 0.20 \times \sin 65°$

$0.65F = 13.4$

$F = 21$ N clockwise, perpendicular to the door.

1 A torque of 55 Nm is required to loosen a nut on an engine. Calculate the minimum force with which this can be achieved, if the length of the spanner (wrench) used is 25 cm.

2 Figure A4.7 shows a side-view of one pedal on a bicycle. The distance from the pedal to the axis is 21 cm.

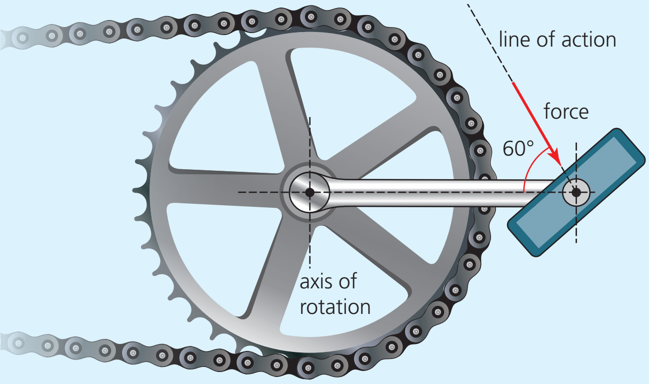


Figure A4.7 Bicycle pedal

a If a cyclist pushes down the line of action, as shown, with a force of 48 N, determine the torque that is being applied.

b Sketch a copy of the diagram but move the pedal to the position where the cyclist can probably apply the greatest torque.

c In which part(s) of each rotation should no torque be applied by the cyclist?

3 **a** Determine the torque provided by the couple shown in Figure A4.5 if the force is 37 N, $r = 7.7$ cm and the angle θ is 49°.

b Sketch a graph to show how the torque would vary if the object moved from horizontal to vertical, as seen (assume that the directions of the forces do not change).

c Does the magnitude of the torque provided by the couple depend on the position of the axis of rotation? Explain your answer.

Angular displacement, velocity and acceleration

SYLLABUS CONTENT

- The rotation of a body can be described in terms of angular displacement, angular velocity, and angular acceleration.

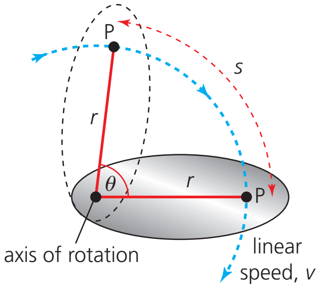


Figure A4.8 Angular displacement, θ, of a point P on a rotating body

◆ Angular displacement The angle through which a rigid body has been rotated from a reference position.

Angular displacement

Any point on a rigid rotating body will be moving along a circular path. See Figure A4.8 for an example.

Angular displacement is defined as the total angle, θ, through which a rigid body has rotated from a fixed reference position. It is measured in radians (or degrees).

In Figure A4.8, the point P is a distance r from the axis of rotation and it has travelled a distance s along the circumference of the circle (*arc length*). So that angular displacement in radians:

$$\theta = \frac{s}{r}$$

Tool 1: Experimental techniques

Understand how to accurately measure quantities to an appropriate level of precision: angle

Describe how you would measure the total angular displacement (in radians) through which the car jack handle seen in Figure A4.9 would need to be rotated in order to raise the side of the car exactly 5.0 cm. Estimate the percentage uncertainty in your measurement.

Figure A4.9 A car jack

Angular velocity

Angular velocity is a vector quantity, but its direction will not be important here, so that angular speed and angular velocity can be considered to be equivalent.

Angular velocity has already been discussed in Topic A.2, where it was considered to be constant during uniform circular motion. More precisely:

Angular velocity, ω, is defined as the change of angular displacement divided by the time taken: $\omega = \frac{\Delta\theta}{\Delta t}$ SI unit: rad s^{-1}, radians (or degrees) per second

Angular velocities are often quoted in rotations per minute (rpm). 1 rpm is equal to $0.10\,\text{rad s}^{-1}$ (to 2 significant figures).

Figure A4.10 A car's tachometer displays rpm/1000

All points on a rigid rotating object will have the same angular velocity, but their linear speeds will be greater if they are further from the axis of rotation. We have seen in that for a body rotating with constant angular speed:

$$\omega = \frac{2\pi}{T}$$

and:

$$v = \frac{2\pi r}{T}$$

so that (as seen in Topic A.2):

linear speed, $v = \omega r$

WORKED EXAMPLE A4.3

In a fairground ride (Figure A4.11) which is moving with a constant linear speed, the passengers complete one rotation in 3.9 s.

Figure A4.11 A fairground ride

a Calculate their angular velocity.

b What is their total angular displacement after one minute?

c Determine the angle between their current position and their starting position.

Answer

a $\omega = \frac{\Delta\theta}{\Delta t} = \frac{360}{3.9} = 92°\,\mathrm{s}^{-1}$ or $\frac{2\pi}{3.9} = 1.6\ \mathrm{rad\ s}^{-1}$

b In 60 s they will have completed 60/3.9 = 15 rotations. (15.38... seen on calculator display)

Total angle moved through = $15.38 \times 2\pi = 97$ rad (5.5×10^{3}°)

c They are 0.38 of a complete rotation from their position at the beginning of the minute.

displacement = $0.38 \times 2\pi = 2.4$ radians (1.4×10^{2}°)

◆ **Angular acceleration, *α*** The rate of change of angular velocity with time, $\Delta\omega/\Delta t$ (SI unit: $\mathrm{rad\,s}^{-2}$). It is related to the linear acceleration, a, of a point on the circumference by $\alpha = a/r$.

Common mistake

Do not confuse angular acceleration with centripetal acceleration.

Angular acceleration

Angular acceleration, *α*, is defined as the rate of change of angular velocity with time:

$$\alpha = \frac{\Delta\omega}{\Delta t} \quad \text{SI unit: rad s}^{-2} \text{ (or degrees per second squared)}$$

There is a simple relationship between angular acceleration and linear acceleration of a point which is a distance r from the axis of rotation.

Since:

$$\Delta\omega = \frac{\Delta v}{r} \Rightarrow \alpha = \frac{\Delta v}{\Delta t r}$$

$$\alpha = \frac{a}{r}$$

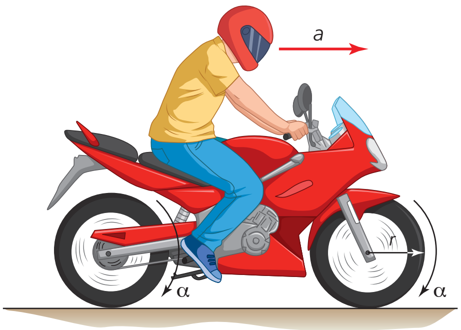


Figure A4.12 Comparing linear and angular acceleration $\alpha = \frac{a}{r}$

For example, Figure A4.12 shows a motor cyclist accelerating linearly at a rate $a = 5.3\ \mathrm{m\,s}^{-2}$. This is also the rate at which the edge of the tyre is accelerating. If the wheel and tyre have an outer radius of 34 cm, the angular acceleration of a wheel, $\alpha = \frac{a}{r} = 5.3 / 0.34 = 16\ \mathrm{rad\,s}^{-2}$.

WORKED EXAMPLE A4.4

A motor spinning at 24 rotations per second accelerates uniformly to 33 rotations per second in 6.7 s.

Calculate its angular acceleration.

Answer

Initial angular velocity = $24 \times 2 \times \pi = 151\,\text{rad}\,\text{s}^{-1}$

Final angular velocity = $33 \times 2 \times \pi = 207\,\text{rad}\,\text{s}^{-1}$

Acceleration = $(207 - 151) / 6.7 = 8.4\,\text{rad}\,\text{s}^{-2}$

Equations of motion for angular acceleration

SYLLABUS CONTENT

- Equations of motion for uniform angular acceleration can be used to predict the body's angular position, θ, angular displacement $\Delta\theta$, angular speed, ω, and angular acceleration.

ATL A4A: Thinking skills

Write 'Linking questions' for the end of Topic A.1, A.2 and A.3 relating to the content of this topic (A.4) so far.

By direct analogy we can write down the equations for uniform angular acceleration. See Table A4.2. ω_i is the initial angular velocity (speed) at the start of time t. ω_f is the final angular velocity at the end of that time.

Table A4.2 Equations of motion

Equations for uniform linear acceleration	**Equations for uniform angular acceleration**
$s = \frac{(u+v)}{2}t$	$\boldsymbol{\Delta\theta = \frac{(\omega_f + \omega_i)}{2}t}$
$v = u + at$	$\boldsymbol{\omega_f = \omega_i + \alpha t}$
$s = ut + \frac{1}{2}at^2$	$\boldsymbol{\Delta\theta = \omega_i t + \frac{1}{2}\alpha t^2}$
$v^2 = u^2 + 2as$	$\boldsymbol{\omega_f^2 = \omega_i^2 + 2\alpha\Delta\theta}$

WORKED EXAMPLE A4.5

An object which is rotating with an angular velocity of $54\,\text{rad}\,\text{s}^{-1}$ accelerates uniformly for 3.2 s and reaches an angular velocity of $97\,\text{rad}\,\text{s}^{-1}$.

a Calculate its angular acceleration.

b What was its angular displacement during the acceleration?

c It then decelerated to rest during an angular displacement of 156 rad. Determine the angular deceleration (negative acceleration).

Answer

a $\omega_f = \omega_i + \alpha t$

$97 = 54 + 3.2\alpha$

$$\alpha = \frac{(97 - 54)}{3.2} = 13\,\text{rad}\,\text{s}^{-2}$$

b $$\Delta\theta = \frac{(\omega_f + \omega_i)}{2}t = \frac{(97 + 54) \times 3.2}{2} =$$

$2.4 \times 10^3\,\text{rad}$

c $\omega_f^2 = \omega_i^2 + 2\alpha\Delta\theta$

$0^2 = 97^2 + (2 \times \alpha \times 156)$

$$\alpha = -\frac{9409}{312} = -30\,\text{rad}\,\text{s}^{-2}$$

4 A carriage on the London Eye (Figure A4.13) can rotate continuously at a speed of 26 cm s^{-1}. The wheel has a radius of 60 m.

a Calculate its angular velocity.

b Calculate how many minutes it takes the wheel to complete one revolution.

■ **Figure A4.13** The London Eye

5 A very large wind turbine (similar to those seen in Figure A4.14) has blades of length 80 m and has a maximum rotational speed of 15 rpm.
Determine the linear speed and the angular velocity of:

a the end of the blade

b a point 10 m from the axis of rotation

c Suggest why engineers limit the speed of rotation of the blades.

■ **Figure A4.14** Wind turbines

6 The outer rim of a bicycle wheel of radius 32 cm has a linear acceleration of 0.46 m s^{-2}.

a Calculate the angular acceleration of the wheel.

b If it starts from rest, determine the time needed for the wheel to accelerate to a rate of three rotations every second.

7 A wheel accelerates uniformly from rest at 5.2 rad s^{-2}.

a What is its angular velocity at the end of 5.0 s?

b Calculate its total angular displacement in this time.

c How many rotations does it complete in 5.0 s?

d After 5.0 s the accelerating torque is removed and the wheel decelerates at a constant rate to become stationary again after 18.2 s.
Calculate how many rotations are completed during this time.

8 A blade of a rotating fan has an angular velocity of 7.4 rad s^{-1}. It is then made to accelerate for 1.8 s, during which time it passes through a total angle of 26.1 rad. Calculate the angular acceleration of the fan blade.

9 A machine spinning at 3000 rpm is accelerated to 6000 rpm while the machine made 12 revolutions.

a Convert 3000 rpm to rad s^{-1}.

b Calculate the angular acceleration.

LINKING QUESTION

- How are the laws of conservation and equations of motion in the context of rotational motion analogous to those governing rectilinear motion?

This question links to understandings in Topics A.1, A.2 and A.3.

Graphs of rotational motion

Graphs for rotational motion can be interpreted in the same way as graphs for linear motion. See Figure A4.15.

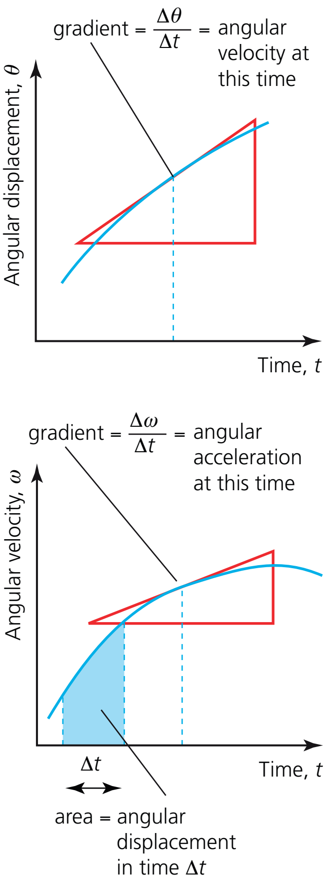


Figure A4.15 Interpreting graphs of rotational motion

10 Sketch an angular displacement–time graph for the following rotational motion: an object rotates at $3\ \mathrm{rad\,s^{-1}}$ for $4\ \mathrm{s}$, it then very rapidly decelerates and then remains stationary for a further $6\ \mathrm{s}$. The rotation is then reversed so that it accelerates uniformly back to its original position after a total time of $15\ \mathrm{s}$.

11 Figure A4.16 shows how the angular velocity of an object changed during $6.0\ \mathrm{s}$.

a Determine the angular acceleration during the first $2.0\ \mathrm{s}$.

b Through what total angle did the object rotate in $6.0\ \mathrm{s}$?

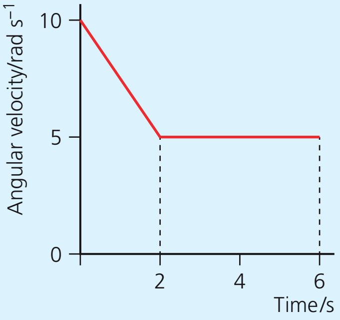


Figure A4.16 Change in angular velocity

Moment of inertia

SYLLABUS CONTENT

- The moment of inertia, I, depends on the distribution of mass of an extended body about an axis of rotation.
- The moment of inertia for a system of point masses: $I = \Sigma mr^2$.

◆ Moment of inertia, I
The resistance to a change of rotational motion of an object, which depends on the distribution of mass around the chosen axis of rotation. The moment of inertia of a point mass is given by $I = mr^2$ (SI unit: kg m^2). The moment of inertia of any real, extended mass can be determined by the addition of the individual moments of inertia of its particles. This is represented by $I = \Sigma mr^2$.

In Topic A.2 we saw that when a resultant *force* is applied to an object, the result is a linear *acceleration*, the magnitude of which depends on the *mass* of the object. Resistance to a change of motion (acceleration) is called *inertia*.

However, in the case of rotational motion, we also need to consider how the mass is distributed around the axis. Consider Figure A4.17.

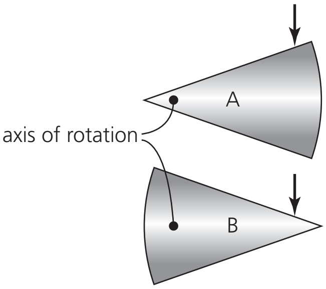


Figure A4.17

Object A will require more force to produce a certain acceleration than object B, which has the same mass and shape but a different axis of rotation. We say that A has a larger **moment of inertia** than B.

Resistance to a change of rotational motion of an object is quantified by its moment of inertia, I, which depends on the distribution of mass around the chosen axis of rotation.

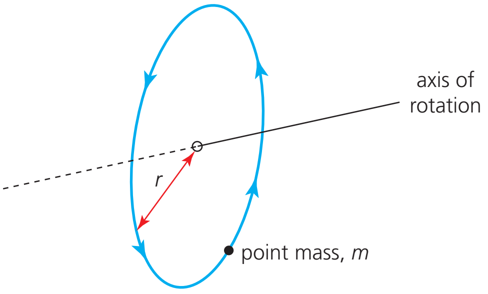
The simplest object to consider is a *point mass*, as seen in Figure A4.18.

The moment of inertia of a point mass, m, rotating at a distance r from its axis is given by:

$$I = mr^2$$

The SI unit of moment of inertia is kg m^2. Most spherical objects can be considered to behave like masses concentrated at their centre points. That is, their centre of mass is at the centre of the sphere.

HL ONLY



■ **Figure A4.18** Rotation of a point mass

WORKED EXAMPLE A4.6

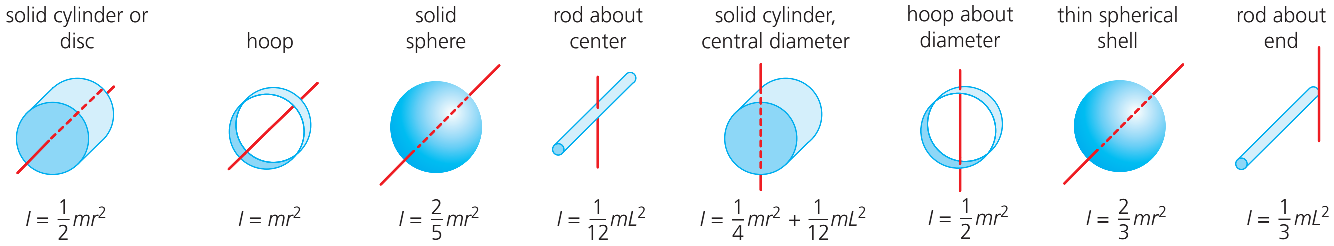
Determine the moment of inertia of a 24 g simple pendulum of length 90 cm.

Answer

$I = mr^2 = 0.024 \times 0.90^2 = 1.9 \times 10^{-2}$ kg m^2 (Assuming that the pendulum can be considered to act as a point mass, and the effect of the string or connecting rod is negligible.)

In principle, the moment of inertia of any real, extended mass can be determined by the addition of the individual moments of inertia of its point masses: $\boldsymbol{I = \Sigma mr^2}$ (The symbol Σ means 'sum of'.)

In practice, the moments of inertia of most simple-shaped objects about specific axes are well known. Some examples are shown in Figure A4.19, but there is no need to remember them, or to know how they were derived, because equations will be provided in the examination paper if needed.



■ **Figure A4.19** Examples of moments of inertia (*r* represents radius and *L* represents length)

Tool 3: Mathematics

Calculate and interpret percentage change and percentage difference

Example: using Figure A4.19, determine the percentage difference between the moments of inertia (about a central axis) of a solid sphere of mass 1.0 kg and a thin spherical shell of mass 100 g. Assume they have the same radius of 12 cm. Use 3 significant figures for all answers.

Moment of inertia of solid sphere = $2/5 \times 1.0 \times 0.12^2 = 5.76 \times 10^{-3}$ kg m^2

Moment of inertia of shell = $2/3 \times 0.10 \times 0.12^2 = 9.60 \times 10^{-4}$ kg m^2

The difference between these two = 4.80×10^{-3} kg m^2

The mean of these two = 3.36×10^{-3} kg m^2

Percentage difference = 100 × (actual difference) / mean = $100 \times (4.80 \times 10^{-3}) / (3.36 \times 10^{-3}) = 143\%$

Perhaps more often, we are concerned with percentage changes. For example, consider a torque which increased from 12 Nm to 18 Nm: the change is 6 Nm (18 − 12)

Percentage change = 100 × (change / original value) = 100 × 6/12 = +50%

Alternatively, if the torque changed from 18 Nm to 12 Nm,

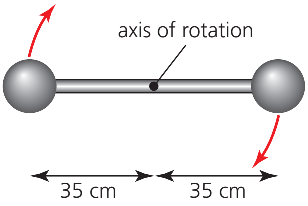
Percentage change = 100 × (change / original value) = 100 × −6/18 = −33%

WORKED EXAMPLE A4.7

Figure A4.20 shows a 'dumb-bell' arrangement in which two spherical masses, each of mass $m_1 = 2.0\,\text{kg}$, are rotating about an axis that is a distance $r_1 = 35\,\text{cm}$ from both of their centres.

a If the rod has a mass of $m_2 = 400\,\text{g}$, length $L = 56\,\text{cm}$ and radius $r_2 = 1.2\,\text{cm}$, determine the overall moment of inertia of this arrangement.

b Calculate the percentage that the rod contributes to the overall moment of inertia of the system.



■ Figure A4.20 'Dumb-bell' arrangement

Answer

a $I = (2 \times m_1 r^2) + \left(\frac{1}{12} \times m_2 L^2\right) = (2 \times 2.0 \times 0.35^2) + \left(\frac{1}{12} \times 0.400 \times 0.56^2\right)$

$= 0.490 + (1.045 \times 10^{-2}) = 0.50\,\text{kg}\,\text{m}^2$ (0.50045... seen on calculator display)

b $100 \times \left(\frac{0.01045}{0.50045}\right) = 2.1\%$

◆ **Flywheel** Dense, cylindrical mass with a high moment of inertia – added to the axes of rotating machinery to resist changes of motion and/or to store rotational kinetic energy.

ATL A4B: Research skills

Comparing, contracting and validating information

Flywheels

Flywheels are added to the axes of rotating machinery to resist changes of motion and/or to store rotational kinetic energy. They need to have large moments of inertia and are used in modern machinery, but Figure A4.21 shows an old-fashioned example, a potter's wheel. The large wheel at the bottom is kicked for a while until it is spinning quickly. After that, because it has a large moment of inertia, there will be no need to keep kicking the wheel continuously to maintain its motion.

■ **Figure A4.21** A flywheel on a potter's wheel

Flywheels can be useful for maintaining rotations in machines that do not have continuous power supplies. To do this they will usually need to be able to store relatively large amounts of kinetic energy. If the potter's flywheel had a mass of 20 kg and radius 50 cm, it would store about 200 J of rotational kinetic energy if it was spinning with a frequency of 2 Hz. (This can be confirmed by using $E_k = \frac{1}{2} I\omega^2$, which is discussed later in this topic.)

A modern flywheel can be seen in Figure A4.22.

In Topic A.3 we outlined the technology of regenerative braking.

1 Use a search engine to find out how flywheels are used:
   - a in vehicles which employ regenerative braking
   - b in wind turbines for generating electricity.

2 Compare and contrast the use of the flywheel in each application. What is similar; what is different?

■ **Figure A4.22** Flywheel on a two-wheeled tractor

Inquiry 1: Exploring and designing

Exploring

Formulating a research question

A student is planning to investigate the behaviour of a flywheel, using the apparatus shown in Figure A4.23. Suggest a possible research question for this investigation.

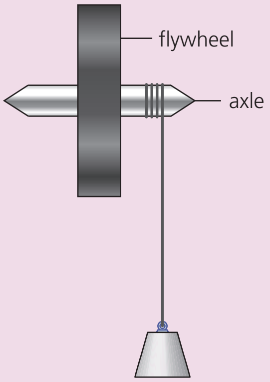


Figure A4.23 Flywheel investigation

12 Estimate the moment of inertia of the Earth in its orbit around the Sun (mass of Earth ≈ 6×10^{24} kg, distance to Sun ≈ 150 million km).

13 a Calculate the moment of inertia of the rotating dumb–bell arrangement seen in Figure A4.24. Assume that the connecting rod has no significant effect.

b By what factor would the moment of inertia change if 20 cm was increased to 30 cm?

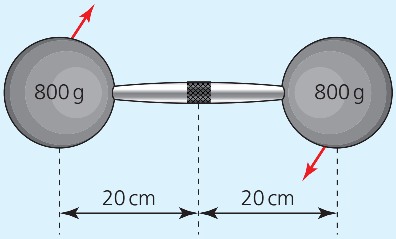


Figure A4.24 A dumb-bell

14 a Suggest why the equation $I = mr^2$ could be used to determine an approximate value for the moment of inertia of a bicycle wheel.

b Estimate a value for the moment of inertia of a typical bicycle wheel.

15 The flywheel shown in Figure A4.23 may be considered to be a steel cylinder (density 7800 kg m^{-3}) of outer radius 40 cm and thickness 12 cm. Use this simplification to estimate its moment of inertia. Ignore other features.

16 Figure A4.25 shows a thick-walled tube and its axis of rotation. Determine its moment of inertia about axis *Y* if its length, *L*, is 20.0 cm..

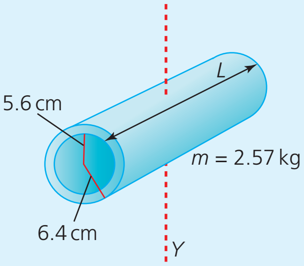


Figure A4.25 Thick-walled tube and its axis of rotation

17 Figure A4.26 shows a boy of mass 25 kg on a playground merry-go-round of mass 370 kg. Estimate a value for the moment of inertia of the system.

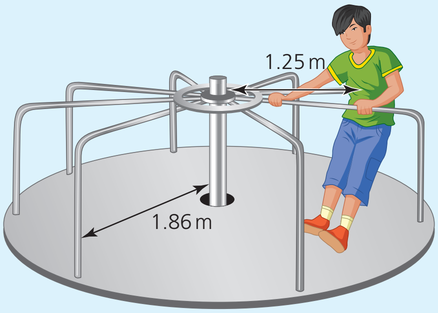


Figure A4.26 Boy on a playground merry-go-round

Newton’s second law for rotational motion

◆ **Newton’s second law for angular motion** $\tau = I\alpha$

◆ **Inverse proportionality** Two quantities are inversely proportional if, when one increases by a factor x, the other decreases by the same factor. For example: $x \propto \frac{1}{y}$ (xy = constant).

SYLLABUS CONTENT

- Newton’s second law for rotation, as given by: $\tau = I\alpha$, where τ is the average torque.

For linear motion, a resultant force, F, acting on an object of mass, m, produces an acceleration, a, as given by: $F = ma$.

For rotational motion, a resultant torque, τ, acting on an object which has a moment of inertia I, produces an acceleration, α.

torque, $\boldsymbol{\tau = I\alpha}$

Tool 3: Mathematics

Understand direct and inverse proportionality

The simplest possible relationship between two variables (like α and τ for a constant moment of inertia) is that they are directly proportional to each other (often just called proportional). This means that if one variable, x, doubles, then the other variable, y, also doubles; if y is divided by five, then x is divided by five; if x is multiplied by 17, then y is multiplied by 17, and so on. In other words, the ratio of the two variables ($\frac{x}{y}$ or $\frac{y}{x}$) is constant. Proportionality is shown using the following symbol:

Proportionality:

$$y \propto x \text{ and } \frac{x}{y} = \text{constant}$$

To check if two variables are proportional to each other, we can either

1. calculate their ratios for different values to confirm that they are constant (see Table A4.3), or
2. draw an x–y graph to determine if it is a straight line through the origin (as discussed later in this chapter).

■ **Table A4.3** The data in either of the last two columns confirms $y \propto x$ (allowing for experimental uncertainties)

x	y	$\frac{x}{y}$	$\frac{y}{x}$
0	0	-	-
0.32	1.6	0.20	5.0
0.81	4.2	0.19	5.2
1.4	6.9	0.20	4.9
2.5	12.8	0.20	5.1
6.4	30.0	0.21	4.7
10.9	55.2	0.20	5.1

If one variable increases while the other decreases (like α and I for a constant torque), we describe it as an *inverse relationship*. The simplest inverse relationship is when one variable, x, doubles, while the other variable, y, halves. if y is divided by five, then x is multiplied by five; if x is multiplied by 17, then y is divided by 17, and so on. In other words multiplying the two variables together always produces the same result: xy = constant. This is called **inverse proportionality**.

Inverse proportionality:

$$y \propto \frac{1}{x} \text{ and } xy = \text{constant}$$

To check if two variables are inversely proportional to each other, we can either

1. calculate values for when they are multiplied together, to confirm that they are constant (see Table A4.4), or
2. draw an $x = \frac{1}{y}$ graph to determine if it is a straight line through the origin (as discussed later in this chapter).

■ **Table A4.4** The data in the last column confirms $y \propto \frac{1}{x}$ (allowing for experimental uncertainties)

x	y	xy
0	0	-
2.0	17	34
11	3.1	34
22	1.6	35
37	0.94	35
43	0.79	34
64	0.55	35

This topic provides another example. Do a quick mathematical check to see if the following data represents an inversely proportional relationship and sketch a graph to show this relationship.

Moment of inertia / 10^{-2} kg m^{-2}	Angular acceleration / rad s^{-2}
0.058	5.3
0.051	5.8
0.039	7.8
0.029	9.9
0.021	14.6

WORKED EXAMPLE A4.8

a Calculate the acceleration produced when a system that has a moment of inertia of 1.23 kg m^2 is acted on by a resultant torque of 0.83 Nm.

b If the system was already rotating at 2.7 rad s^{-1}, determine its maximum angular velocity if the torque is applied for exactly 4 s.

c State the assumption that you made when answering these questions.

Answer

a $\tau = I\alpha$

$$\alpha = \frac{\tau}{I} = \frac{0.83}{1.23} = 0.67\,\text{rad}\,\text{s}^{-2}$$

b $\omega_f = \omega_i + \alpha t = 2.7 + (0.67 \times 4.0) = 5.4\,\text{rad}\,\text{s}^{-1}$

c There are no frictional forces acting on the system.

WORKED EXAMPLE A4.9

Figure A4.27 shows a falling mass, m, attached to a string which is wrapped around a cylinder of radius r and moment of inertia $I = \frac{1}{2}Mr^2$. Derive an equation for the downward linear acceleration, a, of the mass.

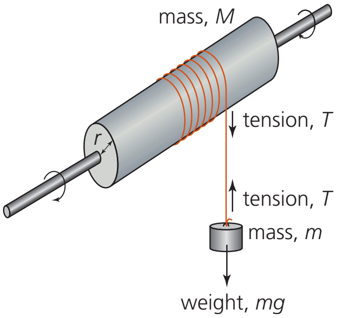


Figure A4.27 A falling mass attached to a string wrapped around a cylinder

Answer

Since $\alpha = \frac{a}{r}$, torque acting on cylinder, $\tau = Tr = I\alpha = \frac{1}{2}Mr^2 \times \frac{a}{r}$

So that, $T = \frac{1}{2}Ma$

Resultant downwards force acting on falling mass, $F = mg - T$

Linear acceleration of falling mass, $a = \frac{F}{m} = \frac{(mg - T)}{m} = \frac{\left(mg - \frac{1}{2}Ma\right)}{m}$

Rearranging gives: $a = \frac{mg}{\left(m + \frac{1}{2}M\right)}$

Common mistake

Note that, because the mass is accelerating downwards, the tension in the string is not equal to the weight of the mass on the end of the string.

Tool 3: Mathematics

Propagate uncertainties in processed data

The term **processed data** is used to describe the results obtained after calculations have been made using **raw data**.

In this section we will consider how uncertainties in raw data affect the results of processed data.

Processed data should not have more significant figures than the raw data used to calculate it.

Consider a simple example: a trolley moving with constant speed was measured to have travelled a distance of 76 cm ± 2cm (±2.6%) in a time of 4.3 s ± 0.2 s (±4.7%).

The speed can be calculated from distance / time = 76 / 4.3 = 17.674…, which is 18 m s^{-1} when rounded to 2 significant figures, consistent with the experimental data.

To determine the uncertainty in this answer we consider the uncertainties in distance and time. Using the largest distance and shortest time, the largest possible answer for speed is 78 / 4.1= 19.024…. m s^{-1}. Using the smallest distance and the longest time, the smallest possible answer for speed is 74 / 4.5 = 16.444…. m s^{-1}. (The numbers will be rounded at the end of the calculations.)

The speed is therefore between 16.444 cm s^{-1} and 19.024 cm s^{-1}. The value 19.024 has the greater difference (1.350) from 17.674. So, the final result can be expressed as 17.674 ± 1.350 cm s^{-1}, which is a maximum uncertainty of 7.6%. Rounding to 2 significant figures, the more realistic result is 18 ± 1 cm s^{-1}.

Uncertainty calculations like these can be very time consuming and, for this course, approximate methods are acceptable. For example, in the calculation for speed shown above, the uncertainty in the data was ±2.6% for distance and ±4.7% for time. The percentage uncertainty in the final result is approximated by adding the percentage uncertainties in the data: 2.6 + 4.7 = 7.3%. This gives approximately the same value as calculated using the largest and smallest possible values for speed. Rules for finding uncertainties in calculated results are given below.

- For quantities that are added or subtracted: add the absolute uncertainties:

$$\text{if } y = a \pm b\text{, then } \Delta y = \Delta a + \Delta b$$

- For quantities that are multiplied or divided: add the individual fractional or percentage uncertainties:

$$\text{if } y = abc\text{, then } \frac{\Delta y}{y} = \frac{\Delta a}{a} + \frac{\Delta b}{b} + \frac{\Delta c}{c}$$

- For quantities that are raised to a power, n:

$$\text{if } y = a^n\text{, then } \frac{\Delta y}{y} = \left|n\left(\frac{\Delta a}{a}\right)\right|$$

- For other functions (such as trigonometric functions or logarithms) calculate the highest and lowest absolute values possible and compare with the mean value, as shown in Worked example A4.10. But note that, although such calculations can occur in connection with laboratory work, they will not be required in examinations.

◆ **Processed data** Data produced by calculations made from raw experimental data.

◆ **Raw data** Measurements made during an investigation.

WORKED EXAMPLE A4.10

An angle, θ, was measured to be 34° ± 1°. Determine the uncertainty in the tangent of this angle.

Answer

$\tan 34° = 0.6745$, $\tan 33° = 0.6494$, $\tan 35° = 0.7002$

Larger absolute uncertainty = 0.7002 – 0.6745 = 0.0257

(0.6745 − 0.6494 = 0.0251, which is smaller than 0.0257)

So, $\tan\theta = 0.67 \pm 0.03$ (using the same number of significant figures as in the original data).

Nature of science: Falsification

Uncertainties

Most people believe that science deals with 'facts'. That is a reasonable comment – but it also gives an incomplete impression of the nature of science. The statement is misleading if it suggests that scientists believe they are always discovering 'truths' that will last forever. In reality, scientific knowledge is open to change, if and when we make new discoveries. More than that, it is the essential nature of science and good scientists to encourage the re-examination of existing 'knowledge' and 'truths' and to look for improvements and progress.

'All scientific knowledge is uncertain…' Richard P. Feynman (1998), *The Meaning of It All: Thoughts of a Citizen-Scientist*.

'One aim of the physical sciences has been to give an exact picture of the material world. One achievement of physics in the twentieth century has been to prove that this aim is unattainable.' Jacob Bronowski

18 A resultant torque of 2.4 Nm ± 0.2 N, accelerated a large metal hoop (with axis of rotation through a diameter) of radius 42 cm ± 1 cm from rest to 5.7 rad s^{-1} ± 0.1 rad s^{-1} in 3.2 ± 0.2 s. Determine a value for the mass of the hoop and the absolute uncertainty in your answer.

19 Calculate the torque needed to accelerate a rotating object which has a moment of inertia 3.2 kg m from 1.3 rad s^{-1} to 4.9 rad s^{-1} in 8.8 s.

20 An object was accelerated from 300 rpm (revolutions per minute) to 1100 rpm in 2.3 s when a resultant torque of 112 Nm was applied. Determine its moment of inertia.

21 Two parallel forces each of 26 N are separated by a distance of 8.7 cm. If this couple provides the resultant torque to a rotating system that has a moment of inertia of 17.3 kg m^2, determine the angular acceleration produced.

22 A torque of 14.0 Nm is applied to a stationary wheel, but resistive forces provide an opposing torque of 6.1 Nm. If the wheel has a moment of inertia of 1.2 kg m^2, show that the total angular displacement after 2.0 s is about two rotations.

23 Consider Figure A4.27. What mass, *m*, will produce a linear acceleration of 2.5 m s^{-2} when acting on an 8.3 kg cylinder?

Conservation of angular momentum

SYLLABUS CONTENT

- An extended body rotating with an angular speed has an angular momentum, L, as given by: $L = I\omega$.
- Angular momentum remains constant unless the body is acted upon by a resultant torque.

◆ **Angular momentum, *L***
Moment of inertia multiplied by angular velocity: $L = I\omega$ (SI unit: kg $m^2 s^{-1}$).

Angular momentum, L, of a rotating object is the rotational equivalent of linear momentum ($p = mv$). It depends on the moment of inertia, I, of the object and its angular velocity (speed), ω.

angular momentum, $\boldsymbol{L = I\omega}$ SI unit: kg $m^2 s^{-1}$

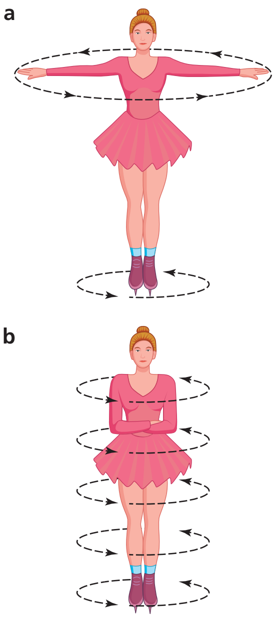


Figure A4.28 Ice-skater

◆ **Conservation of angular momentum** The total resultant angular momentum of a system is constant provided that no resultant external torque is acting on it.

The law of conservation of linear momentum (Topic A.2), which has no exceptions, was shown to be very useful when predicting the outcome of interactions between masses exerting forces on each other. In a similar way, the law of **conservation of angular momentum** (as follows) has no exceptions and can be used to predict changes to rotating systems.

> The total angular momentum of a system is constant provided that no resultant (net) external torque is acting on it.

Figure A4.28 shows a spinning ice-skater in two positions, **a** and **b**. In moving from position **a** to position **b**, the skater lowers her arms and brings them closer to her body, and so reduces her moment of inertia. Assuming there are no external torques acting, her rotational momentum will be constant so that her angular velocity must increase. Similar rotational behaviour can be seen in the motions of gymnasts, divers and ballet dancers.

WORKED EXAMPLE A4.11

A sphere of mass 2.1 kg and radius 38 cm is spinning around a diameter at a rate of 44 rpm. Calculate its angular momentum.

Answer

$$L = I\omega = \frac{2}{5}mr^2 \times \frac{2\pi}{T}$$

$$= \frac{2}{5} \times 2.1 \times 0.38^2 \times 2 \times \frac{\pi}{(60/44)}$$

$$= 0.56\,\text{kg}\,\text{m}^2\,\text{s}^{-1}$$

WORKED EXAMPLE A4.12

A solid metal disc of mass 960 g and radius 8.8 cm is rotating horizontally at 4.7 rad s^{-1}.

a Calculate the moment of inertia of the disc.

b Calculate the new angular velocity after a mass of 500 g is dropped quickly and carefully on to the disc at a distance of 6.0 cm from the centre.

Answer

a $I = \frac{1}{2}mr^2 = \frac{1}{2} \times 0.96 \times \left(8.8 \times 10^{-2}\right)^2 = 3.7 \times 10^{-3}\,\text{kg}\,\text{m}^2$

b moment of inertia of added mass $= mr^2 = 0.5 \times (6.0 \times 10^{-2})^2 = 1.8 \times 10^{-3}\,\text{kg}\,\text{m}^2$

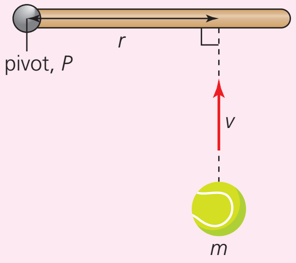
$L = I\omega = \text{constant}$

$(3.7 \times 10^{-3}) \times 4.7 = [(3.7 \times 10^{-3}) + (1.8 \times 10^{-3})] \times \omega$

$\omega = 3.2\,\text{rad}\,\text{s}^{-1}$

Common mistake

Students often believe that an object travelling in a straight line must have zero angular momentum, but consider Figure A4.29. A ball of mass m and speed v is just about to strike a stationary rod perpendicularly at a distance r from where it is pivoted.



■ **Figure A4.29** Ball striking a pivoted rod (seen from above)

The rod will be made to rotate anticlockwise, gaining angular momentum. From the law of conservation of angular momentum in the rod and ball 'system', the ball must have lost angular momentum. The ball initially had angular momentum, L, about point P, not around any point within the ball itself. Instantaneous value:

$$L = I\omega = (mr^2) \times \left(\frac{v}{r}\right) = mvr$$

LINKING QUESTION

- How does conservation of angular momentum lead to the determination of the Bohr radius?

This question links to understandings in Topic E.1 for HL students.

24 Calculate the angular momentum of a 1.34 kg disc of radius 56 cm spinning at an angular speed of 37 rad s^{-1} around an axis passing perpendicularly through its centre.

25 The three blades of a rotating fan each have a moment of inertia of 0.042 kg m^2. If they have a combined angular momentum of 0.74 kg m^2 s^{-1}, determine how many times the fan rotates every minute.

26 An unpowered merry-go-round of radius 4.0 m and moment of inertia 1200 kg m^2 is rotating with a constant angular velocity of 0.56 rad s^{-1}. A child of mass 36 kg is standing close to the merry-go-round and decides to jump onto its edge.

a Predict the new angular velocity of the merry-go-round. State any assumptions you made.

b Discuss whether the merry-go-round would return to its original speed if the child jumped off again.

27 Neutron stars are the very dense collapsed remnants of much larger spinning stars. Suggest why they have extremely high rotational velocities.

LINKING QUESTION

- How does rotation apply to the motion of charged particles or satellites in orbit?

This question links to understandings in Topics D.1 and D.3.

Angular impulse

SYLLABUS CONTENT

▶ The action of a resultant torque constitutes an angular impulse, ΔL, as given by: $\Delta L = \tau t = \Delta(I\omega)$.

In Topic A.2, we noted that the effect of an (average resultant) force is greater if it acts for a longer time. So, it was convenient to introduce the term linear impulse, $J = F\Delta t$. Using Newton's second law, a change of linear momentum, $\Delta p = \Delta(mv)$ occurs because of a linear impulse $F\Delta t$.

Similarly, for rotational motion: an average resultant torque, τ, acting for a time Δt, produces a change of angular momentum, ΔL, called angular impulse (no symbol).

A change of angular momentum, $\boldsymbol{\Delta L = \Delta(I\omega)}$ occurs because of an angular impulse, $\tau\Delta t$:

$\boldsymbol{\Delta L = \tau\Delta t}$ SI unit: kg m^2 s^{-1}

N m s is an equivalent and alternative unit.

Nature of science: Science as a shared endeavour

Communicating inter-connected concepts

It could be argued that we do not need to give [force (or torque) × time] a name: impulse. Have we added to our understanding if we say that impulse = change of momentum, rather than force × time = change of momentum, or is it just easier to say one word instead of three? Would it be possible to understand this topic without ever referring to 'impulse'?

Perhaps the main reason why impulse is important is that forces are rarely constant. Forces usually change to, or from, zero over time and they may also vary depending on many other factors. It is much simpler to refer to the overall effect.

If the applied resultant torque changes, an average value should be used to determine an impulse. For a torque which varies in a regular way, this can be assumed to be midway between the starting and final values. In other examples, we may need to determine an average value from looking at a torque–time graph. We know from Topic A.2, that the area under a force–time graph is equal to the change of linear momentum of the system (impulse). Similarly, in rotational dynamics:

The area under a torque–time graph is equal to the change in angular momentum (angular impulse). This is true for any shape of graph.

WORKED EXAMPLE A4.13

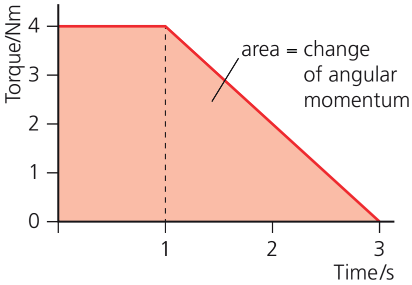


Figure A4.30 Example of a torque–time graph

Consider Figure A4.30.

a Determine the angular impulse represented.

b If this impulse accelerated an object already rotating at 25 rad s^{-1}, calculate the final angular velocity if the object had a moment of inertia of 0.51 kg m^2.

Answer

a Area under graph = angular impulse = change of angular momentum

$$\Delta L = (4.0 \times 1.0) + \left(\frac{1}{2} \times 4.0 \times 2.0\right) =$$

8.0 N m s (or kg m^2 s^{-1}).

b $\Delta L = \Delta(I\omega) = I(\omega_f - \omega_i) = 8.0$

$\omega_f - 25 = 8.0/0.51$

$\omega_f = 41$ rad s^{-1}

28 An object which has a moment of inertia of 4.8 kg m^2 is initially at rest. A torque is then applied which increases uniformly from zero until the object is rotating with an angular velocity of 79 rad s^{-1} after 23 s.
Show that the maximum torque applied was approximately 30 Nm.

29 A solid sphere of mass 1.47 kg and radius 12 cm is rotating about a diameter with an angular speed of 57 rad s^{-1}. Determine what constant torque will bring it to rest in 10.0 s.

30 Figure A4.31 shows the variation of torque applied to a stationary system that has a moment of inertia of 0.68 kg m^2.

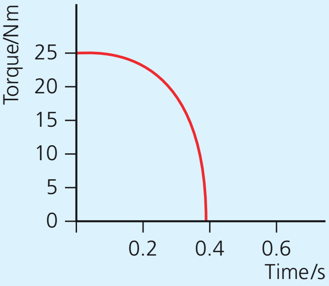


Figure A4.31 Variation of torque applied to a stationary system

a Estimate the change of angular momentum of the system.

b Predict a value for its final angular velocity.

Rotational kinetic energy

SYLLABUS CONTENT

- The kinetic energy of rotational motion, as given by: $E_k = \frac{1}{2}I\omega^2 = \frac{L^2}{2I}$

◆ **Rotational kinetic energy, E_k** Kinetic energy due to rotation, rather than translation. $E_k = \frac{1}{2}I\omega^2$

Knowing, from Topic A.3, that linear kinetic energy:

$$E_k = \frac{1}{2}mv^2 = \frac{p^2}{2m}$$

we can simply write down the equivalent equations for **rotational kinetic energy** by analogy.

Rotational kinetic energy: $\boldsymbol{E_k = \frac{1}{2}I\omega^2 = \frac{L^2}{2I}}$

It is common for an object to have both linear and kinetic energy, the wheel on a bicycle, for example.

WORKED EXAMPLE A4.14

A car of mass 1340 kg is moving with a speed of 12 m s^{-1}.

a Calculate its linear kinetic energy.

b If each wheel and tyre (of four) has a radius of 29 cm and a moment of inertia of 0.59 kg m^2, determine its rotational kinetic energy.

c What is the total kinetic energy of the car?

Answer

a $E_k \text{ (linear)} = \frac{1}{2}mv^2 = \frac{1}{2} \times 1340 \times 12^2 = 9.6 \times 10^4\text{ J}$

b $\omega = \frac{v}{r} = \frac{12}{0.29} = 41.4\text{ rad s}^{-1}$ (41.379... seen on calculator display)

$E_k \text{ (rotational)} = \frac{1}{2}I\omega^2 = \frac{1}{2} \times 0.59 \times 41.379^2 = 5.1 \times 10^2\text{ J}$

c $(9.6 \times 10^4) + (4 \times 5.1 \times 10^2) = 9.8 \times 10^4\text{ J}$

◆ **Slipping (wheel)** Occurs when there is not enough friction between a wheel and the surface to maintain a rolling motion.

◆ **Sliding** Surfaces moving over each other without any rotation involved.

Common mistake

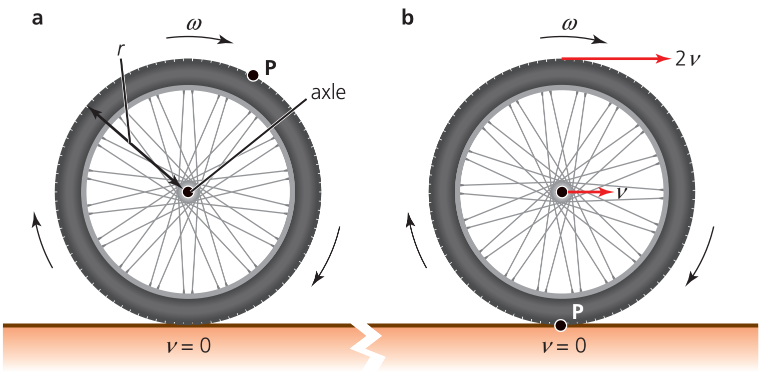
When a vehicle accelerates, many students think that there is just one frictional force, acting against the direction of motion. But *all* forces occur in pairs: because of friction, there is a force backwards on the road *and* an equal and opposite force acting forwards on the vehicle. This force accelerates the vehicle or opposes air resistance if it is moving with constant speed. When there is no forward propulsion, the frictional force on the vehicle from the road will be in the opposite direction.

◆ **Roll** Rotation of an object along a surface in which the lowest point of the object is instantaneously stationary. Requires friction. Compare with slipping.

Rolling (without slipping)

We will assume that there is sufficient surface friction to prevent any **slipping** or **sliding**. This means that there is no relative motion between the lower surface of the rotating object and the surface on which it is moving.

Consider a wheel of radius r rotating with an angular speed ω. All points on its circumference will be moving with linear speed $v = \omega r$. Figure A4.32a shows the wheel of a motor bike (for example), on a road surface, and Figure A4.32b shows the same wheel a short time later. If point P is moving with linear speed v, this must also be the overall speed of the motor bike, as shown on the central axle.



■ **Figure A4.32** Moving wheel showing instantaneous speeds relative to the road surface

The wheel is rotating clockwise due to the action of the engine, and it is pushing backwards (to the left) on the road surface because of friction. Friction with the road surface pushes the car forward. (Newton's third law.)

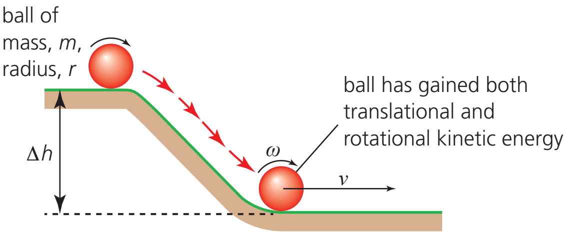
Under normal circumstances, because of friction, there will be no slipping of the wheel on the road surface, which would be dangerous. This means that point P on the wheel in Figure A4.32b must be *momentarily* stationary.

A point on the top of the wheel will be moving with speed $2v$ relative to the road surface.

Because there is no movement, the frictional force does not do any work, so no energy is dissipated at that point. (This is a simplified interpretation.)

Rolling down a slope

An object, such as a ball or a wheel, which can **roll** down a hill will transfer its gravitational potential energy to both translational kinetic energy and rotational kinetic energy.



■ **Figure A4.33** Rolling down a slope

$$mg\Delta h = \frac{1}{2}mv^2 + \frac{1}{2}I\omega^2$$

HL ONLY

At the bottom of a slope, a sliding object will reach a higher speed than a rolling object. Also, rotating objects that have bigger moments of inertia will travel slower at the bottom of the same slope. If the angle of the slope is too steep, rolling will not be possible.

Consider the example of a solid sphere, for which:

$$I = \frac{2}{5}mr^2$$

Remembering that $v = \omega r$, the equation above becomes:

$$\cancel{m}g\Delta h = \frac{1}{2}\cancel{m}\omega^2 r^2 + \left(\frac{1}{2}\right)\frac{2}{5}\cancel{m}r^2\omega^2 = \frac{7}{10}\omega^2 r^2$$

Note that, with the assumptions made, the angular velocity at the bottom of the slope does not depend on the slope angle or the mass of the ball.

WORKED EXAMPLE A4.15

Determine the:

a angular speed

b linear speed of the centre of mass of a 500 g ball,

which has a radius of 10 cm, after it has rolled down a slope of vertical height 1.0 m.

Answer

a

$$g\Delta h = \frac{7}{10}\omega^2 r^2$$

$$9.8 \times 1.0 = 0.7 \times \omega^2 \times 0.1^2$$

$$\omega = 37\,\text{rad s}^{-1}$$

b $v = \omega r = 37 \times 0.1 = 3.7\,\text{m s}^{-1}$

31 Calculate the rotational kinetic energy of a tossed coin if it has a mass of 8.7 g, radius 7.1 mm and completes one rotation in 0.52 s.

32 Calculate the rotational kinetic energy of the Earth spinning on its axis. (Research relevant information.)

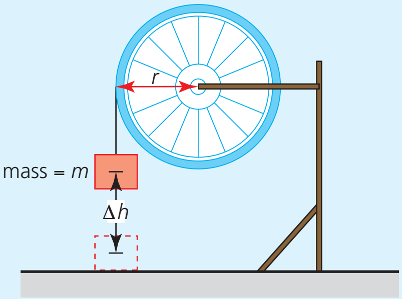
33 A flywheel is spinning with rotational kinetic energy of 4.6×10^4 J. Calculate its moment of inertia if it has an angular momentum of $98\,\text{kg m}^2\,\text{s}^{-1}$.

34 The moment of inertia of the windmill seen in Figure A4.34 is $96\,\text{kg m}^2$. Estimate how many rotations every minute are needed for it to have 250 J of kinetic energy.

Figure A4.34 A windmill

35 Figure A4.35 shows an experimental arrangement that could be used to determine a value for the moment of inertia of a wheel. A string is wrapped around the outside of a wheel and provides a torque as the attached mass accelerates downwards and starts the wheel rotating.

- **a** Write down an equation to represent the transfer of energy when the mass has fallen a distance h.
- **b** Calculate a value for the moment of inertia of the wheel, of radius 24 cm, if a mass of 500 g is moving down with a speed of 1.14 m s^{-1} after falling a distance of 50 cm.



■ **Figure A4.35** Determining the moment of inertia of a wheel

36 The cyclist seen in Figure A4.36 is moving to the right with a constant linear speed of 4.0 m s^{-1}.

- **a** State the linear speed of all points on the circumference of the wheel (with respect to the bicycle).
- **b** State the speed of the lowest point of the wheel with respect to the ground.
- **c** Use the picture to show that the angular speed of the wheel is approximately 15 rad s^{-1}.
- **d** What is the instantaneous velocity of the top of the wheel, with respect to the ground?

■ **Figure A4.36** Cyclist

37 A solid ball and a hollow ball of the same mass and radius roll down a hill. At the bottom, discuss which ball

- **a** will be rotating faster
- **b** has the greater linear speed.

38 a Calculate the greatest

- **i** angular speed
- **ii** linear speed of a solid ball of radius 1.2 cm rolling down a slope from a vertical height of 6.0 cm.

- **b** What assumption did you make?
- **c** Compare your answer to the greatest speed of the same ball dropped the same vertical distance.

39 Use the analogies between linear and rotational mechanics to write down equations for rotational work and power.

Inquiry 1: Exploring and designing

Designing

Identify variables

A student wishes to investigate balls rolling down slopes. Identify all the possible variables involved, and select one independent and one dependent variable that could be investigated. State how the other variable would be controlled.

A.5 Relativity

Guiding questions

- How do observers in different reference frames describe events in terms of space and time?
- How does special relativity change our understanding of motion compared to Galilean relativity?
- How are space–time diagrams used to represent relativistic motion?

◆ **Relativistic motion** Travelling at a significant fraction of the speed of light.

◆ **Reference frame** A coordinate system from which events in space and time are measured.

◆ **Coordinate system** An agreed numerical way of identifying the location and time of an event.

◆ **Event** Single incident that occurs exactly at a precise time and location.

Einstein's theory of special relativity (1905) was a complete revolution in scientific thinking. Before then, Newtonian Mechanics (as explained in Topics A.1 to A.3) had accurately described and predicted motion in the Universe as it was understood at that time. However, the theories of Newton, Galileo and others cannot be applied accurately to objects which are moving very fast (with speeds close to the speed of light called **relativistic motion**). The (unexpected) discovery that the speed of light is always the same for all observers was one reason that Einstein proposed his theory of relativity, one aspect of which is that all motions are observed relative to each other, there is no 'correct point' of view, there is nowhere at absolute rest.

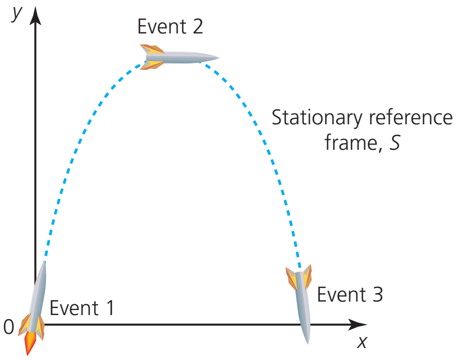
Reference frames

When making calculations on motion (in Topics A.1 to A.3) you will usually have assumed that the objects concerned were moving over the stationary surface of the Earth. The Earth's surface was the *reference frame* and values of displacement, velocity and acceleration were relative to a stationary point on that surface.

A **reference frame** is a **coordinate system** that allows a single value of time and position to be assigned to an event.

Precise **events** are used to simplify our understanding:

An **event** is considered to be a single, instantaneous incident that occurs at a specific time and point in space.



■ **Figure A5.1** Graphical representation of the Earth's reference frame for a rocket in flight

Examples of events could be a flash of light, the moment when two objects collide, or the high point of an object in parabolic flight. Lightning strikes and balloon bursts are commonly used visualizations.

Reference frames are often represented by a set of axes, usually given the label S or S', as shown for two dimensions (only) in Figure A5.1.

To fully define a four-dimensional reference frame, we must specify the origin, the directions of the x-, y- and z-axes, and the event from which the measurement of time, t, is started. The example shown in Figure A5.1 is limited to the x- and y-axes (for simplicity) and it uses the obvious reference frame that is the Earth's surface. However, if we wanted, we could alternatively consider the rocket's reference frame, in which the rocket is stationary and it is the Earth that is seen to move (at the same speed in the opposite direction).

The success of Newtonian mechanics is that it allows the accurate calculation of properties such as displacement, velocity, acceleration and time using the equations of motion (Topic A.1), as Worked example A5.1 illustrates.

WORKED EXAMPLE A5.1

In reference frame, *S*, shown in Figure A5.1, calculate for the three events shown the *x*-, *y*- and *t*-coordinates of an unpowered rocket with an initial vertical velocity of $400\,\text{m}\,\text{s}^{-1}$ and a horizontal velocity of $100\,\text{m}\,\text{s}^{-1}$. Ignore the effects of air resistance.

Answer

Event 1: This is the event that defines the origin and also the start of the timing, so, $x = 0\,\text{m}$, $y = 0\,\text{m}$ and $t = 0\,\text{s}$. The coordinates are (0.0 m, 0.0 m, 0.0 s).

Event 2: This is the event defined by the rocket reaching its maximum height. We can use an equation of motion ($v^2 = u^2 + 2as$) to calculate *x*, *y* and *t*: this gives us the height, $y = 8200\,\text{m}$.

The equation $s = \frac{(u + v)}{2}t$ can be used to show that the time to reach the top of the flight, $t = 41\,\text{s}$.

Since there is no horizontal acceleration, it is straightforward to calculate the horizontal position, *x*, using $s = ut = 4100\,\text{m}$.

Hence the (*x, y, t*) coordinates of Event 2 are (4100 m, 8200 m, 41 s).

Event 3: This event occurs when the rocket is the same height as it was originally. The symmetry of parabolic motion means that it occurs at (8200 m, 0 m, 82 s).

In the rest of Topic A.5, to explain principles without involving extra complications, we will restrict the discussion of reference frames to just the *x*-direction and time.

1 A car is caught by a speed camera travelling at $35.0\,\text{m}\,\text{s}^{-1}$. If the speed camera photograph is taken at point (0.00 m, 0.0 s) determine the coordinates of the car 23.0 s later.

2 A naughty child throws a tomato out of a car at a stationary pedestrian the car has just passed. The car is travelling at $16\,\text{m}\,\text{s}^{-1}$ and the child throws the tomato directly towards the pedestrian so that it leaves the car with a speed of $4\,\text{m}\,\text{s}^{-1}$.
Explain why the tomato will not hit the pedestrian.

Different reference frames

◆ **Observer:** Often a hypothetical person able to use their senses, or instrumentation, to record information about events.

Have you ever walked along a moving train and wondered what your speed was? If you happened to bang your head twice on bags that stuck out too far from the luggage rack as you walked, what were the coordinates of these two events and the distance between them? The answer to this depends on the reference frame from which an **observer** is taking the measurements.

In the study of relativity, an *observer* is a hypothetical person who takes measurements from only one specific reference frame. An observer is always stationary relative to their own reference frame.

In the example of the train, there are three possible reference frames that could be occupied by three different observers:

1 an observer taking measurements sitting on the platform as the train moves past
2 an observer taking measurements sitting on a seat in the train
3 an observer taking measurements walking up the train at the same velocity as you.

According to Newton, each of these three observers will record different values for how fast you are moving and your position when you bang your head. However, they will all agree on the time between the two events occurring and the distance you have moved up the carriage between the two events.

Inertial reference frames

The problem that most students have when trying to understand relativity is that, when they observe movement while they are standing on the ground, they instinctively think that the Earth's surface is not moving, which makes it the 'correct' frame of reference (point of view). We think that the points of view of others – in planes and boats and trains – are temporary and misleading. It needs to be repeatedly stressed that, in physics:

all reference frames are equally valid; there is no 'correct' reference frame.

Figure A5.2 Observer in a rocket

It may help to think about a thought experiment, far away from the Earth: consider an observer in a space vehicle (without windows) in deep space, where any effects of gravity are negligible. See Figure A5.2.

Can the observer determine if the vehicle is 'stationary', moving with constant velocity, or accelerating?

If the observer slowly and carefully releases an object (which has no weight in deep space) in mid-air, the object will appear to stay in exactly the same place if the vehicle and all its contents are moving with constant velocity. If the rocket was considered to be 'stationary' the same observations would be made, but most importantly, there is no such thing as being absolutely 'stationary' – it can never be distinguished from motion at constant velocity.

A resultant force is needed for acceleration. That force originates with the rocket engines and the observer is accelerated by contact forces with the vehicle. An object released in mid-air would not have any resultant force acting on it, so that it will maintain its original motion while the vehicle accelerates around it. To the observer, the object moves 'backwards' compared to them and the accelerating rocket.

This last point is very important: For an observer in the frame of reference of an accelerating vehicle, Newton's laws of motion appear to be broken.

◆ **Inertial reference frame** A frame of reference that is neither accelerating nor experiencing a gravitational field, in which masses obey Newton's laws of motion.

An **inertial reference frame** is one which is not accelerating and in which Newton's laws of motion can be applied.

If there were places which were truly stationary, they would be perfect inertial reference frames, the ideal background for observations of motions. (The adjective 'inertial' suggests lack of movement.) However, a frame of reference moving with constant velocity (zero acceleration) fulfils the same purpose.

The reference frames discussed in the rest of Topic A.5 will all be inertial (non-accelerating) reference frames.

For non-relativistic applications we usually use the Earth's surface as our inertial reference frame (despite its movement).

◆ **Global positioning system (GPS)** A navigation system that provides accurate information on the location of the GPS receiver, by continually communicating with several orbiting satellites.

3 State which of the the following can be thought of as truly inertial reference frames, almost inertial reference frames (objects measured over a small distance appear to be travelling at constant velocity) or clearly not inertial reference frames (unbalanced forces or gravity are clearly present):

a a rocket stationary in deep space so that it is a long way from any gravitational fields.
b a rocket travelling through deep space in a straight line with constant speed
c a **GPS** communication satellite in orbit around the Earth
d a space probe hovering above the surface of the Sun
e a proton travelling close to the speed of light through a straight section of tubing in the CERN particle accelerator in Geneva.

Newton's postulates concerning time and space

◆ **Postulate** See *axiom*.

◆ **Axiom** An unproven assumption that is accepted to be true, which is then used as starting point for further discussion. Similar in meaning to a postulate.

◆ **Simultaneous events** Events that occur at the same time in a specific reference frame, so that in this reference frame they have the same time coordinates. Events that are simultaneous in one frame may not be simultaneous in another frame.

TOK

The natural sciences

- Do the natural sciences rely on any assumptions that are themselves unprovable by science? How is an axiomatic system of knowledge different from, or similar to, other systems of knowledge?

Mathematics and the arts

A **postulate** is a starting point for the development of more advanced reasoning and discussions.

An historical example, from Euclid more than two thousand years ago, concerning geometry: a straight line can be drawn from any one point to any other point. Such fundamental postulates in mathematics are referred to as **axioms**.

Postulates and axioms may not be directly provable, but they are usually simple and unambiguous statements which are agreed by everybody (at that time). They are considered to be necessarily true, in the sense that they are logically necessary.

Before Einstein's theory of relativity, Newton's description of the Universe had made important assumptions. These assumptions are still used by everybody in their everyday lives.

1 The universality of time: All observers agree on the time interval between two events. In particular, they will agree on whether two events are **simultaneous**, or not.
2 The universality of distance: All observers agree on the distance between two simultaneous events.

Simultaneous means that two events are observed to occur at exactly the same time. That is, the time interval between the two events is zero.

To understand these postulates better, imagine a universe with a tiny clock placed in the centre of every cubic metre as shown in Figure A5.3.

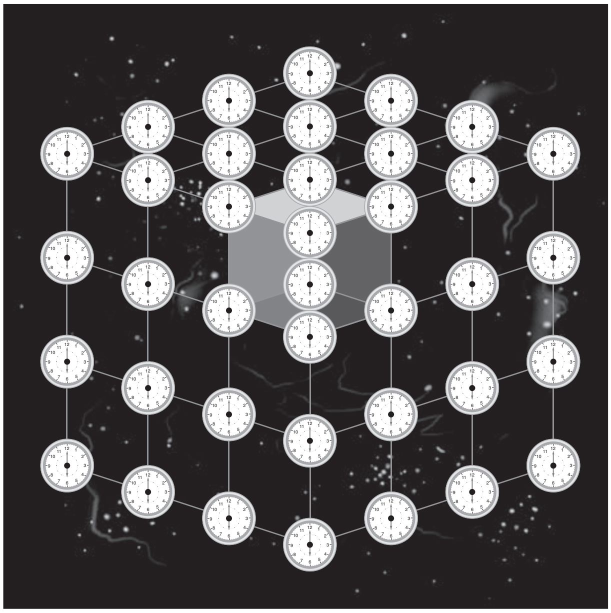


Figure A5.3 A cubic matrix of clocks spreading out regularly throughout space and all reading exactly the same time

The first postulate implies that every clock would always be reading the same time and ticking at exactly the same rate. Any observer moving through the Universe carrying a clock would find that their clock also read the same time as the background clocks and would tick at the same rate. If an observer also carried a metre rule with them as they moved around, they would find that it always exactly matched the shortest distance between any two adjacent clocks.

The central theme of this topic is that Newton's postulates are not totally accurate: they do not apply if relativistic effects are significant.

Galilean relativity

SYLLABUS CONTENT

- Newton's laws of motion are the same in all inertial reference frames and this is known as Galilean relativity.
- In Galilean relativity the position x' and time t' of an event are given by: $x' = x - vt$ and $t' = t$.
- Galilean transformation equations lead to the velocity addition equation as given by: $u' = u - v$.

◆ Galilean relativity How relative motions were described before the discovery of special relativity.

Galilean relativity refers to relative motion as described and explained (using the principles we have already used in Topics A.1 to A.3) by Galileo, Newton and others. That is, relativity as understood before special relativity, which was first introduced by Einstein in 1905 (see below). Galilean relativity can be assumed to be a very good approximation for special relativity for speeds which are low compared to the speed of light.

In Galilean relativity, Newton's laws of motion are the same in all (inertial) frames of reference.

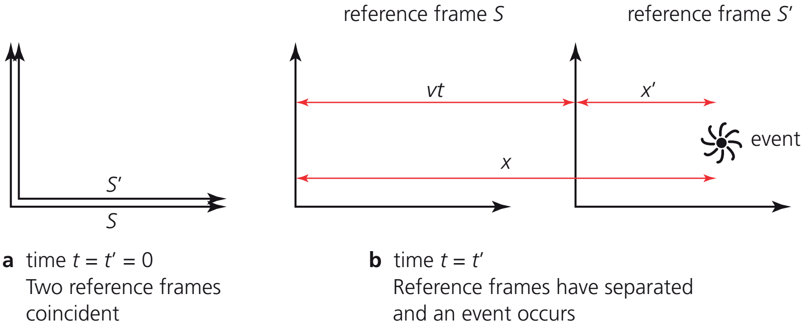
Whenever we move our point of view from one reference frame to another, we need to do what is called a *transformation* by applying standard equations. This becomes very important when we study relativity, so it is worth ensuring that Galileo and Newton's simpler vision of the Universe (as follows) is expressed in a similar way and fully understood.

Observations of position, distance and speed made within our own reference frame are straightforward, but if we make measurements from our reference frame of similar quantities in another reference frame which is moving compared to us, we need to know in what way our measurements are different from measurements made in the other reference frame.

Consider first, the widely used example of an event that occurs on a moving train and how it is seen by observers on the train and on the ground. Consider that both observers start their clocks when the observer sitting on the train travelling with a constant velocity of 20 m s^{-1} passes the observer sitting on the ground. After 5 s, the observer on the ground sees a flash of light (an event) at a point on the train 120 m away. The coordinates of this event in his frame of reference are $x = 120$ m, $t = 5$ s.

The train has moved forward 5 × 20 = 100 m in the 5 s since the timings began, so that, in the frame of reference of the observer on the train, the flash of light occurs a distance of 20 m in front of her. The coordinates of this event in her frame of reference are $x = 20$ m, $t = 5$ s.

More generally, Figure A5.4 compares where an event is observed to occur in two different reference frames (S and S') which were coincident (in the same place) at time $t = t' = 0$, and before they separated.



a time $t = t' = 0$
Two reference frames coincident

b time $t = t'$
Reference frames have separated and an event occurs

Figure A5.4 Comparing an event in two frames of reference

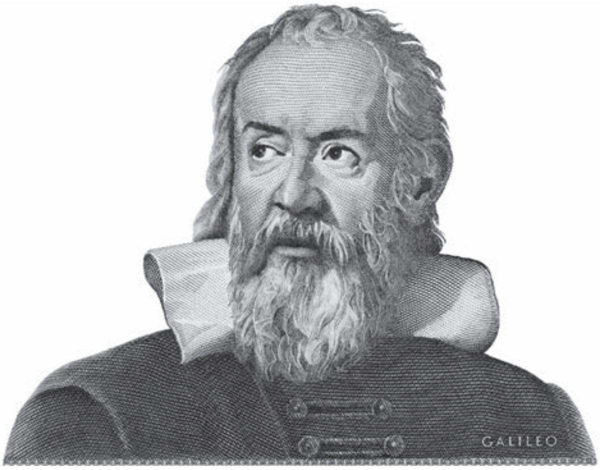


Figure A5.5 Galileo Galilei (1564–1642)

◆ **Galilean transformation** The non-relativistic method of mathematically relating observations between reference frames.

If in reference frame *S* the coordinates of an event are (x, t), in reference frame S', which has relative velocity v compared to S, the coordinates (x', t') of the same event are

$\boldsymbol{x' = x - vt}$ and $\boldsymbol{t' = t}$

These are known as the **Galilean transformation** equations.

Common mistake

It is easy to think that an event primarily occurs in one reference frame and then it is observed in another, but we need to remember that any given event occurs in all reference frames.

Remember that this assumes that the two reference frames are coincident when $t = t' = 0$.

Velocity addition equation

In this topic we will often need to calculate the velocity of an object as observed from different reference frames. This called **velocity addition**. In Galilean relativity this is straightforward. Returning to our previous example of the train: suppose that an object on a train (moving with velocity v) is moving forward with a constant speed and an observer on the ground (reference frame *S*) records this as a velocity u. This will be greater than the velocity of the object, u', recorded on the train (reference frame S'). $u' = u - v$

If in reference frame *S* the velocity of an object is u, in a reference frame S', which has relative velocity of v compared to *S*, the same movement will be recorded as having a velocity, $\boldsymbol{u' = u - v}$

◆ **Velocity addition** Equation which connects the velocities of the same object as observed in two different reference frames.

WORKED EXAMPLE A5.2

A train is moving with a constant velocity of 16.0 m s^{-1}. A ball is rolling along the floor of the train in the direction of travel with a constant velocity of 3.0 m s^{-1}.

a Calculate the velocity of the ball as recorded by an observer on the ground outside.

b Determine how your answer would change if

- **i** the ball was rolling towards the back of the train with the same speed
- **ii** the train was moving in the opposite direction (with the ball moving towards the front of the train). .

Answer

a
$$u' = u - v$$
$$3.0 = u - 16.0$$
$$u = +19\,\text{m s}^{-1}$$

b **i**
$$u' = u - v$$
$$-3.0 = u - 16.0$$
$$u = +13\,\text{m s}^{-1}$$

ii
$$u' = u - v$$
$$-3.0 = u - (-16.0)$$
$$u = -19\,\text{m s}^{-1}$$

If we choose to reverse which reference frames are *S* and S', the same answers will be obtained (as we should expect)

WORKED EXAMPLE A5.3

In deep space, rocket A leaves a space-station with a constant velocity of 300 m s^{-1}. At the same time rocket B travels in the same direction with a constant velocity of 200 m s^{-1}.

a Calculate the distance between rocket A and the space-station after one hour.

b According to an observer in rocket B, what is the distance to rocket A after one hour?

c In rocket B's reference frame, determine how fast an observer would measure the speed of rocket A.

Answer

a $x = ut$ where $t = 1 \times 60 \times 60 = 3600\,\text{s}$

$= 300 \times 3600 = 1.08 \times 10^6\,\text{m}$

b $x' = x - vt$

$= 1.08 \times 10^6\,\text{m} - (200 \times 3600)$

$= 3.6 \times 10^5\,\text{m}$

c $u' = u - v$

$= 300 - 200 = 100\,\text{m s}^{-1}$

Assume that the Newtonian model of the Universe is correct and use Galilean transformations to answer the following questions. *(Note that the answers to some of these questions will contradict the rules of relativity that are introduced later.)* The speed of light, $c = 3.00 \times 10^8\,\text{m s}^{-1}$.

4 In Worked example A5.3 the rockets travel in the same direction. Use the Galilean transformation equations to calculate the answers to Worked example A5.3 questions **b** and **c**, if the rockets travel in opposite directions.

5 A rocket travelling at one-tenth of the speed of light away from Earth shines a laser beam forwards into space.

- **a** Determine how fast an observer inside the rocket measures the light beam photons to be travelling.
- **b** Calculate how fast an observer floating stationary, relative to the Earth, measures the light beam photons to be travelling.

6 Two rockets travelling towards each other are measured by an observer on Earth to each be moving with a speed of 0.6c. Calculate how fast an observer in one rocket thinks that the other rocket is travelling.

7 If you were in an incredibly fast spaceship that was travelling past a space-station at 0.35c and you accelerated a proton inside the ship so that it was travelling forwards through the ship at 0.95c, what speed would an observer in the space-station measure the proton to be travelling?

Limitations of Galilean relativity

The discovery of the constancy of the speed of light (in a vacuum) – see opposite – was evidence that Galilean relativity could not be applied under all circumstances.

Consider the Galilean velocity addition equation applied to light: if a beam of light was sent forward by a passenger on the train (instead of a ball as in Worked example A5.2), they would *correctly* record the speed of light leaving them to be $c = 3.00 \times 10^8\,\text{m s}^{-1}$, but using the Galilean velocity equation *incorrectly* predicts that the observer on the ground would record a higher speed.

◆ **Michelson–Morley experiment** An experiment designed to measure the Earth's speed through the ether. The famous null result was the prime reason for the abandonment of the ether idea, which then contributed to the development of special relativity.

◆ **Ether (or aether)** A hypothetical substance, proposed (falsely) to be the medium through which electromagnetic waves travel.

◆ **Special relativity** Theory connecting space and time developed by Albert Einstein based on two postulates concerning relativistic motion in inertial reference frames. The consequences lead to time dilation, length contraction and the equivalence of mass and energy.

● Nature of science: Hypotheses, and falsification

The Michelson–Morley experiment

In the nineteenth century it was assumed that light needed a medium through which to travel as with other types of waves, such as sound. It was thought that there was a not-understood or detected, 'luminiferous aether', that was present in all space, including vacuum.

Many experiments were designed to discover and investigate this **aether**. The most famous were the experiments carried out in 1887 by Albert Michelson and Edward Morley. The experiments were technically difficult because of the high speed of light, but it was expected that it would be possible to detect small differences in the speed of light beams sent in different directions through the aether (because of the motion of the Earth).

The **Michelson-Morley experiment** failed to demonstrate (verify) the hypothesis that the speed of light would be affected by its direction of travel through the aether. Repeated tests then, and subsequently, have not detected any difference in the speed of light. It has been called 'the most famous failed experiment in history', although perhaps it would be fairer to refer to Michelson and Morley's important result as a 'null finding.'

The results of experiments similar to this confirm that the speed of light is *always* observed to have the same value, regardless of the motions of the source or observer.

Introducing special relativity

Towards the end of the nineteenth century, the classical physics of Newton and Galileo faced two very big problems:

1. The work of James Clerk Maxwell on electromagnetism in the middle of the nineteenth century had combined the phenomena of electricity, magnetism and light. However, Maxwell's (correct) theories of electromagnetism contradicted classical physics in some important respects. James Clerk Maxwell was undoubtedly one of the greatest physicists/mathematicians of all time but his work is *not* included in this course.
2. Experiments were unable to show that light travelled at different speeds depending on its direction of travel with respect to the rotating Earth.

■ **Figure A5.6** Albert Einstein in 1905

Einstein proposed the theory of **special relativity**, connecting space and time, in 1905 in order to resolve these problems. His theory adjusts the Galilean / Newtonian model for speeds close to the speed of light, but classical physics is still valid for slower speeds.

It is called 'special' relativity because it is restricted to inertial frames of reference.

The effects of special relativity only become significant at speeds close to the speed of light, but none of us have any direct experiences of such phenomena in our everyday lives. Throughout the rest of Topic A.5, we will be using examples of events involving such speeds: the motion of imaginary rockets and atomic particles, with imaginary observers travelling with them.

LINKING QUESTION

- Why is the equation for the Doppler effect for light so different from that for sound?

This question links to understandings in Topic C.5.

◆ **Postulates of special relativity** The speed of light in a vacuum is the same for all inertial observers. The laws of physics are the same for all inertial observers.

◆ **Invariant quantity** A quantity that has a value that is the same in all reference frames. In relativity, examples are the speed of light in a vacuum, space–time interval, proper time interval, proper length, rest mass and electrical charge.

The two postulates of special relativity

SYLLABUS CONTENT

- The two postulates of special relativity.

First postulate: the laws of physics are identical (**invariant**) in all inertial reference frames.

The first postulate does not initially appear to be profound. However, it can be interpreted as:

- Observations in different inertial reference frames all have equal worth; there is no single 'correct' frame of reference.
- The Universe has no unique stationary reference frame.
- No experiment is possible that can show an observer's absolute velocity through the Universe.

Second postulate: the speed of light in a vacuum is a constant, $\boldsymbol{c = 3.00 \times 10^8\,\mathrm{m\,s^{-1}}}$, in all inertial reference frames.

This simple statement has enormous implications and needs to be carefully considered (as explained in the rest of this topic). It does not appear to make any sense as judged by our experiences from everyday life. However, many experiments have confirmed it to be true.

It implies that if a rocket in deep space passes a space-station at a tenth of the speed of light, and fires a laser beam forwards as it does so, then both the observer in the rocket and on the space-station must measure the speed of light to be $3.00 \times 10^8\,\mathrm{m\,s^{-1}}$, even though they are moving relative to each other. For this to be the case, space and time must behave in profoundly different ways to how we have learnt to expect.

LINKING QUESTION

- Special relativity places a limit on the speed of light. What other limits exist in physics? (NOS)

TOK

Knowledge and the knower

- How do our expectations and assumptions have an impact on how we perceive things? Is the truth what the majority of people accept?

The constancy of the speed of light (and its implications) conflicts with our expectations and experiences but is undoubtedly true. This counterintuitive knowledge cannot be denied and is a cornerstone of modern physics. Scientists are faced with similar issues when dealing with quantum physics.

Implications of the two postulates

Since the constant speed of light equals distance travelled / time taken, and we know from Galilean relativity that observers in different frames of reference can measure different distances between the same events, then the observers must also measure different times for the travel of a light beam. This means that observers moving relative to each other will disagree about the measurement of time.

Time cannot be an invariant quantity in a relativistic universe.

Our earlier model of the matrix of clocks (Figure A5.2) was wrong. Not only do clocks read different times and tick at different rates but for any pair of events different clocks can record different time *intervals*. In other words, the time interval between two events is not the same for different observers taking measurements from different (inertial) reference frames.

◆ **Space–time** The combination of space and time into a single entity that is used to describe the fabric of the Universe. Fundamentally, in relativity, time and space are not independent of each other. They are observed differently depending on the relative motion of an observer.

◆ **Paradigm** The complete set of concepts and practices etc. that characterize a particular area of knowledge at a particular time. When these change significantly, it is described as a **paradigm shift**.

Space and time are linked as **space–time**. Space and time are not independent of each other, as they were assumed to be in Newtonian mechanics.

The rest of this topic will provide more details about 'space–time' and the implications of special relativity, including

- time dilation
- length contraction
- relativistic velocity additions
- invariant space–time intervals
- space–time diagrams
- relativity of simultaneous events.

TOK

The natural sciences

- What role do paradigm shifts play in the progression of scientific knowledge?

When referring to a particular area of knowledge, a **paradigm** is the name given to the collected, relevant and widely accepted theories, understandings and practices (and so on) that characterize that topic at that time. In the late nineteenth century, the paradigm of knowledge about motion had been centred on the work of Galileo and Newton (and others).

In his 1962 work 'The Structure of Scientific Revolutions' the American philosopher of science Thomas Kuhn suggested that a **paradigm shift** occurs when, for whatever reason, an existing paradigm is replaced with a new paradigm. In science, this might occur after a significant new discovery, or after a completely new, and probably widely unexpected, theory is developed.

A totally new way of thinking about existing knowledge requires great imagination and individuality, which many would describe as genius. Einstein's theory of relativity was possibly the greatest paradigm shift in the history of physics.

Human nature is such that it is usually difficult for people, who may have spent many years living with a particular paradigm, to accept something totally new, which contradicts what they had previously believed.

ATL A5A: Thinking skills

Providing a reasoned argument to support a conclusion

Apart from relativity, another famous paradigm shift in physics knowledge was the confirmation that the Earth was not at the centre of the Universe.

Describe another paradigm shift (of any kind, perhaps from your studies in other IB Diploma subjects) and explain why you think it had significant and far-reaching effects.

Lorentz transformations

SYLLABUS CONTENT

▶ The postulates of special relativity lead to the Lorentz transformation equations for the coordinates of an event in two inertial reference frames as given by:

$$x' = \gamma(x - vt)$$

$$t' = \gamma\left(t - \frac{vx}{c^2}\right)$$

where

$$\gamma = \frac{1}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}}$$

◆ **Lorentz transformation** The equations, involving the Lorentz factor, used to calculate the new position and time coordinates, or spatial and temporal intervals, when transferring from one relativistic reference frame to another.

◆ **Lorentz factor, γ** Scaling factor that describes the distortion of non-invariant quantities when moving between different relativistic reference frames: $\gamma = \frac{1}{\left(\sqrt{1-\frac{v^2}{c^2}}\right)}$

In Newton's model of the Universe, we used Galilean transformation equations to move from one reference frame to another, allowing us to change one coordinate into another (from x, y, z and t to x', y', z' and t'). We can do the same in Einstein's relativistic Universe, but we must instead use the **Lorentz transformation** equations, as follows (for the x-direction only). The equations can be used to transform x-coordinates and t-coordinates of a single event, but only if the origins of the two reference frames coincided at $t = 0$ s.

If in reference frame S the coordinates of an event are x and t, in reference frame S', which has relative velocity of v compared to S, the coordinates (x', t') of the same event are given by:

$$x' = \gamma(x - vt)$$

and

$$t' = \gamma\left(t - \frac{vx}{c^2}\right)$$

γ is known as the **Lorentz factor**. It can be calculated using the following equation.

Lorentz factor:

$$\gamma = \frac{1}{\left(\sqrt{1-\frac{v^2}{c^2}}\right)}$$

The factor involves a ratio, so it does not have any units.

For the purposes of your IB Diploma Physics course, you do not need to know the origin of these three equations.

γ is always greater than one. For everyday macroscopic speeds, $v << c$, so that γ has a value very close to one, which shows us that relativistic effects are not significant in our daily lives. For speeds close to the speed of light, γ becomes significantly greater than one, so that relativistic effects dominate. See Figure A5.7.

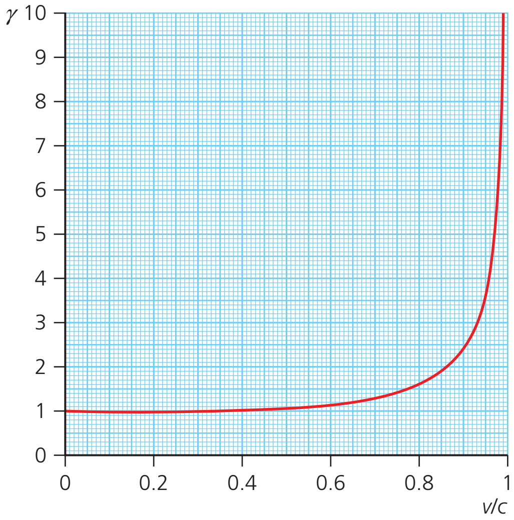


Figure A5.7 Graph showing how the Lorentz factor, γ, varies with speed, v (shown as v/c).

WORKED EXAMPLE A5.4

a Calculate the Lorentz factor for a relative speed of $1.50 \times 10^8\,\mathrm{m\,s^{-1}}$ ($0.50c$).

b An event in reference frame S occurs at $x = 5000\,\mathrm{m}$ and $t = 2.0\,\mathrm{s}$. Calculate when and where the same event occurs as observed from a rocket (reference frame S') which has a relative velocity of $0.50c$.

c State what assumptions you have made to answer part **b**.

Answer

a $\gamma = \dfrac{1}{\left(\sqrt{1-\dfrac{v^2}{c^2}}\right)} = \dfrac{1}{\left(\sqrt{1-\dfrac{(1.50\times10^8)^2}{(3.00\times10^8)^2}}\right)} = 1.15$

b $x' = \gamma(x - vt) = 1.15 \times (5000 - [0.50 \times 3.00 \times 10^8 \times 2.0]) = -3.5 \times 10^8\,\mathrm{m}$

$t' = \gamma\left(t - \dfrac{vx}{c^2}\right) = 1.15 \times \left(2.0 - \dfrac{0.50 \times 3.00 \times 10^8 \times 500}{(3.00 \times 10^8)^2}\right) = 2.3\,\mathrm{s}$

c It was assumed that the two frames of reference were coincident at $t = 0\,\mathrm{s}$.

Tools 3: Mathematics

Use units where appropriate: light year

Astronomical distances are huge. The nearest star to Earth, other than our Sun, (*Alpha Proxima*), is $4.02 \times 10^{16}\,\mathrm{m}$ away. It becomes convenient to use larger units than metres and kilometres in astronomy. The following are non-SI units.

The **light-year (ly)** is the *distance* travelled by light in one year: **1 ly =** $(3.00 \times 10^8) \times 365 \times 24 \times 360$ **=** $\mathbf{9.46 \times 10^{15}\,m}$

(A light-year is defined to be exactly a distance of 9 460 730 472 580 800 m.)

In light-years, the distance to *Alpha Proxima* is 4.25 ly.

Using the light-year as the unit of distance makes many relativity questions easier to answer.

The *parsec* is another widely used unit for distance in astronomy (see Topic E.5).

◆ **Light-year, ly** Unit of distance used by astronomers equal to the distance travelled by light in a vacuum in 1 year.

WORKED EXAMPLE A5.5

According to a rest observer in reference frame S, a rocket reaches a point 20 light-years away after 30 years. This gives (x, t) coordinates for the rocket as (20 ly, 30 y). Another reference frame S' is moving at $0.50c$ relative to S. Determine the coordinates of the rocket according to an observer in S'. (The two reference frames were coincident at $t = 0$.)

Answer

We have already calculated $\gamma = 1.15$ for $v = 0.50c$.

$x' = \gamma(x - vt) = 1.15 \times [20\,\mathrm{ly} - (0.50c \times 30\,\mathrm{y})] = 1.15 \times (20\,\mathrm{ly} - 15\,\mathrm{ly}) = 5.8\,\mathrm{ly}$

$t' = \gamma\left(t - \dfrac{vx}{c^2}\right) = 1.15 \times \left(30\,\mathrm{y} - \dfrac{0.50c \times 20\,\mathrm{ly}}{c^2}\right) = 1.15 \times (30\,\mathrm{y} - 10\,\mathrm{y}) = 23\,\mathrm{y}$

Therefore, according to an observer in reference frame S', the rocket has only travelled 5.8 ly in 23 years, which means that it is travelling at only $0.25c$. This example is straightforward because the unit being used (ly) allows c to be cancelled easily.

WORKED EXAMPLE A5.6

One observer records an event at $x = 250\,\text{m}$ and $t = 1.7 \times 10^{-6}\,\text{s}$. Determine the coordinates of this event as recorded by a second observer travelling at $0.75c$ to the right according to the first observer. Assume the frames of reference were coincident at $t = 0$.

Answer

$$\gamma = \frac{1}{\left(\sqrt{1 - \frac{v^2}{c^2}}\right)} = \frac{1}{\left(\sqrt{1 - 0.75^2}\right)} = 1.51$$

$$x' = \gamma(x - vt)$$

$$= 1.51 \times (250 - [0.75 \times (3.00 \times 10^8) \times 1.7 \times 10^{-6}] = -200\,\text{m}$$

$$t' = \gamma\left(t - \frac{vx}{c^2}\right)$$

$$= 1.51 \times \left(1.7 \times 10^{-6} - \frac{0.75 \times (\,3.00 \times 10^8) \times 250}{\left(3.00 \times 10^8\right)^2}\right) = 1.6 \times 10^{-6}\,\text{s}$$

Equations for transforming distances and time intervals between two events

Often, we are interested in the differences in x-coordinates and t-coordinates between events. That is, distances and time *intervals*. The transformation equations (Δx to $\Delta x'$) and Δt to $\Delta t'$) are similar to those shown above:

Distance between two events:

$$\Delta x' = \gamma(\Delta x - v\Delta t)$$

Time interval between two events:

$$\Delta t' = \gamma\left(\Delta t - \frac{v\Delta x}{c^2}\right)$$

For each of these problems assume one-dimensional motion and assume that, in each case, the observers start timing when the origins of the two reference frames coincide.

8 Imagine a situation where a rocket passes the Earth at $0.5c$. There are two observers – one in the Earth's frame of reference and the other in the rocket's frame of reference.

- **a** Calculate the value of γ.
- **b** A star explosion event occurs at a point 20 light-years from the Earth. The rocket passes the Earth heading towards the star. According to the Earth-based observer the rocket passes the Earth 20 years before the light arrives.
  Determine x' and t' coordinates of the explosion event for the observer in the rocket's reference frame.

9 From Earth, the Milky Way galaxy is measured to be 100 000 light-years in diameter, so the time taken for light to travel from one side of the Milky Way to the other is 100 000 years. Calculate the diameter of the Milky Way for an observer in a distant galaxy moving at a speed of $0.2c$ away from Earth.
Assume that they are travelling in the same plane as the measured diameter of the Milky Way.

◆ Inertial observer An observer who is neither accelerating nor experiencing a gravitational field.

◆ Spatial To do with the dimensions of space. A spatial interval is a length in space.

◆ Temporal To do with time. A temporal interval is an interval of time.

10 According to Earth-based astronomers a star near the centre of the Milky Way exploded 800 years before a star 2000 ly beyond it.
Determine how much later the second explosion is according to a rocket travelling towards the explosions at $0.2c$.

11 In a laboratory an electron is measured to be travelling at $0.9c$. According to an observer in the laboratory at $t = 9.6 \times 10^{-9}$ s it is at a position of $x = 2.6$ m down the length of a vacuum tube.
Calculate the value of the Lorentz factor and use it to work out the time and position of the electron according to an observer in the electron's reference frame.

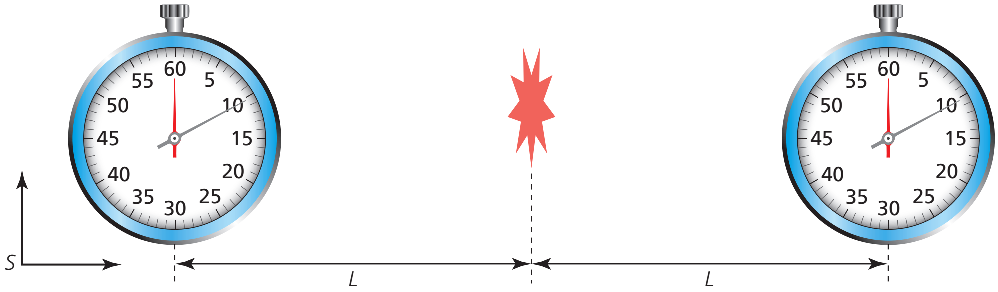
12 Two **inertial observers** are travelling with a relative velocity of $0.8c$ and both see two events occur. According to one observer the events occur 4.2 m apart and with a time interval of 2.4×10^{-8} s between them.
According to the other observer, determine the **spatial** ($\Delta x'$) and **temporal** ($\Delta t'$) intervals between the two events.

Clock synchronization

◆ Synchronized Two clocks are said to be synchronized if according to an observer they are reading the same time.

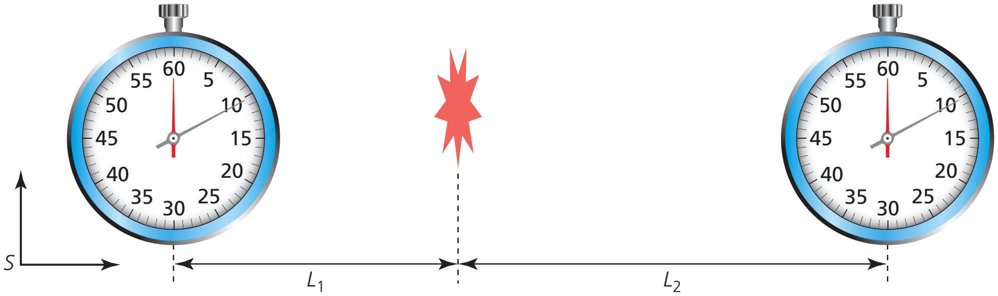
If we wish to compare times of events at different locations in the same reference frame, we need to **synchronize** the clocks. That is, we need to make sure that both / all clocks in the same reference frame always show exactly the same time.

The simplest method is for a flash of light (or a sound) to originate exactly midway between the two clocks and the clocks are both started when the flash of light is detected. The flashes of light are received simultaneously.



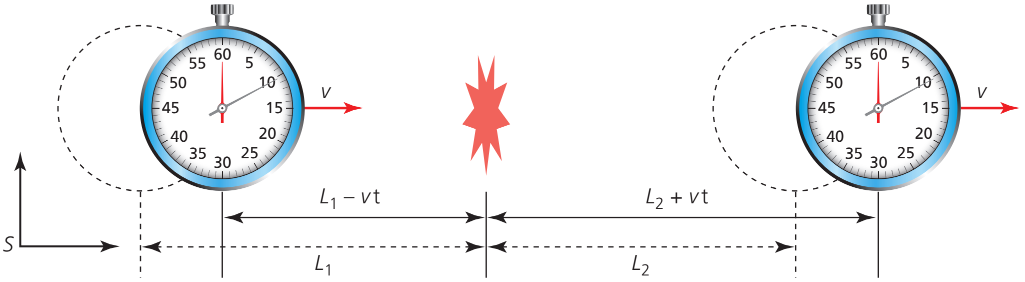
■ **Figure A5.8** Synchronizing two equidistant clocks

The arrangement seen in Figure A5.8 may not be practicable and it is more likely that the clocks are placed at different distances from the flash of light. See Figure A5.9. If the distances are known (L_1, L_2), then the clocks can be synchronized by setting them initially to $\frac{L_1}{c}$ and $\frac{L_2}{c}$, and then starting the clocks from those values when the flash is received.



■ **Figure A5.9** Synchronizing two clocks at different distances

Now consider how an observer in a different reference frame, *S′*, would detect the process seen in Figure A5.8. See Figure A5.10. An observer who sees the clocks moving will see one clock moving towards the flash and the other clock moving away from the flash. Remember that the speed of light, *c*, is the same in all reference frames.



■ **Figure A5.10** Clocks cannot be synchronized by an observer with a velocity relative to them

As seen in reference frame *S′*, the light takes time to travel outwards from the flash and, in this time, one clock is further from the source of the flash than the other clock, so that the light cannot reach the two clocks at the same time in *S′* (simultaneously). In reference frame *S′* the two clocks will not be synchronized, and the clock on the left will read an earlier time than the one on the right.

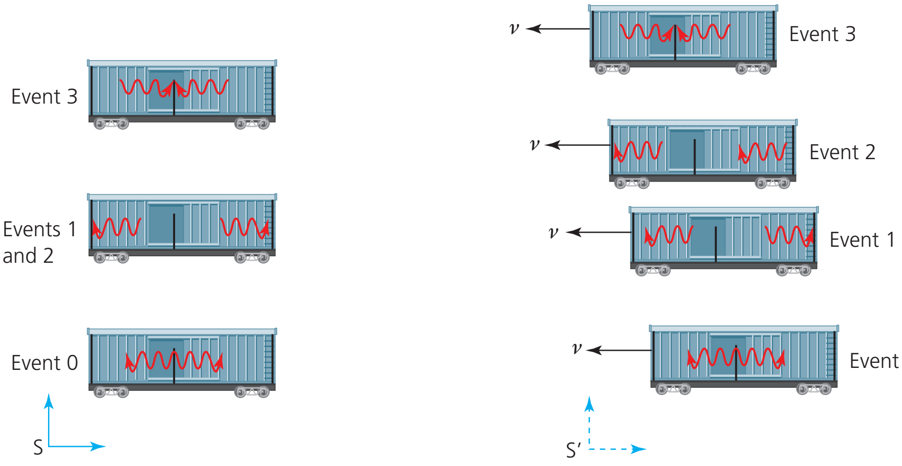
In other words, the observer in reference frame *S* thinks that two events (light arriving at the two clocks) are simultaneous, but the observer in reference frame *S′* records that the event at the left-hand clock occurred first.

Simultaneity

SYLLABUS CONTENT

- The relativity of simultaneity.

The previous section has shown that we need to reconsider our ideas about whether two or more events occur at the same time (are simultaneous) and the order of events. Consider Figure A5.11, which shows a similar situation to Figure A 5.10, but this time the figure shows a flash of light sent from the centre of a train carriage (Event 0) in both directions to mirrors at the end of the carriage. The light beams are then reflected back (Events 1 and 2) to their origin (Event 3).



■ **Figure A5.11** Simultaneous and non-simultaneous events

The reference frame shown on the left-hand side (S) is that of an observer on the train. Both beams of light travel equal distances in equal times. The observer will record that:

- The pulses are sent out simultaneously.
- The pulses reach each end of the carriage simultaneously.
- The pulses return to observer S and are recorded simultaneously.

The reference frame shown on the right-hand side is that of an observer on the ground outside the train (S'). The observer in S' will record that:

- The pulses are sent out simultaneously.
- The pulse that travels down the carriage against the motion of the carriage must arrive at the end of the carriage (Event 1) before the pulse that travels up the carriage (Event 2) because it must have travelled a shorter distance at the same speed.
- However, the observer in S' still sees the pulses return to observer S simultaneously because the effect is reversed for the reflected rays.

If two (or more) events occur at the same place at the same time, then they are simultaneous for all observers in all reference frames.

If two (or more) events occur at different places, then it is possible that they could be simultaneous for one observer in one reference frame, but not be simultaneous for other observers in other reference frames.

We will return to the concept of simultaneity when discussing space–time diagrams.

WORKED EXAMPLE A5.7

A spacecraft travelling at 0.6c passes a space-station and at that moment clocks in both locations are set to zero. A short time later the spacecraft passes a point which is 3000 m away from the origin of the space-station's reference frame. Determine what time a clock on the spacecraft will record for this event.

Answer

$$t' = \gamma\left(t - \frac{vx}{c^2}\right)$$

with

$$\gamma = \frac{1}{\left(\sqrt{1 - \frac{v^2}{c^2}}\right)} = 1.25$$

and time in reference frame of space-station,

$$t' = 1.25 \times \left(1.67 \times 10^{-5} - \frac{0.6 \times 3.00 \times 10^8 \times 3000}{\left(3.00 \times 10^8\right)^2}\right) = 1.34 \times 10^{-5}\,\text{s}$$

13 A *light clock* is a concept sometimes used as a way of comparing observations that are made by observers in two different inertial reference frames. A light clock is a very simple device that reflects a light beam between two parallel mirrors separated by a fixed distance, d. The speed of light in a vacuum is constant for all observers, but the path length taken by the light varies. See Figure A5.12; one physicist (Rachel) is in the rocket, while another (Mateo) is hiding in the cloud.

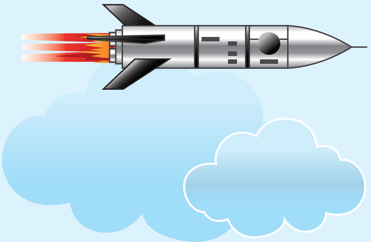


Figure A5.12

One of the diagrams in Figure A5.13 shows the path of the light beam as seen by Rachel, while the other is seen by Mateo, who sees the rocket moving to the right with speed v. Which is which?

14 According to Mateo the time that the light pulse takes to travel from M_1 to M_2 is Δt. Therefore, state how far the rocket moves sideways in this time.

15 Use a Galilean transformation to determine the speed of the light beam according to Mateo.

16 Using Newtonian physics, calculate how far the light beam has to travel when reflecting between M_1 and M_2, according to Mateo.

17 Mateo sees the rocket moving sideways with speed v. In terms of c and Δt, how far has the light beam travelled from M_1 and M_2 according to Mateo?

18 According to Rachel in the rocket, the time taken to travel from M_1 to M_2 is $\Delta t'$. Utilizing Pythagoras's theorem, use your understanding of the postulates of Newtonian physics to derive an expression for Δt in terms of $\Delta t'$, v and c.

19 Explain in terms of the constancy of the speed of light, why the two observers must disagree about the time it takes for the light beam to travel between M_1 and M_2.

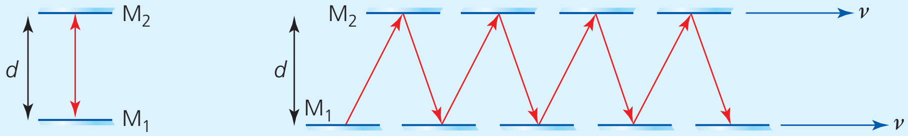


Figure A5.13

Velocity addition transformations

SYLLABUS CONTENT

- Lorentz transformation equations lead to the relativistic velocity addition equation as given by:

$$u' = \frac{u - v}{\left(1 - \frac{uv}{c^2}\right)}$$

Apart from times and distances, we need to know how to make velocity transformations between different frames of reference. Earlier in this topic we used the Galilean velocity transformation: $u' = u - v$, but for situations in which relativistic effects are significant, we need to use the following equation:

$$\boldsymbol{u' = \frac{u - v}{\left(1 - \frac{uv}{c^2}\right)}}$$

u is a velocity of an object in one reference frame (S), u' is the velocity of the same object as seen by an observer who is in a reference frame (S') moving with a velocity v with respect to the first reference frame. Remember that u' must always be less than c.

Note that if u and/or v are small compared to c, the equation reduces to the Galilean form.

There is no need to understand the origin of this equation, although it is linked to the Lorentz transformations.

WORKED EXAMPLE A5.8

Two particles are seen from an external reference frame to be travelling towards each other, each with a velocity of 0.60*c* (Figure A5.14).

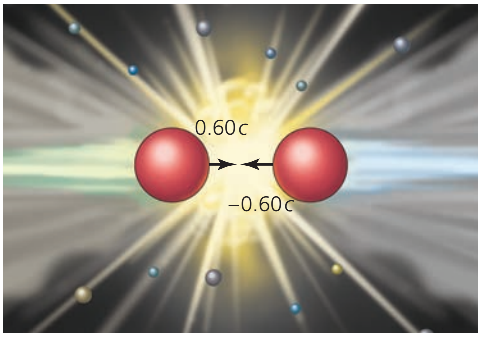


Figure A5.14 Two particles

An observer with one particle measures the velocity of the other particle; determine what speed they record.

Answer

$$u' = \frac{u - v}{\left(1 - \frac{uv}{c^2}\right)}$$

$$= \frac{0.60c - -0.60c}{1 - \frac{(0.60c \times -0.60c)}{c^2}}$$

$$= \frac{1.2c}{1 - \frac{-0.36c^2}{c^2}} = \frac{1.2}{1.36}c = 0.88c$$

The situation is symmetrical. Observers with each particle travelling at 0.60*c* will measure the speed of the other to be 0.88*c*.

It is very easy to miss out the negative signs when doing this calculation. Remember that *u* and *v* are both vectors and so can be positive or negative depending on their direction.

WORKED EXAMPLE A5.9

Two rockets are observed from an external reference frame (*S*) to be travelling in the same direction – the first is measured to be travelling through empty space at 0.75*c*, and a second rocket, which is sent after it, is measured to be travelling at 0.95*c* (Figure A5.15).

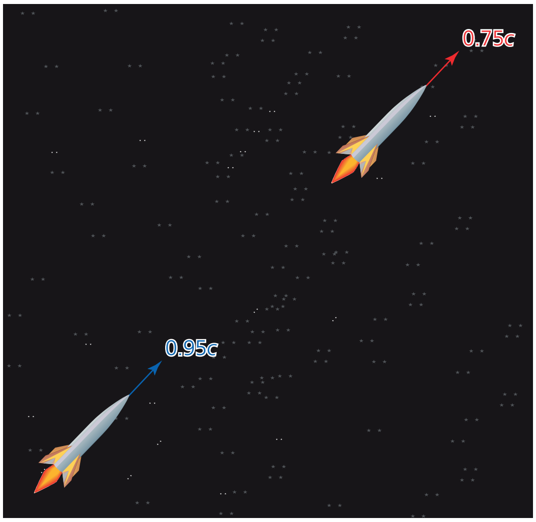


Figure A5.15 Two rockets

Calculate what an inertial observer travelling with the first rocket would measure the approach of the second rocket to be.

Answer

$$u' = \frac{(u - v)}{\left(1 - \frac{uv}{c^2}\right)}$$

$$= \frac{(0.95c - 0.75c)}{1 - \frac{(0.95c \times 0.75c)}{c^2}} = \frac{0.20c}{1 - \left(\frac{0.71c^2}{c^2}\right)}$$

$$= \frac{0.20}{0.29}c = 0.69c$$

Some of these questions refer to photons (light 'particles') and atomic particles, which have not yet been discussed in this course. However, no detailed knowledge is needed to answer the questions.

20 A rocket travelling at one-tenth of the speed of light away from Earth shines a laser beam forwards into space.

a An observer inside the rocket accurately measures the speed of the light beam photons. What value would you expect them to obtain?

b An observer floating stationary, relative to the Earth, also accurately measures the light beam photons. State what value they will obtain.

21 Two rockets are flying towards each other; each are measured by an observer on Earth to be moving with a speed of $0.7c$.
How fast does an observer in one rocket think that the other rocket is travelling?

22 If you were in an incredibly fast spaceship that was travelling past a space-station at $0.35c$ and you accelerated a proton inside the ship so that it was travelling forwards through the ship at $0.95c$, relative to the ship. Determine the speed an observer in the space-station would measure the proton to have.

23 In an alpha-decay experiment the parent nucleus may be considered to be stationary in the laboratory. When it decays, the alpha particle travels in one direction with a velocity of $0.7c$ while the daughter nucleus travels in exactly the opposite direction at $0.2c$.
According to an observer travelling with the daughter nucleus, calculate how fast the alpha particle is travelling.

24 In a beta particle decay experiment an electron and an anti-neutrino that are produced happen to travel away in exactly the same direction. In the laboratory reference frame, the anti-neutrino has a velocity of $0.95c$ and the electron has a velocity of $0.75c$.
What is the anti-neutrino's velocity according to an observer travelling in the electron's reference frame?

25 Protons in CERN's Large Hadron Collider travel in opposite directions around the ring at over $0.999\,000\,0c$.
According to an observer travelling with one group of protons, how fast are the approaching protons travelling?

26 Two light beams are travelling towards each other in exactly opposite directions. According to a laboratory observer their relative velocity is $2c$.
Calculate how fast an observer travelling in the reference frame of one of the light beam's photons measures the speed of the approaching light beam's photons to be.

27 In a space race two spaceships pass a mark and are measured by the race officials at the mark to be travelling in the same direction and travelling at $0.6c$ and $0.7c$ respectively.
According to the faster spaceship, calculate how fast the other ship is moving.

Time dilation

SYLLABUS CONTENT

- Time dilation as given by: $\Delta t = \gamma \Delta t_0$.
- Proper time interval.

We have already seen that observers in relative motion will not agree about measurements of time. We will now consider the measurement of time intervals in more detail.

An observer in reference frame S measures a time interval, Δt, between successive ticks of a clock that is stationary in their reference frame. Another observer in reference frame S' moving with relative velocity, v, will measure a greater time interval for the same ticks, as given by the Lorentz transformation:

$$\Delta t' = \gamma\left(\Delta t - \frac{v\Delta x}{c^2}\right)$$

(As shown previously.)

However, since the ticks of the clock occur at the same place, $\Delta x = 0$, so that the equation can be simplified:

If in reference frame S the time interval between two events that occur at the same place, which is stationary in that reference frame, is Δt, then in reference frame S', which has relative velocity of v compared to S, the time interval between the same two events will be observed to be $\Delta t' = \gamma \Delta t$.

◆ **Time dilation** Relative to an observer who sees the two events occurring in the same place. All other observers measure an increase in the time interval between two events.

◆ **Twin paradox** A paradox that appears to challenge special relativity, based on the impossibility that two twins should each find that they are older than the other. One twin remains on Earth while the other travels at high speed to a distant star and returns.

Since the Lorentz factor, γ, is always greater than one, the time intervals between the ticks of a clock in a reference frame (S') that is moving relative to where the clock is at rest (reference frame S) are greater. This is known as **time dilation**.

WORKED EXAMPLE A5.10

Imagine that a rocket is travelling at $2.0 \times 10^8\,\mathrm{m\,s^{-1}}$ away from Earth.

a The rocket carries a clock which ticks once every second. Calculate the time interval between ticks of the clock as observed on Earth.

b How would your answer change if the clock was on Earth and observed from the rocket?

Answer

a $\Delta t' = \gamma \Delta t$

$$\gamma = \frac{1}{\left(\sqrt{1 - \frac{v^2}{c^2}}\right)} = 1.7$$

$\Delta t' = \gamma \Delta t = 1.7 \times 1.0 = 1.7\,\mathrm{s}$

1.0 s on the rocket corresponds to 1.7 s on Earth.

b The situation is symmetrical: 1.0 s on Earth corresponds to 1.7 s on the rocket.

It is important to realize that time dilation refers to all processes, not just the ticking of a clock. Time proceeds at different rates for frames of reference that are moving relative to each other, but this only becomes relevant at very high speeds. In this course we will *not* be considering what happens to accelerating frames of reference, or frames of reference that are moving closer together.

ATL A5B: Research skills

The '**twin paradox**' is a widely used situation used when discussing time dilation. Research into this thought experiment, explain why it is described as a paradox, and outline how it is resolved.

TOK

Knowledge and the knower

- Does knowledge always require some kind of rational basis?

A paradox is a situation, question or statement that contains an embedded contradiction. For example: Perhaps the most famous example of a simple paradox is 'this statement is false'. What might paradoxes suggest to us about the nature or reason and its value in determining what is true?

Proper time interval

◆ **Proper time interval** The time interval between two events as measured by an observer who records the two events occurring at the same point in space. It is the shortest time interval between events measured by any observer.

It should be clear that the magnitude of a time interval between two events depends on whether the observer is moving with respect to the events being observed. If the observer is in the same frame of reference as the events, this corresponds to the shortest possible time interval, which is called the **proper time**. If the observer is moving with respect to the events, the time interval will be greater.

A proper time interval, Δt_0, is the time interval between two events that take place at the same location as the observer.

We can then re-write the **time-dilation formula**:

time dilation:

$$\Delta t = \gamma \Delta t_0$$

◆ **Time dilation formula** $\Delta t = \gamma \Delta t_0$, where Δt_0 represents the proper time interval as measured by an observer who sees the first and second events occur in the same place.

You need to try to imagine yourself in the reference frame of each observer – are the *x*, *y* and *z* coordinates of the two events the same? If they are, then this observer measures the proper time between the two events.

28 In a laboratory, an electron is accelerated by a potential difference of 100 kV. Its speed is then measured by timing how long it takes to pass between two different points measured in the laboratory as being 5.00 m apart. Is the observer in the electron's reference frame or the observer in the laboratory reference frame recording proper time?

29 A rod measured in its rest frame to be one metre in length is accelerated to 0.33*c*. The rod is then timed as it passes a fixed point. Is the observer at the fixed point or the observer travelling with the rod measuring proper time?

30 The same rod is timed by both observers as it travels between two fixed points in a laboratory. If the observers are recording when the front of the rod passes each fixed point, is either observer measuring the proper time?

31 In a third experiment the two observers start timing when the front of the rod passes the first point and stop timing when the end of the rod passes the second point. Is either observer measuring the proper time?

Length contraction

SYLLABUS CONTENT

- Length contraction as given by:

$$L = \frac{L_0}{\gamma}$$

- Proper length.

Suppose that we observe an object in another reference frame, *S*′, which is moving relative to our reference frame, *S*. The length of the object, a rod for example, $\Delta x'$, can be determined in reference frame *S*′ from a measurement of the position of its ends.

We know from the Lorentz transformation equation that $\Delta x' = \gamma(\Delta x - v\Delta t)$ (as seen above), where Δx is the length of the rod as observed from reference frame *S*. Assuming that the measurement of the positions of both ends of the rod are done at the same time, $\Delta t = 0$, so that the equation simplifies:

If in reference frame *S*′, the length of an object, which is stationary in that reference frame, is recorded to be $\Delta x'$, then in reference frame *S*, which has relative velocity of *v* compared to *S*, the length will be recorded as:

$$\Delta x = \frac{\Delta x'}{\gamma}$$

Since γ is always greater than one, any length in another reference frame, which is moving relative to us, will always be less than the length we would measure if the object was in our own reference frame. This is called **length contraction**.

◆ **Length contraction** The contraction of a measured length of an object relative to the proper length of the object due to the relative motion of an observer.

Top tip!

It may seem strange that relativistic effects *increase* time intervals but *decrease* lengths. For example, from observations of a spacecraft travelling at very high speed away from Earth, time expands but lengths contract (both as seen from Earth). This is because we are considering a clock at rest in the Earth's reference frame, but a length at rest in the spacecraft's reference frame.

WORKED EXAMPLE A5.11

The length of an object measured in a spacecraft travelling at 20% of the speed of light away from Earth is 1.00 m. Determine what length would be detected by an observer on Earth.

Answer

$$\gamma = \frac{1}{\left(\sqrt{1-\frac{v^2}{c^2}}\right)} = 1.02$$

$$\Delta x = \frac{\Delta x'}{\gamma} = \frac{1.00}{1.02} = 0.980\,\text{m}$$

Because the situation is symmetrical, a 1.00 m length on Earth would be recorded as 0.980 m from the spacecraft.

Proper length

The magnitude of a length between two points depends on whether the observer is moving with respect to the observations. If the observer is in the same frame of reference as the object, this will correspond to the longest possible length, which is called the **proper length**, L_0. If the observer is moving with respect to the object, the length will decrease.

The proper length of an object, L_0, is its length when stationary in the same reference frame as the observer.

We can then re-write the **length contraction formula** as:

length contraction: $\boldsymbol{L} = \frac{\boldsymbol{L_0}}{\gamma}$

◆ **Proper length** The proper length of an object is the length measured by an observer who is at rest relative to the length being measured. The proper length is always the longest length measurable by any observer.

◆ **Length contraction formula** $L = L_0/\gamma$, where L represents the length and L_0 represents the proper length.

32 In a laboratory, an electron is accelerated by a potential difference of 100 kV. Its speed is then measured by timing how long it takes to pass between two different points measured in the laboratory as being 5.00 m apart.
Is an observer in the electron's reference frame, or the observer in the laboratory reference frame recording the proper length between the two points?

33 A rod measured in its rest frame to be one metre in length is accelerated to 0.33c. The rod is then timed as it passes a fixed point.
Is the observer at the fixed point or the observer travelling with the rod measuring proper length between the start and finish events?

34 The same rod is timed by both observers as it travels between two fixed points in the laboratory.
If the observers are recording when the front of the rod passes each fixed point, is either observer measuring the proper length for the distance between the start and finish events?

35 In a third experiment the two observers start timing when the front of the rod passes the first point and stop timing when the end of the rod passes the second point.
Is either observer measuring the proper length for the distance between the start and finish events?

36 A rod is measured to have a proper length of exactly 1.00 m.
Calculate how long you would measure it to be if it was to fly past you at 0.80c.

37 The time taken for the rod in question 36 to pass a fixed point in the laboratory is 2.5×10^{-9} s.
Determine what time interval an observer travelling with the rod would measure between the same two events.

38 Fatima flies through space and, according to Fatima, her height is 1.60 m. Fatima flies headfirst past an alien spaceship and the aliens measure her speed to be 0.80*c*.

- **a** How tall will the aliens on their spaceship measure Fatima to be?
- **b** Oliver takes 6.1×10^{-9} s to fly past the same aliens at 0.90*c*. According to the aliens what time interval does it take Oliver to fly past them?

39 In a space race, a spaceship, measured to be 150 m long when stationary, is travelling at relativistic speeds when it crosses the finish line.

- **a** According to the spaceship it takes 7.7×10^{-7} s to cross the finishing line. Calculate how fast it is travelling in terms of *c*.
- **b** Determine what time interval the spaceship takes to cross the finishing line according to an observer at the finishing line.
- **c** According to an observer at the finishing line, how long is the spaceship?
- **d** According to the observer at the finishing line, how fast is the spaceship travelling?

40 In the same race as question 39 a sleek new space cruiser takes only 2.0×10^{-6} s to cross the finish line according to the race officials at the line. They measure the space cruiser to be 450 m long. How long is the space cruiser according to its sales brochure?

The muon-decay experiment: a test of special relativity

SYLLABUS CONTENT

▶ Muon-decay experiments provide experimental evidence for time dilation and length contraction.

The muon particle effectively provides us with a tiny clock that travels at a speed very close to the speed of light.

A **muon** (μ) is an unstable subatomic particle. Muons are produced naturally in the Earth's atmosphere as a result of collisions between atmospheric particles and very high-energy cosmic radiation that is continually bombarding the planet. This occurs at about 10 km above the Earth's surface; the muons produced have an average speed of around 0.995*c* in the Earth's frame of reference.

◆ **Muon decay experiment** An important experiment supporting both time dilation and length contraction. The experiment compares the levels of high-energy muons found in the atmosphere at around 10 km with those found at the Earth's surface, using the muon half-life as a means of measuring time.

◆ **Muon** Unstable elementary subatomic particle.

◆ **Half-life (radioactive)** The time taken for the number of unstable particles to be reduced to half.

◆ **Radioactive decay (radioactivity)** Spontaneous and random change of an unstable particle.

Top tip!

You do not need to understand or remember the nature of muon particles. However, an understanding of this section requires some knowledge of the concept of **half-life**, which is not described fully until Topic E.3. A brief outline follows:

Because they are unstable, muons exhibit random **radioactive decay** into other particles. Mathematically, decay is expressed by the concept of *half-life* which is the time taken for half of any given number of muons to decay. The half-life of muons as observed in the frame of reference of the Earth's surface is 1.56×10^{-6} s. This means that:

- After one half-life (1.56×10^{-6} s), half of the muons are still undecayed and half of the muons have decayed.
- After two half-lives ($2 \times 1.56 \times 10^{-6}$ s), $\frac{1}{2} \times \frac{1}{2} = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$ of the muons are still undecayed.
- After three half-lives ($3 \times 1.56 \times 10^{-6}$ s), $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$ of the muons are still undecayed.
- After *n* half-lives ($n \times 1.56 \times 10^{-6}$ s), $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \ldots \left(= \frac{1}{2}\right)^n$ of the muons are still undecayed.

Muon decay and the predictions of classical physics

The time taken for the muons to travel the 10 km from the upper atmosphere down to the Earth's surface is calculated simply as:

$$t = \frac{x}{v} = \frac{(10 \times 10^3)}{(0.995 \times 3.00 \times 10^8)} = 3.35 \times 10^{-5}\,\text{s}$$

During this time, most of the muons will have decayed, so that only a very small fraction arrives at the Earth's surface. We can determine that fraction by calculating the number of half-lives involved as shown after the next Tools box.

Tool 3: Mathematics

Carry out calculations involving logarithmic and exponential functions

If we wish to determine the value of y in the exponential equation $y = x^a$, where a is not a whole number, we need to take logarithms:

$$\log y = a \times \log x$$

For example, if $y = 3.4^{7.4}$,

$$\log y = 7.4 \times \log 3.4 = 7.4 \times 0.5315 = 3.933$$

$$y = \text{antilog of } 3.933 = 8.6 \times 10^3$$

Alternatively, natural logarithms, ln, could be used.

Number of half-lives elapsed as muons travel down to the Earth's surface, n

$$\frac{\text{total time}}{\text{half-life}} = \frac{(3.35 \times 10^{-5})}{(1.56 \times 10^{-6})} = 21.5$$

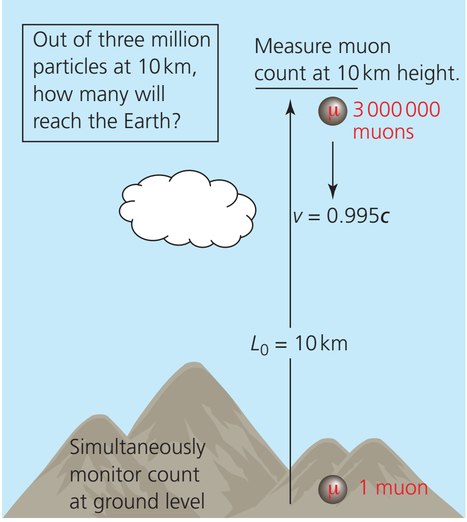
The fraction reaching surface,

$$f = \left(\frac{1}{2}\right)^{21.5}$$

$$\log f = 21.5 \log \left(\frac{1}{2}\right) = -6.47$$

Fraction (f) = 3.37×10^{-7}.

◆ **Classical physics**
Physics theories that predated the paradigm shifts introduced by relativity and quantum physics.



■ **Figure A5.16** Using classical physics to predict muon count on Earth's surface

Classical physics predicts that only about 1 muon reaches the surface for every 3 million that are created in the upper atmosphere.

But, in reality, a far greater proportion of muons are detected than this classical physics calculation predicts. In fact, the fraction of the muons from the upper atmosphere reaching the Earth's surface is approximately 0.2 (two in every 10). In order to understand this, we need to replace classical physics with a relativistic interpretation.

Muon decay using relativity and time dilation

In the Earth reference frame, the proper length is 10.0 km and the speed of the muons is 0.995c. An observer in this frame of reference would also measure the time interval for the muons' travel, Δt, to be 3.35×10^{-5} s. However, an observer in the muons' frame of reference would be measuring the proper time interval (Δt_0).

$$\Delta t = \gamma \Delta t_0$$

with:

$$\gamma = \frac{1}{\left(\sqrt{1 - \frac{v^2}{c^2}}\right)} = \frac{1}{\left(\sqrt{1 - \frac{0.995^2 c^2}{c^2}}\right)} = 10$$

so that:

$$3.35 \times 10^{-5} = 10\Delta t_0$$

$$\Delta t_0 = 3.35 \times 10^{-6}\,\text{s}$$

Alternatively, this can be calculated from the time dilation of the half-life.

$$\text{number of half-lives} = \frac{\text{total time}}{\text{half-life}} = \frac{(3.35 \times 10^{-6})}{(1.56 \times 10^{-6})} = 2.15$$

$$\text{The fraction reaching surface} = \left(\frac{1}{2}\right)^{2.15} = 0.23.$$

This is consistent with actual measurements, thus providing experimental evidence for time dilation.

Muon decay using relativity and length contraction

In the muon's frame of reference, the 10.0 km thickness of the lower atmosphere is contracted.

$$L = \frac{L_0}{\gamma} = \frac{10 \times 10^3}{10} = 1000\,\text{m}$$

Then:

$$t = \frac{x}{v} = \frac{1000}{(0.995 \times 3.00 \times 10^8)} = 3.35 \times 10^{-6}\,\text{s}$$

So, the fraction remaining to reach the Earth's surface is once again 0.23. In this analysis, from the muon's reference frame, the rest observer perceives what we measure to be 10 km of atmosphere to be only 1 km.

Experimental confirmation of this result has provided evidence in support of the concept of length contraction.

41 Some muons are generated in the Earth's atmosphere 8.00 km above the Earth's surface as a result of collisions between atmospheric molecules and cosmic rays. The muons that are created have an average speed of 0.99c.

- **a** Calculate the time it would take the muons to travel the 8.00 km through the Earth's atmosphere to detectors on the ground according to Newtonian physics.
- **b** Calculate the time it would take the muons to travel through the atmosphere according to a relativistic observer travelling with the muons.
- **c** Muons have a very short half-life. Explain how measurements of muon counts at an altitude of 8.0 km and at the Earth's surface can support the theory of special relativity.

42 In a high-energy physics laboratory electrons were accelerated to a speed of 0.950c.

- **a** How long would scientists in the laboratory record for these electrons to travel a distance of 2.00 km at this speed?
- **b** Calculate the time this takes in the reference frame of the electrons.

Space–time

In Newton's Universe, time and space are both invariant – they have fixed intervals that do not vary throughout either space or time. This means that in classical physics we can measure space and time independently. As we have seen, this is not true in a relativistic universe. If time and distance are not fixed, unvarying quantities in a relativistic universe, what quantities can we rely on? The answer is: space–time intervals.

In relativity, space and time are joined to form a single four-dimensional (x, y, z and t) concept called space–time.

◆ **Space–time diagrams** Graphs showing variations of objects' positions with time, adapted to compare different frames of reference.

This is very difficult to visualize, but **space–time diagrams** (discussed later) are very helpful.

TOK

The natural sciences

- What is the role of imagination and intuition in the creation of hypotheses in the natural sciences?

We have seen throughout this topic and elsewhere in Theme A that sometimes physics produces results that seem to contradict our commonsense expectations but must logically be true. What are the different roles of imagination and intuition in making scientific knowledge? How do they relate to reason as a way of knowing?

In essence: things cannot move through space without also moving through time, simply because it takes time to move anywhere. Conversely, we cannot measure time without referring to things moving through space. If everything stayed in exactly the same place and nothing moved, we would have no indication that time was passing.

Space–time was a concept first introduced by Minkowski in 1908. He was Einstein's former mathematics teacher. Einstein initially rejected the idea of space–time but then realized its importance and used it as a major stepping stone in the discovery of general relativity.

◆ **Space–time interval, Δs** A space–time interval combines both the spatial and temporal elements of space–time into a single value.

Space–time interval

SYLLABUS CONTENT

- The space–time interval Δs between two events is an invariant quantity as given by: $(\Delta s)^2 = (c\Delta t)^2 - \Delta x^2$.

In classical physics, a distance, Δs, between two points, the length of a rod for example, in three-dimensional space (x, y, z) can be calculated using Pythagoras's theorem: $(\Delta s)^2 = \Delta x^2 + \Delta y^2 + \Delta z^2$. Although the coordinates of its ends may change when seen in different frames of reference, the length of the rod is always the same (invariant).

Similarly, a 'distance' (Δs) between two points in four-dimensional space–time (x, y, z, t) can be calculated from $(\Delta s)^2 = c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$. (Note: do not worry about the signs used here; different sources of information commonly use different sign conventions.) You are not expected to explain the origin of this equation.

The 'distance' Δs is called a **space–time interval**.

Since we are restricting direction to the x-direction only, space–time interval squared:

$$(\Delta s)^2 = (c\Delta t)^2 - \Delta x^2$$

Different (inertial) observers may measure different time intervals and different distances between events, but they will agree on the space–time interval between events.

We know that an event that is measured to have coordinates (x, t) in the S reference frame can have different coordinates (x', t') in the S' frame of reference. But observers in both reference frames will agree that the quantities $[c^2(\Delta t)^2 - (\Delta x)^2]$ and $[c^2(\Delta t')^2 - (\Delta x')^2]$ are equal.

$$c^2(\Delta t')^2 - (\Delta x')^2 = c^2(\Delta t)^2 - (\Delta x)^2$$

The importance of this will become clearer when we draw space–time diagrams in the next section.

WORKED EXAMPLE A5.12

A single laser pulse is made to trigger two explosion events as it travels through a long vacuum tube. The two events are 99 m apart and the time for the light to travel this distance is 3.3×10^{-7} s. Determine the space–time interval squared between the two events.

Answer

$$(\Delta s)^2 = (c\Delta t)^2 - \Delta x^2 = \left[\left(3.00 \times 10^8\right)^2 \times \left(3.3 \times 10^{-7}\right)^2\right] - 99^2 = 0.0\,\text{m}^2$$

The space–time interval for any two events linked by a photon travelling in a vacuum is always zero. Two events linked by an object travelling slower than c will have a positive space–time interval, while two events that are too far apart for a photon to travel between the two events in the time interval between them have a negative space–time interval.

WORKED EXAMPLE A5.13

Determine the space–time interval squared for an electron that is fired with a speed of $5.93 \times 10^{-7}\,\text{m}\,\text{s}^{-1}$ across a gap of 5.00 m.

Answer

$$\Delta t = \frac{\Delta x}{v} = \frac{5.00}{5.93 \times 10^7}$$

$$= 8.43 \times 10^{-8}\,\text{s}$$

$$(\Delta s)^2 = (c\Delta t)^2 - \Delta x^2$$

$$= \left(3.00 \times 10^8\right)^2 \times \left(8.43 \times 10^{-8}\right)^2 - 5.00^2$$

$$= 6.14 \times 10^2\,\text{m}^2$$

The fact that the space–time interval between any two events is constant for all observers allows us to calculate how long a time an observer travelling in the electron's reference frame will record between the two events. In this reference frame the electron is stationary and the start and finish lines move towards it with the start and finish events occurring at the electron.

This means that this observer is recording proper time and $\Delta x' = 0$:

$$(\Delta s)^2 = (c\Delta t')^2 - (\Delta x')^2$$

$$(\Delta t')^2 = \frac{(\Delta s)^2}{c^2}$$

$$\Delta t' = \sqrt{\frac{6.14 \times 10^2}{\left(3.00 \times 10^8\right)^2}} = 8.26 \times 10^{-8}\,\text{s}$$

We have seen that proper time interval and proper length can be considered as invariant quantities.

- a proper time interval, Δt_0, is the time between two events that take place at the same location as the observer, so that $\Delta x = 0$. The space–time interval can then be written as:

  $$\Delta s^2 = c^2\Delta t_0^{\ 2} - \Delta x^2 = c^2\Delta t_0^{\ 2} - 0^2$$

  $$\Delta s^2 = c^2\Delta t_0^{\ 2}$$

- the proper length of an object, L_0, is its length when stationary in the same reference frame as the observer, so that $\Delta t = 0$. The space–time interval can then be written as:

  $$\Delta s^2 = c^2\Delta t^2 - \Delta x^2 = c^2 0^2 - \Delta x^2 = -\,\Delta x^2$$

  $$\Delta s^2 = -\,\Delta x^2$$

The fact that Δs^2 is negative is to do with how Δs^2 is defined.

Space–time diagrams

SYLLABUS CONTENT

- Space–time diagrams.
- The angle between the worldline of a moving particle and the time axis on a space–time diagram is related to the particle's speed as given by: $\tan\theta = \frac{v}{c}$.

The graphical representation of events seen in Figure A5.17 is called a space–time diagram. There are some similarities (but not many) with distance–time graphs. The axes represent the reference frame (coordinate system) of a specific inertial observer.

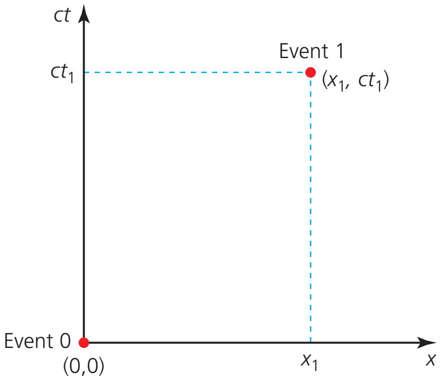


Figure A5.17 Space–time diagram for an inertial reference frame, *S*, showing two events and their coordinates

Space–time diagrams can be a very powerful method of explaining relativistic physics. They contain a lot of information, so we will try to build up the parts before putting a complete diagram together.

Space–time diagrams are normally drawn with the distance, *x*, on the horizontal axis and time, *t*, on the vertical axis. Although the vertical axis could just show time, more commonly it shows the speed of light multiplied by time, *ct*, because, as we will explain, this simplifies interpretation of the diagram with respect to space–time intervals. The units of the vertical axis, *ct,* are m, which is the same as on the horizontal axis. The scales on both axes are the same.

Events are represented as points in space–time. Just like an ordinary graph, the coordinates of the event are read off from the axes. In Figure A5.17 it is clear that, for an inertial observer in reference frame *S*, Event 0 occurs before Event 1. Event 0 occurs at $x = 0$, $ct = 0$. Event 1 occurs at $x = x_1$, $t = t_1$ (ct_1).

Events that occur on the same horizontal line are simultaneous. Events that occur on the same vertical line occur at the same location.

◆ **World line** The path that an object traces on a space–time diagram.

World lines

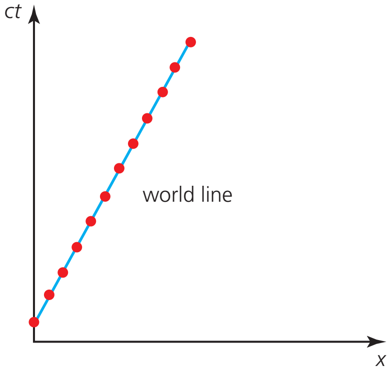


Figure A5.18 Space–time diagram showing how a series of events joined together produces an object's world line

An object travelling through space–time can be imagined as a series of consecutive events. If we join up these events with a line, then we are plotting an object's path through space–time. We call this path the object's **world line**. In Figure A5.18 a straight world line is drawn showing that the object is moving through space with constant velocity relative to the observer. This particular world line does not pass through the origin because the object is observed a short time after the observer started their clock.

Consider Figure A5.19, which shows world lines with different gradients.

Since:

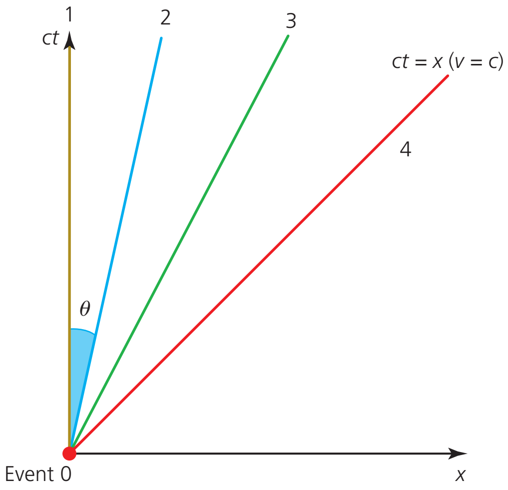
$$v = \frac{x}{t}$$

the gradient of a world line is given by:

$$\frac{ct}{x} = \frac{c}{v}$$

So, the steeper the gradient, the slower the object is travelling. The units are the same on each axis so that the gradient has no units.

An object that is stationary as seen by an observer in the same reference frame will appear as a vertical line because its *x*-coordinate does not change (Line 1). Line 2 represents a moving object. Line 3 represent faster movement.



■ **Figure A5.19** Straight world lines

Line 4 is a central feature of space–time diagrams. It has a gradient of 1. (Assuming equal scales, the line will be at angle of 45° to both axes.)
It represents $v = c$, that is, motion at the speed of light, which can only be light itself.

All inertial observers agree on the value of c, so all observers must agree on the world line for light.

Note that it is not possible for a world line to have an angle θ of greater than 45°, because that would represent an object moving with a speed greater than the speed of light.

Angles between world lines and *ct*-axis

Looking at Figure A5.19, we can see that:

$$\tan\theta = \frac{x}{ct} = \frac{v}{c}$$

The angle between the world line of a moving object and the *ct*-axis on a space–time diagram is related to the object's speed as given by:

$$\boldsymbol{\tan\theta = \frac{v}{c}}, \quad \theta = \tan^{-1}\left(\frac{v}{c}\right)$$

Adding another frame of reference to a space–time diagram

This is what makes space–time diagrams so useful. Representing a second inertial reference frame on the same diagram is straightforward because the background space–time does not change and events remain in the same places, making it possible to compare how different observers perceive the same events.

The world lines seen in Figure A5.19 represent observations made of four objects in a particular frame of reference, S. Now let us consider how to represent the motion of any of those objects, number 2 for example, in its own frame of reference, S', which is moving to the right with a speed v (= $c\tan\theta$).

If we were to draw a separate space–time diagram for S', the world line of object 2 would be a vertical line, showing that it was stationary in its own frame of reference. However, the intention here is to draw and compare two (or more) frames of reference on the same diagram.

When we add the S' frame of reference to the original S frame of reference, its space–time axis (ct'), which represents being stationary in its own frame of reference, must coincide with its world line in S. Its other space–time axis (x') must be placed so that the pair of axes are symmetrical about the $ct = x$ and $ct' = x'$ line. See Figure A5.20.

The axes of the S' frame of reference are tilted and not perpendicular to each other. To determine the coordinates of an event we draw lines parallel to the S' axes, just as we do in the S frame of reference, as seen in Figure A5.17.

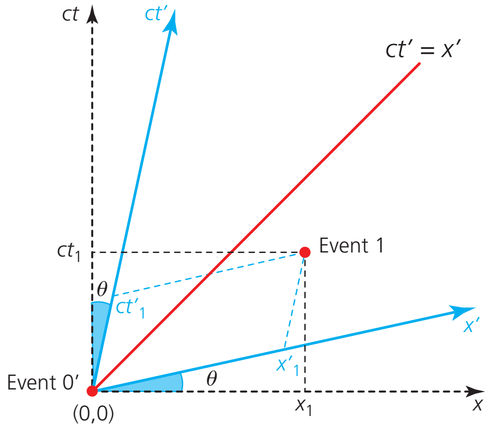
The coordinates of event 1 in S, x_1 and ct_1, transform to x'_1 and ct'_1 in S'.

Space–time diagrams like these give an immediate visualization that $x_1 \neq x'_1$ and $t_1 \neq t'_1$.

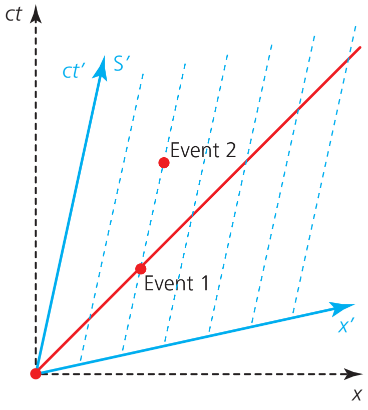
Space–time diagrams can be understood in a similar way to the graphs that you are used to, except that the grid lines are skewed rather than vertical and horizontal.

See Figure A5.21 in which, in reference frame S', Events 1 and 2 occur at the same place, but at different times. The same two events in reference frame S occur at different places and different times.

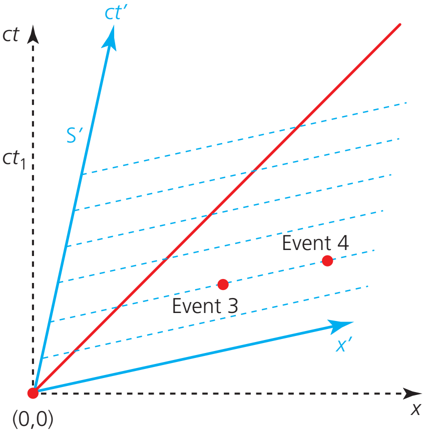
Now consider Figure A5.22, in which Events 3 and 4 are simultaneous for an observer in reference frame *S*′, but at different locations. The same two events in reference frame *S* occur at different locations and different times.



■ **Figure A5.20** Space–time diagram showing the additional axes for reference frame *S*′ in blue



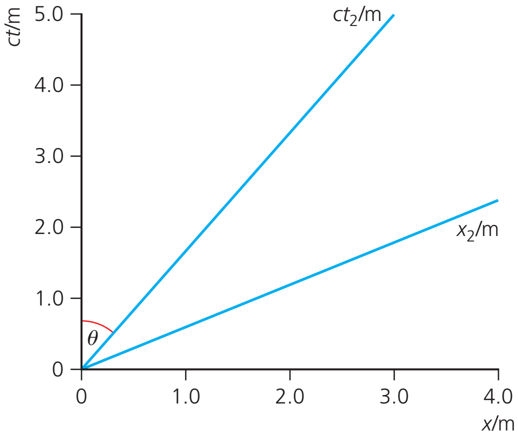
■ **Figure A5.21** Space–time diagram with dashed blue lines representing separate world lines for different points that are each stationary in reference frame *S*′



■ **Figure A5.22** Space–time diagram with dashed blue lines representing separate world lines for different points that occur at the same time in reference frame *S*′

WORKED EXAMPLE A5.14

Use Figure A5.23 to determine the speed of the reference frame shown in blue relative to the other reference frame.



■ **Figure A5.23**

Answer

$\tan\theta = \frac{v}{c}$

$\frac{3.0}{5.0} = \frac{v}{c}$

$v = 0.60c$

WORKED EXAMPLE A5.15

Use the Lorentz transformation equations to show that the x'-line has a gradient of v/c, and hence confirm that the angle between the x- and x'-axes is given by: $\tan\theta^{-1} = \left(\frac{v}{c}\right)$, as shown in Figure A5.24.

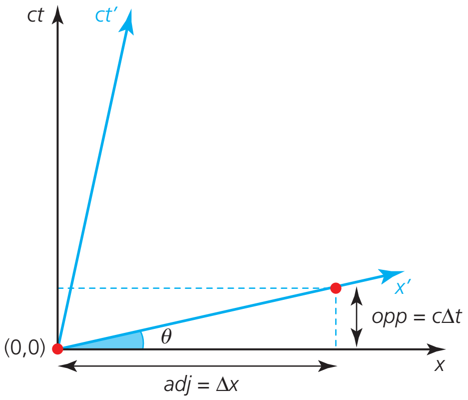


Figure A5.24 Space–time diagram showing the calculation of the angle formula

Answer

The equation for the x'-axis in terms of x and t can be found by setting $t' = 0$ and using the Lorentz transformation for t':

$$t' = \gamma\left(t - \frac{vx}{c^2}\right) = 0$$

So, the bracket = 0

$$t = \frac{vx}{c^2}$$

$$ct = \left(\frac{v}{c}\right)x$$

which is of the form $y = mx$

$$\text{gradient} = \left(\frac{v}{c}\right) = \frac{c\Delta t}{\Delta x} = \frac{opp}{adj}$$

$$\tan\theta = \frac{opp}{adj} = \left(\frac{v}{c}\right)$$

$$\theta = \tan^{-1}\left(\frac{v}{c}\right) \text{ as required}$$

WORKED EXAMPLE A5.16

Remember how an observer could demonstrate that two events were simultaneous (see Figure A5.11). Draw a space–time diagram with reference frame S representing the observer on the train and S' representing the reference frame on the platform.

The light rays are sent out in opposite directions, so we need to draw a positive and a negative x-axis to allow us to position both the events (Figure A5.25).

Observer S sees:

- The pulses are sent out simultaneously (Event 0).
- The pulses reach each end of the carriage simultaneously (Events 1 and 2).
- The pulses return to observer S simultaneously (Event 3).

Observer S' sees:

- The pulses are sent out simultaneously (Event 0).
- The pulse that is fired down the carriage against the motion of the carriage must arrive at the end of the carriage (Event 1) before the pulse that is fired up the carriage arrives at the other end of the carriage (Event 2).
- However, S' still sees the pulses return to observer S simultaneously (Event 3); the geometry of the space–time diagram gives us exactly the same outcome, demonstrating that events with no space–time interval are simultaneous for all observers, but events that occur in two separate places can be simultaneous for some observers but not for others.

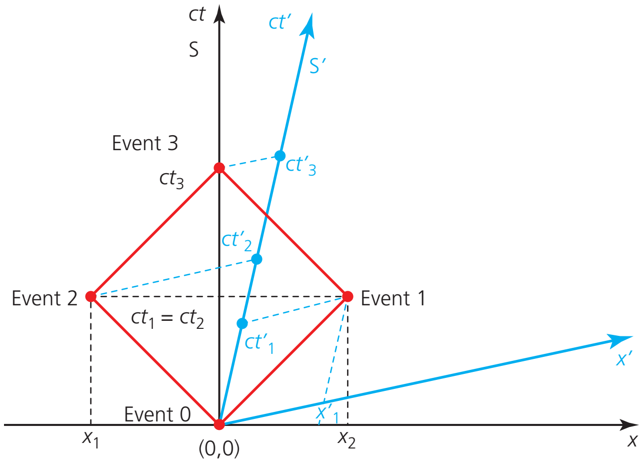


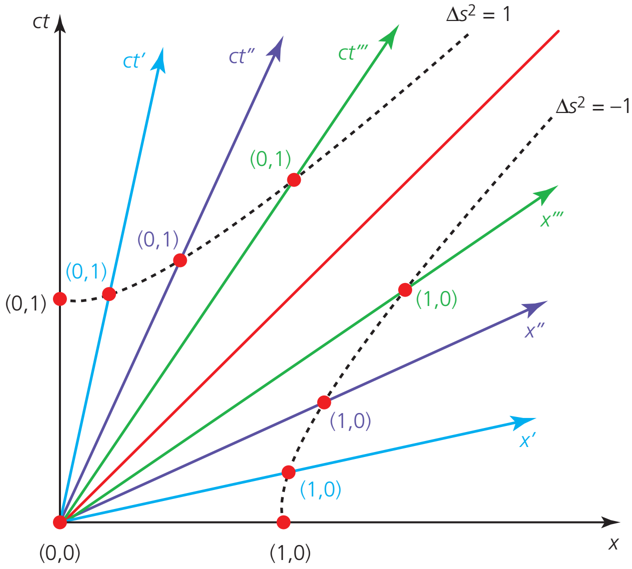
Figure A5.25 Space–time diagram for the thought experiment considering simultaneity, carried out in Figure A5.11. The red lines represent the world lines of the two reflecting light rays. The grey axes represent the inertial reference frame of the train carriage, *S*, while the blue axes represent the inertial frame of the platform, *S'*, with the train moving to the left. The dashed intersections with the timeline of each observer give their version of the order of events.

Lines of constant space–time interval

We have seen that space–time interval is in invariant quantity:

$$\Delta s^2 = c^2\Delta t^2 - \Delta x^2 = c^2\Delta t'^2 - \Delta x'^2$$

This means that we can draw lines of constant space–time interval (sometimes called invariant hyperbole) on space–time diagrams.



■ **Figure A5.26** Two lines of constant space–time interval

Figure A5.26 shows four different frames of reference (travelling at different velocities) in different colours. The dotted lines are two examples of lines of constant space–time: $\Delta s^2 = 1$ and $\Delta s^2 = -1$.

> Calculations using the coordinates (in any frame of reference) of any point on a line of constant space–time interval will give the same numerical result.

First consider the $\Delta s^2 = -1$ line. The points at which the dotted line crosses all the x-axes corresponds to $t = 0$ and so on, so that all these points, in their different frames of reference, must have coordinates of (1, 0). $c^2\Delta t^2 - \Delta x^2 = -1$ and so on, with $t = 0$.

If observers in each of the reference frames were measuring the length of a stationary rod, with one end at the origin of their reference frame (0, 0), they would all record the same proper length of the rod, 1 m.

We can see immediately that:

> The scales of the axes on space–time diagrams are not equal for different frames of reference.

The scales expand with greater velocity, as can be seen by comparing the lengths between the red dots in Figure A5.26.

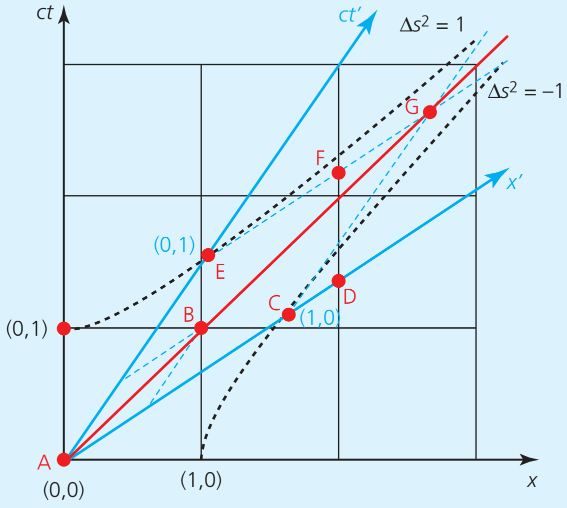
Now consider the $\Delta s^2 = +1$ line. The points at which the dotted line crosses all the ct-axes corresponds to $x = 0$ and so on, so that all these points, in their different frames of reference must have coordinates of (0, 1). $c^2\Delta t^2 - \Delta x^2 = 1$ and so on, with $x = 0$.

43 a Use a ruler, calculator and Figure A5.27 to complete Table A5.1.

■ **Table A5.1**

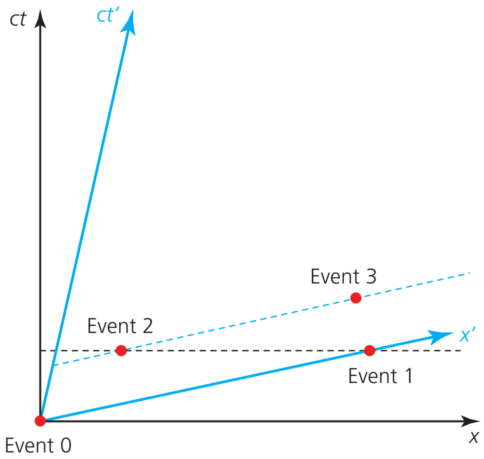
Event	Coordinates in S (x, ct)	Coordinates in S' (x', ct')
A	(0, 0)	(0, 0)
B		(0.4, 0.4)
C	(1.6, 1.1)	(1, 0)
D		
E		(0, 1)
F		
G		

b List the order in which the events occur according to observers in both reference frame S and reference frame S'.



■ **Figure A5.27** Space–time diagram showing seven events, labelled A to G, from two different reference frames

HL ONLY



■ **Figure A5.28** Space–time diagram comparing simultaneity in different reference frames

Simultaneity in space–time diagrams

Remember that all inertial observers will agree that two events are simultaneous if they occur in the same place, but they may disagree as to the order of two events that occur at two different points in space. Figure A5.28 shows a space–time diagram with four different events. According to one observer, Event 0 occurs first followed by Events 1 and 2 occurring simultaneously, with Event 3 happening last. However, for the other observer Events 0 and 1 both occur simultaneously followed by Events 2 and 3 occurring simultaneously.

Nature of science: Models

Visualization of models

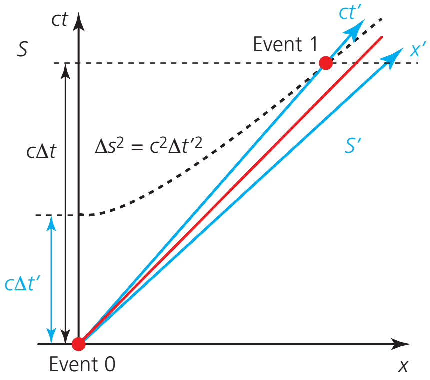
The visualization of events in terms of space–time diagrams is an enormous advance in understanding the concept of space–time.

The Lorentz transformations we use to describe how we transfer from one reference frame to another are highly mathematical and make the topic very difficult to interpret in a non-mathematical way. The geometry of space–time diagrams appears initially quite confusing, but with practice provides an entirely different way of approaching relativity. This new dimension means that aspects of relativity become significantly more accessible – in particular, space–time diagrams readily explain whether or not events are simultaneous in different reference frames and explain the order of events seen by different observers.

With more practice, space–time diagrams also explain concepts such as time dilation and length contraction but they can also be used to understand relativistic velocity additions and to visualize why it is impossible to exceed the speed of light in a vacuum.

Time dilation in space–time diagrams

Look at a space–time diagram for the muon experiment (Figure A5.29). Using the angle formula actually gives an angle of 44.9° for $v = 0.995c$, but ct' is drawn at less than this for clarity.



Event 0 is the formation of a muon by the incoming cosmic radiation, while Event 1 is the arrival of the muon at the Earth's surface. The black reference frame S is the Earth reference frame, while the blue reference frame S' is that of the muon. An observer travelling with the muon measures the proper time between Events 0 and 1. To measure this on the scale of the vertical ct-axis we follow the dashed line of constant space–time interval from Event 1 to where it crosses the ct-axis, where it can be easily calculated by measuring the interval labelled $c\Delta t'$. The interval according to the observer in reference frame S can be calculated using $c\Delta t$.

■ **Figure A5.29** Space–time diagram of the muon-decay experiment.

One of the problems with space–time diagrams is that the scales on the axes are not the same. We could use Lorentz transformations to carefully mark the scales on each axis, but there is a neat little trick that allows us to avoid this.

In the muon reference frame, Event 1 occurs a time $\Delta t'$ after Event 0, and both events occur at $x' = 0$. This is because in the muon reference frame the rest observer will see the stationary muon formed in the atmosphere (Event 0) and then the Earth's surface colliding with the stationary muon (Event 1) – the muon is stationary throughout. Thus, the observer measures the spatial separation between the two events to be zero, so the time, $\Delta t'$, is the proper time between the two events, shown correctly to scale on the *ct*-axis.

This occurs at a specific space–time interval, where $\Delta s^2 = c^2\Delta t'^2$, and we can follow the dashed line that joins all the points with this same space–time interval. Where this crosses the vertical *ct*-axis it marks the equivalent interval as measured on the scale of the *ct*-axis. On the space–time diagram this is labelled $c\Delta t'$.

In the Earth reference frame, *S*, the time interval between Events 0 and 1 is significantly longer and can be found from the vertical coordinate of Event 1. This is marked as $c\Delta t$ on the *ct*-axis. Since both measurements have been correctly scaled onto the *ct*-axis their lengths can now be directly compared, and it is clear that the proper time interval $\Delta t'$ is shorter than the stretched (or dilated) Δt time interval. Careful measurement from the *ct*-axis would show that $c\Delta t = \gamma c\Delta t'$. Note that this equation may appear to be the wrong way round because S' is measuring proper time.

44 Use the space–time diagram in Figure A5.30 and the Lorentz transformation equations to do the following calculations.

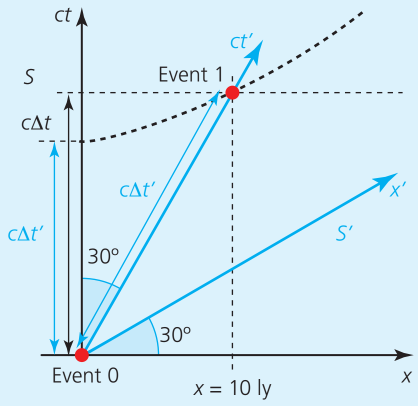
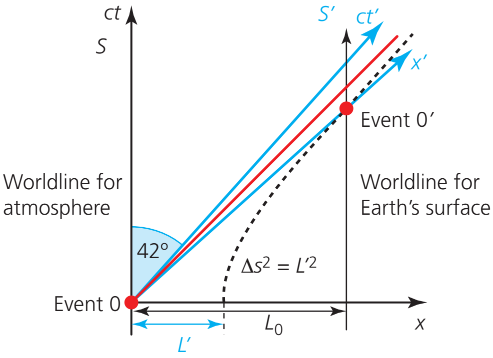


Figure A5.30

- **a** the velocity of an object that has a world line at 30°
- **b** the value of γ for this speed
- **c** the time, *t*, at which an observer in reference frame *S* will record the object to have travelled 10 ly (Event 1) at this speed
- **d** the value of $c\Delta t$ between Event 0 and Event 1.
- **e** The graph is drawn correctly to scale. Measure the length of $c\Delta t$ and $c\Delta t'$ on the *ct*-axis and show that the ratio of the measured lengths $\frac{c\Delta t}{c\Delta t'} \approx \gamma$.
- **f** State which reference frame is measuring proper time.
- **g** Use time dilation to calculate the value of $c\Delta t'$.
- **h** Mark the position of 14 ly on both the vertical black axes (*S* reference frame) and blue axes (S' reference frame) to show that the scales on the axes are different.

Length contraction in space–time diagrams

Once again let us turn to the muon experiment as shown in Figure A5.31. The length being measured is the distance between the formation of the muons in the Earth's atmosphere and the surface of the Earth.



🟥 **Figure A5.31** Space–time diagram for the muon-decay experiment used to demonstrate length contraction. The instantaneous separation between the atmosphere and the Earth's surface in the muon reference frame must be measured in each reference frame.

In the previous section, the length contraction equation was harder to derive than the time-dilation equation because it required one more key piece of information – in order to measure a length correctly we must measure the position of each end of the length at the same time. In other words, the two space–time events used to determine the two ends of the length in a given reference frame must occur simultaneously in that reference frame.

This length is straightforward to measure in the Earth reference frame because it is simply the horizontal separation between the vertical worldline of the Earth's atmosphere and the world line of the Earth's surface. These are shown on the space–time diagram as the two vertical black axes. Because each is stationary in the Earth reference frame, the separation between them is a proper length and is labelled L_0 on the diagram, where it could easily be measured on the x-axis.

In the muon reference frame, S', the distance between the Earth's atmosphere and the Earth's surface can be measured using simultaneous events 0 and 0′, where the world line of the atmosphere and the world line of the Earth's surface each cross the x'-axis. In the muon reference frame both events occur when $t' = 0$, so they can be used to correctly measure the separation, L' – it could be measured off the scale on the x'-axis but we would need to calculate the scale to do this.

Instead, we can sketch on the curve that links all the points with space–time interval $\Delta s^2 = -L'^2$. Extending this to the x-axis gives the separation between Events 0 and 0′ on the x-axis scale, where it can easily be measured. It can clearly be seen that the proper length is much larger than the contracted length, L', confirming that length contraction occurs. Careful measurement would also show that:

$$L' = \frac{L_0}{\gamma}$$

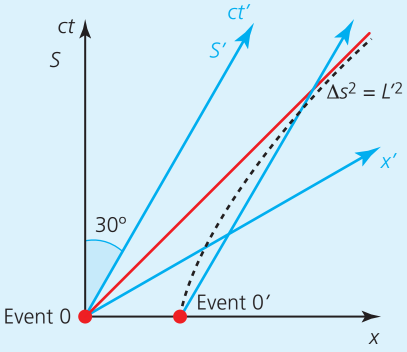
Once again, the space–time diagram's geometry has been able to represent the dynamics of relativity.

45 In Figure A5.31 the angle between the x and x'-axes is also 42°.

- **a** Calculate the relative velocity of the two frames of reference.
- **b** Calculate the value of γ.
- **c** Calculate the length of L', if $L_0 = 1.0\,\text{m}$.
- **d** Using a ruler, measure the ratio of L_0 / L' to confirm that this gives the value of γ.

46 Einstein's first postulate stated that the laws of physics are the same in all inertial reference frames. This means that we should be able to show on a space–time diagram that an object that is stationary in reference frame S' will also be measured as having a contracted length by an observer in S.

Use Figure A5.32 to show that this is indeed the case by marking on the length as measured by S' on the x' axis and using the space–time interval curve to mark on the x'-axis the equivalent length as measured by S. Hence, use the measured lengths along the x'-axis to estimate the value for γ.



■ **Figure A5.32**

B.1 Thermal energy transfers

Guiding questions

- How do macroscopic observations provide a model of the microscopic properties of a substance?
- How is energy transferred within and between systems?
- How can observations of one physical quantity be used to determine the other properties of a system?

Thermal energy, internal energy and heat

These three terms are often used interchangeably, which can be confusing! Unfortunately, different teachers and different books have varying interpretations, so it is important to clarify, from the beginning of this topic, exactly how these terms will be used in this course.

All substances contain particles / molecules which can have individual kinetic energies and potential energies. There are more details about this later in this topic. We will describe the total of all these particle energies as the *internal energy* of a substance. We will not describe the energy inside substances as thermal energy, or heat.

Energy is always transferred from hotter objects to cooler objects. We will describe this transfer as *thermal energy*, although the word 'heat' is also widely used to describe this type of energy transfer.

TOK

The natural sciences

- Does the precision of the language used in the natural sciences successfully eliminate all **ambiguity**?

Ambiguities

There is, perhaps, nowhere else in the study of physics where such important terms have such ambiguous uses. It is interesting to consider how this has arisen, why it is not corrected, and whether it is truly important.

If you understand the theory of particle energies inside matter, does it really matter if your teacher calls it 'internal energy', while a book refers to is as 'thermal energy' and your friend calls it 'heat'? To what extent is precise language important to our understanding of underlying physics?

◆ **Ambiguity** Open to different interpretations.

Kinetic theory of matter

SYLLABUS CONTENT

- Molecular theory in solids, liquids and gases.
- Density, ρ, is given by: $\rho = \frac{m}{V}$.

◆ **Kinetic theory of matter** All matter is composed of a very large number of small particles that are in constant motion.

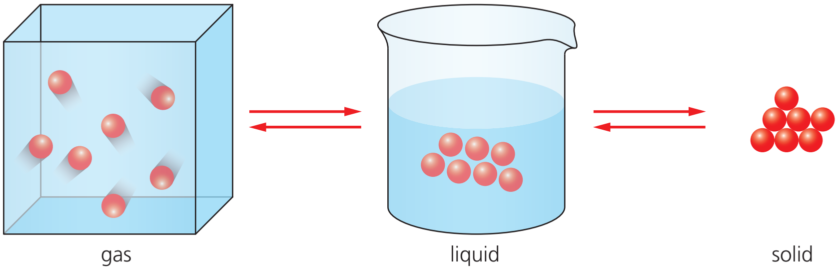
◆ **Ions** An atom or molecule that has gained or lost one or more electrons.

◆ **Atoms** The particles from which chemical elements are composed. They contain subatomic particles.

The essential idea that all matter contains countless billions of particles constantly moving (in various ways), with forces between them when they get close enough, is possibly the most important theory in the whole of science. The microscopic **kinetic theory of matter** is the starting point that can be used to help explain so much of what we observe in our macroscopic everyday life. To begin with, it can explain the different properties of solids, liquids and gases.

Solids, liquids and gases

The 'particles' we are referring to in the kinetic theory are usually molecules, but they could also be **ions**, or **atoms**. Figure B1.1 shows a simplified visual model of particle arrangements. Table B1.1 offers generalized comments on the major differences.



■ **Figure B1.1** Particle arrangements

■ **Table B1.1** Differences between particles in solids, liquids and gases

	Solid	**Liquid**	**Gas**
Arrangement of particles	regular patterns	no pattern	no pattern
Forces between particles	attractive and large enough to keep particles in their positions	some particles have enough energy to overcome attractive forces	negligible (except in collisions) under most conditions
Separation of particles	close together	still close together	much further apart
Motion of particles	vibrate in fixed positions	some limited random movement is possible	all move in random directions and usually with high speeds

Density

Clearly, the more massive the individual particles are, and/or the closer they are together, the greater will be the total mass of a given volume of a substance.

$$\text{density} = \frac{\text{mass}}{\text{volume}} \qquad \rho = \frac{m}{V} \qquad \text{SI unit: kg m}^{-3}$$

◆ **Expand** Increasing in size. An expansion of a gas is an increase in volume.

When solids and liquids are heated to a higher temperature, they will usually **expand** slightly in size because there will be a very small increase in the separation of particles. This means that there will be a small decrease in their densities.

When gases are heated, they will only expand if they are in a container that will allow that to happen.

Table B1.2 Typical values for densities (gases at 0 °C and normal atmospheric pressure)

Substance	**Density / kg m^{-3}**
helium	0.18
air	1.23
carbon dioxide	1.98
wood (pine, approx.)	500
ethanol	810
ice	910
water (at 20 °C)	998
water (at 4 °C)	1000
sea water (approx.)	1030
aluminium	2710
average density of Earth	5520
iron	7870
gold	19 300
black hole	1×10^{15}

The particles in solids and liquids can be considered to be as close together as possible, that is, they are effectively 'touching' each other. Since gases are about 1000 × less dense than solids and liquids, their molecules are typically 10 molecular diameters apart from each other.

1 Calculate the density of olive oil in SI units, if a mass of 125 g has a volume of 137 cm^3.

2 An iron bar has the dimensions 5.0 × 5.0 × 25.0 cm. What is its mass?

3 Explain how we can conclude from Table B1.2 that the molecules in air are approximately 10 times further apart than the molecules in water.

4 Outline why ice floats on water. (Refer to buoyancy forces from Topic A.2.)

5 Water has its maximum density at 4 °C. How does this affect the formation of ice on, for example, a lake in very cold weather?

6 In everyday life the volume of liquids is more often measured in litres, l, (and cl and ml) rather than in m^3 or cm^3. 1 litre is a volume of 1000 cm^3.
Determine the volume, in litres, of some ethanol which has a mass of 1.0 kg.

7 It is common practice in many Asian countries to see people making merit by placing gold leaf on Buddhist images, as shown in Figure B1.2.

Figure B1.2 Placing gold foil on a statue of Buddha

Gold is extremely *malleable* – meaning that it can be hammered relatively easily into different shapes, including *very* thin foil (approximately 2×10^{-7} m).

Predict what area of gold foil of thickness 1.80×10^{-5} cm can be made from each 1.0 g of gold.

Temperature

SYLLABUS CONTENT

- Temperature difference determines the direction of the resultant thermal energy transfer between bodies.
- Kelvin and Celsius scales are used to express temperature.
- A change in temperature of a system is the same when expressed with the Kelvin or Celsius scales.
- Kelvin temperature is a measure of the average kinetic energy of particles: $\bar{E}_k = \frac{3}{2}k_B T$.

◆ **Thermal equilibrium** All temperatures within a system are constant.

◆ **Thermal contact** Objects can be considered to be in thermal contact if thermal energy (of any kind) can be transferred between them.

Macroscopic understanding of temperature

In everyday life, temperature is a number obtained from a thermometer, which informs us how hot, or cold, something is. But this is a long way from an acceptable scientific definition. Thinking a little more deeply, we realize that temperature values tell us something about energy transfer: Energy always flows spontaneously from a hotter place to a colder place. As we have already noted, we will call this a flow of *thermal* energy.

The resultant flow of thermal energy is always from higher temperature to lower temperature.

Consider an isolated system such as shown in Figure B1.3. If there is no net transfer of thermal energy between A and B, then they must be at the same temperature, and we describe them as being in **thermal equilibrium**. Similarly for B and C. If B is in thermal equilibrium with A and C, then A and C are also in thermal equilibrium with each other, and at the same temperature.

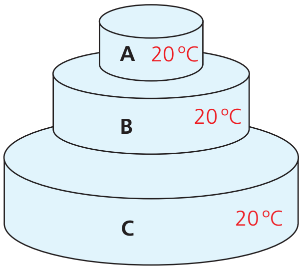


Figure B1.3 Thermal equilibrium: same temperature, no flow of thermal energy

Consider the simple example of a system of two objects (or substances) at different temperatures able to transfer thermal energy between themselves, but isolated from everything around them (their surroundings). The hotter object will transfer energy to the colder object and cool down, while at the same time the colder object warms up. As the temperature difference between the two objects gets smaller, so too does the rate of thermal energy transfer (as represented by the gradients of the graphs). This is represented in Figure B1.5, which shows how the temperature of two objects (A and B) might change when they are placed in good **thermal contact** with each other. Being in 'thermal contact' means that thermal energy can be transferred between them, by any means. But the phrase is most often used when referring to significant physical contact.

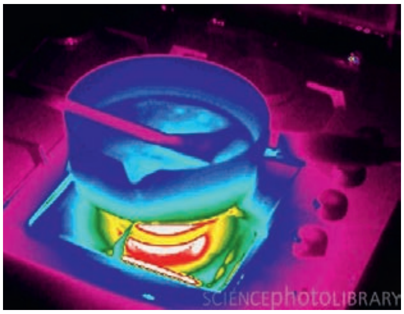


Figure B1.4 This thermogram, taken using infrared radiation, uses colour to show different temperatures in a saucepan on a cooker. The scale runs from white (hottest) through red, yellow, green and blue to pink (coldest)

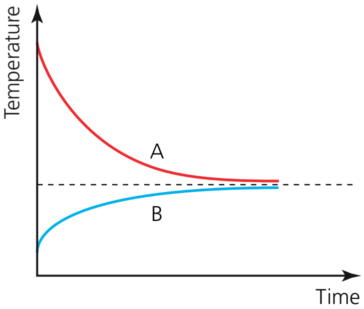


Figure B1.5 Two objects (A and B) at different temperatures, insulated from their surroundings but not from each other, will reach thermal equilibrium

Eventually A and B will reach the same temperature. If the temperatures have stopped changing and both objects are at the same temperature, the objects are in thermal equilibrium and there will be no net flow of thermal energy between them. In any realistic situation, it is not possible to completely isolate / insulate two objects from their surroundings, so the concept of thermal equilibrium may seem to be idealized.

Common mistake

0 °C is simply the freezing point of pure water. It has no other meaning. It is not a 'true' zero. For example, it would be wrong to think that 20 °C was double the temperature of 10 °C.

◆ **Celsius (scale of temperature)** Temperature scale based on the melting point (0 °C) and boiling point (100 °C) of pure water.

◆ **Thermometer** An instrument for measuring temperature.

◆ **Kelvin scale of temperature** Also known as the **absolute temperature scale**. Temperature scale based on absolute zero (0 K) and the melting point of water (273 K). The kelvin, K, is the fundamental SI unit of temperature. T (in K) = 0 °C + 273.

◆ **Absolute zero** Temperature at which (almost) all molecular motion has stopped (0 K or −273 °C).

The concept of hotter objects always getting colder, and colder objects always getting hotter, suggests an important concept: eventually everything will end up at the same temperature.

Temperature scales

Celsius scale of temperature

A temperature scale needs two fixed points. For the **Celsius scale**, these are the freezing point and the boiling point of pure water (under specified conditions). An instrument for measuring temperature, a **thermometer**, can then calibrated by marking these two points as 0 °C and 100 °C, and then dividing the interval between them into one hundred equal divisions. Higher and lower temperatures can then be determined by *extrapolation*.

Kelvin scale of temperature

The choice of the two fixed points on the Celsius scale is arbitrary and mainly for convenience. However, there is a more logical scale – the **Kelvin scale** – that is widely used in science, but in everyday life, people around the world have become used to the Celsius scale (and Fahrenheit scale).

There is a temperature at which the kinetic energy of all particles reduces to (almost) zero. This is known as **absolute zero** and it is discussed in more detail in Topic B.3. On the Celsius scale absolute zero has the value of –273.15 °C. This temperature is the lower fixed point, zero, of the Kelvin temperature scale, which is sometimes described as the **absolute temperature scale**.

Zero kelvin (0 K) is the lowest possible temperature (= –273.15 °C)

The upper fixed point of the Kelvin scale also effectively uses the melting point of pure water which, for convenience, is given the value of +273.15 K. This scale defines the SI unit of temperature, the kelvin, K. Defining the Kelvin temperature scale in this way means that a *change* of temperature has the same numerical value in both the Celsius and the Kelvin scales. Table B1.3 compares some Celsius and kelvin temperatures (to the nearest whole number).

The symbol T is used for temperature in kelvin. Θ is often used for a temperature in degrees Celsius.

■ **Table B1.3** A comparison of temperatures in degrees Celsius and Kelvin

Temperature	**°C**	**K**
absolute zero	−273	0
melting point of water	0	273
body temperature	37	310
boiling point of water	100	373

Temperature in kelvin, T/K = temperature in Celsius, Θ/°C + 273

WORKED EXAMPLE B1.1

a The freezing point of ethanol is –114 °C. Convert this temperature into kelvin.

b The melting point of aluminium is 933 K. Convert this temperature to degrees Celsius.

c On a cold night the temperature dropped from +10 °C to −10 °C. Calculate this change of temperature in kelvin.

Answer

a $T = -114 + 273 = 159\,\text{K}$

b $\Theta = 933 - 273 = 660\,°\text{C}$

c $(-10) - (+10) = -20\,\text{K}$

Tool 2: Technology

Use sensors

Electronic sensors respond to a particular physical quantity by producing a corresponding voltage. That voltage must then be converted to digital form before it can be understood, processed and displayed by appropriate software. There are several possibilities, including where a separate data logger/interface, is connected between the sensor and a computer. See Figure B1.6.

Alternatively and more conveniently, the sensor may be connected to a corresponding all-in-one unit which processes and displays results. In some cases, a mobile phone app can be used. Bluetooth connections are also available.

The advantage of using sensors in this way are obvious:

- Data can be collected over very short times (less than a second), or times which are otherwise inconveniently long.
- A large amount of data can be gathered.
- The data can be stored.
- The data can be very quickly processed and graphs drawn.

The sensors which are in common use in physics experiments at this level include:

- position and motion
- pressure
- temperature
- sound level
- light level
- current and voltage
- magnetic field.

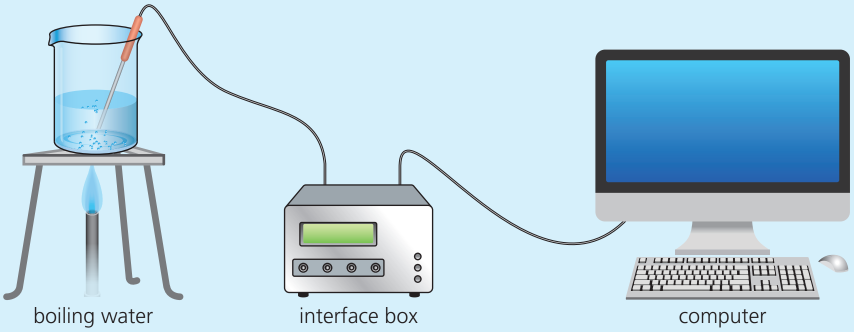


Figure B1.6 Separate sensor, interface and computer

Measuring temperature, using sensors

In principle, any physical quantity which varies with temperature could be used to construct a thermometer. However, it is better to use a physical quantity that varies significantly and regularly over a wide range of temperatures. The most common types of thermometer involve:

- variation in length of a liquid along a thin (capillary) tube
- variation in pressure of a fixed volume of gas
- variation of electrical resistance
- variation in the voltage generated by wires of different metals joined together
- variation in infrared radiation from a surface.

The part of the thermometer which is used to measure the temperature is placed in good thermal contact with the material whose temperature is to be measured (except for infrared thermometers). After sufficient time for the thermometer to reach thermal equilibrium with the material, a reading can be taken. Other sources of thermal energy should be avoided. For example, when measuring air temperature, the thermometer should not be receiving thermal energy directly from the Sun.

Electronic resistance thermometers (sensors) have an obvious advantage in that they can provide immediate digital data when they are connected to a computer through an interface.

8 Convert the following kelvin temperatures into degrees Celsius: **a** 175 **b** 275 **c** 10 000.

9 The world's highest and lowest recorded weather temperatures are reported to be 56.7 °C (California – see Figure B1.7) and −89.2 °C (Antarctica).

- **a** State values for these temperatures in kelvin.
- **b** Some hot water at 68 °C cools down to 22 °C. What is this change of temperature in kelvin?

Figure B1.7 The world's highest temperature to date was recorded in California, USA.

10 The volume of a gas was 67 cm^3 when the temperature was 22 °C.

- **a** If the volume is proportional to the absolute temperature, calculate the volume if the temperature increases to 92 °C.
- **b** Predict the volume *if* the gas temperature could be reduced to 0 K.

11 A student calibrated an unmarked liquid-in-glass thermometer. The liquid expands up a thin capillary tube as it gets hotter. She has correctly marked the upper and lower fixed points as 0 °C and 100 °C. The two marks were 10.7 cm apart.

- **a** She then wanted to use her thermometer to measure room temperature and she left it in the laboratory for 10 minutes. The level of the liquid was then 2.6 cm above the lower fixed point.
  Explain why the thermometer was left undisturbed for 10 minutes.
- **b** Calculate a value for room temperature.
- **c** State an assumption that you have to make to answer part **b**.

Microscopic understanding of temperature

A true understanding of temperature is to be found in the kinetic theory of matter. Consider the example of two samples of the same gas, as shown in Figure B1.8. Suppose that the molecules on the right have a lower average kinetic energy (and speed) than the molecules on the left. After they have collisions, the faster moving molecules will slow down and the slower molecules will speed up (conservation of momentum in elastic collisions). In this way energy is transferred from the left to the right. This is equivalent to a transfer of thermal energy, so we must conclude that the left-hand side of the figure represents a higher temperature. Eventually, the average energies and speeds on both sides will become equal and a macroscopic interpretation would be that they were in thermal equilibrium at the same temperature.

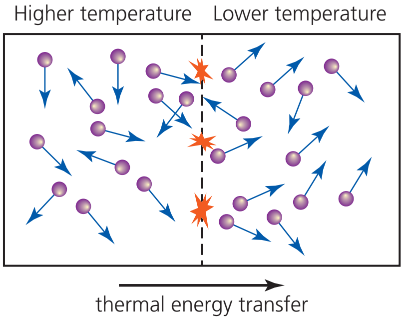


Figure B1.8 Energy transfers between molecules

Average particle speed indicates different temperatures when we compare samples of the *same* gas. Increasing temperature corresponds to greater average speed. More generally, when we compare *different* gases, we need to consider the average kinetic energy of the particles, rather than their speeds.

Temperature (K) is proportional to the average random translational kinetic energy, $\overline{E}_k$, of particles in a gas.

◆ **Vibrational kinetic energy** Kinetic energy due to vibration/oscillation.

◆ **Boltzmann constant, k_B** Important constant that links microscopic particle energies to macroscopic temperature measurements.

All gases, at the same temperature, contain particles with the same average translational kinetic energy.

The particles in most gases are molecules, which means that they also have other forms of kinetic energy (not just translational), for example, rotational kinetic energy and **vibrational kinetic energy**.

The all-important mathematical connection between macroscopic measurements of kelvin temperature, T, and the microscopic concept of individual molecular kinetic energies is provided by the Boltzmann constant in the following equation:

Average random translational kinetic energy of a gas particle: $\boldsymbol{\overline{E} = \frac{3}{2}k_B T}$

k_B is known as the **Boltzmann constant**. It has the value $\mathbf{1.38 \times 10^{-23}\,J\,K^{-1}}$

LINKING QUESTION

- How is the understanding of systems applied to other areas of physics?

Top tip!

In Topics A.2 and A.3 we discussed collisions between *macroscopic* objects, describing the collisions as either elastic (total kinetic energy of the objects is conserved) or inelastic. During inelastic collisions, energy is transferred to the surroundings, dissipated, mostly in the form of internal energy and thermal energy. That is, energy is transferred from the ordered kinetic energy of countless billions of particles moving together in the same direction in the objects as a whole, to the disordered random kinetic energies of individual particles.

Energy dissipation is a macroscopic concept and cannot be applied to microscopic particle collisions. Total kinetic energy can only decrease in a collision between two particles if it is used to cause ionization (see Topic B.5).

WORKED EXAMPLE B1.2

Calculate the average kinetic energy of translation of gas molecules at room temperature.

Answer

Using 20 °C (293 K) as room temperature,

$$\overline{E}_k = \frac{3}{2}k_B T = 1.5 \times (1.38 \times 10^{-23}) \times 293 = 6.1 \times 10^{-21}\,\text{J}$$

The energy of particles in liquids and solids is more complicated because of the significant forces between the particles. In general, however, the following is always true:

A temperature rise is equivalent to the particles gaining kinetic energy.

12 Consider the gas in the previous worked example. If the particles have a mass of 5.3×10^{-26} kg, use the equation for linear kinetic energy to estimate their average speed.

13 The surface temperature of the Sun is 5800 K. Calculate the average kinetic energy of the particles it contains, assuming that the equation in Worked example B1.2 can be applied.

14 A cylinder of gas contains gas molecules moving with an average speed of $400\,\text{m}\,\text{s}^{-1}$, which is characteristic of their temperature of 20 °C. If the cylinder is then put on an aircraft which is moving at $200\,\text{m}\,\text{s}^{-1}$, discuss what will happen to the average speed of the gas molecules and the temperature of the gas.

15 Nitrogen and oxygen are the two principal gases in air. Oxygen molecules are slightly more massive than nitrogen molecules. Explain how the average speeds of the molecules will compare in the air you are breathing.

◆ **Sense perception** How we receive information, using the five human senses.

TOK

Knowledge and the knower

- How do we acquire knowledge?
- To what extent are technologies merely extensions to the human senses, or do they introduce radically new ways of seeing the world?

Sense perception (of temperature)

'Information' received directly by receptors in the human body and then processed by our brains, is described as **sense perception**.

It is often said that we have five senses (hearing, sight, smell, touch and taste), but we also have a limited ability to detect changes in temperature and the flow of thermal energy into, or out of, our bodies. Most people are able to estimate the approximate temperature of the air around them.

However, this 'way of knowing' using sense perception can be unreliable. Whether we are hot or cold is a very common topic of conversation, but people in the same environment can sometimes disagree about the temperature that they sense. Being able to consult an instrument capable of measuring the temperature (a thermometer) has obvious advantages, in everyday life as well as in scientific experiments. However, such reliable measurements were not possible until about 300 years ago.

■ **Figure B1.9** Fahrenheit with his thermometer

A German physicist, Daniel Fahrenheit (Figure B1.9) invented the first accurate thermometer in 1709. He used the expansion of mercury along a thin tube.

Internal energy

All substances contain moving particles. Moving particles have kinetic energy. The particles might be moving in different ways, which gives rise to three different forms of kinetic energy:

- Particles might be vibrating about fixed positions (as in a solid) – this gives the particles *vibrational* kinetic energy.
- Particles might be moving from place to place (translational motion) – this gives the particles *translational* kinetic energy in liquids and gases.
- Molecules might also be rotating – this gives the molecules *rotational* kinetic energy.

Particles can have potential energy as well as kinetic energy. In solids and liquids, it is the electrical forces (between charged particles) that keep particles from moving apart or moving closer together. Wherever there are electrical forces there will be electrical potential energy in a system, in much the same way as gravitational potential energy is associated with gravitational force.

If the average separation of the particles in a solid or liquid increases, so too does their potential energy (and the total internal energy of the substance.)

In gases, however, the forces between molecules (or atoms) are usually negligible because of the larger separation between molecules. This is why gas molecules can move freely and randomly. The molecules in a gas, therefore, usually have negligible electrical potential energy – all the energy is in the form of kinetic energy.

So, to describe the total energy of the particles in a substance, we need to take account of both the kinetic energies and the potential energies. This is called the internal energy of the substance and is defined as follows:

> The internal energy of a substance is the sum of the total random kinetic energies and total potential energies of all the particles inside it.

In the definition of internal energy given above, the word 'random' means that the particle movements are disordered and unpredictable. That is, they are not linked in any way to each other, or ordered – as their motions would be if they were all moving together, such as the particles in a macroscopic motion of a moving object. The particles in a moving object have both the ordered kinetic energy of macroscopic movement together and the random kinetic energy of internal energy.

Nature of science: Theories

Caloric fluid

An understanding of thermal energy and internal energy depends on the kinetic theory of matter, but that theory is less than 200 years old. Before that, 'heat' was often explained in terms of a vague invisible 'caloric fluid' that flowed out of a hot object, where it was concentrated, to a colder place where it was less concentrated.

This is an example of one of many serious scientific theories that were developed to explain observed phenomena, but which were never totally satisfactory because they could not explain all observations. The earlier 'phlogiston' theory of combustion is another such theory related to heat and combustion.

Looking back from the twenty-first century, these theories may seem unsophisticated and inaccurate (but imaginative!). However, they should be judged in the context of their times, and at the time of these theories (seventeenth and eighteenth centuries) the kinetic theory of matter had not been developed, so the current understanding of the flow of thermal energy was not possible.

Thermal energy

SYLLABUS CONTENT

- Conduction, convection and thermal radiation are the primary mechanisms for thermal energy transfer.

Thermal energy is the name we give to the transfer of energy because of a temperature difference: a net flow from hotter to colder.

There are three principal ways in which thermal energy can be transferred:

- **Thermal conduction**. In which kinetic energy is transferred between particles.
- **Convection**. In which differences in the densities of liquids and gases result in their movement.
- **Thermal radiation**. In which electromagnetic radiation is emitted by surfaces.

We will discuss each of these in detail in the next three sections.

Thermal conduction

SYLLABUS CONTENT

- Conduction in terms of the difference in kinetic energy of particles.
- Quantitative analysis of rate of thermal energy transferred by conduction in terms of the type of material and cross-sectional area of the material and the temperature gradient as given by:
  $\frac{\Delta Q}{\Delta t} = -kA\frac{\Delta T}{\Delta x}$

As mentioned earlier in this topic, when gas particles (usually molecules) have elastic collisions, the slower moving particles gain kinetic energy and the faster moving particles lose kinetic energy. In this way, over time, there will be a net transfer of energy from a place where particles, on average, are moving faster to a place where they are moving slower on average. That is, from hotter to colder. Such a transfer will continue until the whole of the gas has particles with the same average kinetic energy, when thermal equilibrium has been reached and a constant temperature reached.

◆ **Conduction (thermal)** Passage of thermal energy through a substance as energy is transferred from particle to particle.

Similar ideas can be applied to the transfer of energy between particles in liquids and solids. This type of thermal energy transfer, from particle to particle, is called **thermal conduction**. Figure B1.10 gives an impression of thermal conduction through a solid, although vibrations and increasing kinetic energy are not easily represented in a single picture!

Thermal conduction occurs because of the transfer of kinetic energy between particles.

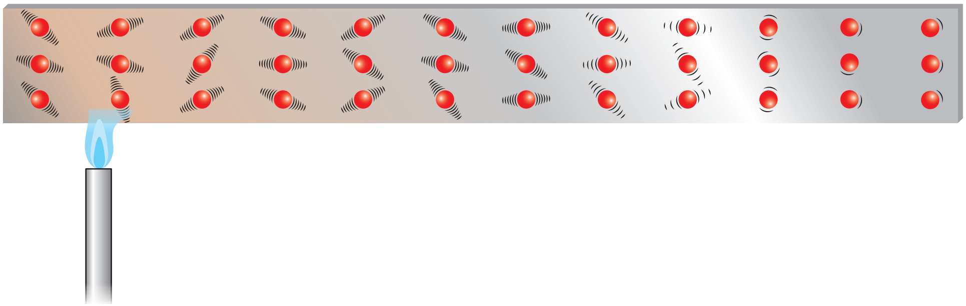


Figure B1.10 Thermal conduction through a solid

Common mistake

Many students think that thermal conduction only occurs in solids. This is not true, although some solids, especially metals, are by far the best thermal conductors. See Table B1.4.

Consider again Figure B1.10. The solid bar has gained its thermal energy by conduction from the hot gas in the flame.

In solids, the particles vibrate in fixed positions, with forces between them. In Figure B1.10, the particles on the left-hand side are vibrating faster and have greater vibrational kinetic energy (on average) because the solid is at a higher temperature. Energy is transferred through the solid, to the right, because of the forces / interactions between particles.

Solids are generally better thermal conductors than liquids, and liquids conduct better than gases. This can be explained by considering the closeness of particles and the strength of forces between them.

Table B1.4 lists various substances and their *thermal conductivities*, which are explained later in this topic. A larger number means that the substance is better at conducting thermal energy: more energy is transferred under similar conditions. (Metals are good conductors because they contain many *free / delocalized electrons*.)

■ **Table B1.4** Typical thermal conductivities at room temperature (some approximate)

Substance	**Thermal conductivity / $W\,m^{-1}\,K^{-1}$**
vacuum	0.00
carbon dioxide	0.15
air	0.025
polyurethane foam	0.03
paper	0.05
rubber	0.13
wood	0.15
common plastics	0.2
water	0.59
concrete and brick	0.72
glass	0.86
carbon	1.7
ice	2.1
iron	84
aluminium	237
copper	385

◆ **Insulator (thermal)** A material that significantly reduces the flow of thermal energy.

We often describe a substance as being either a good thermal conductor or a good **thermal insulator** (poor conductor). However, these are not precisely defined terms, although it should be clear from looking at Table B1.4 that the last three are much better at conducting thermal energy than any of the rest. These three would be described as good thermal conductors; the rest are usually described as insulators.

16 How can you explain that a vacuum has a thermal conductivity of zero?

17 The last three substances in Table B1.4 are all good conductors of thermal energy. State what they have in common.

18 Compare the ability of air, water, glass and copper to conduct thermal energy. (Determine ratios.)

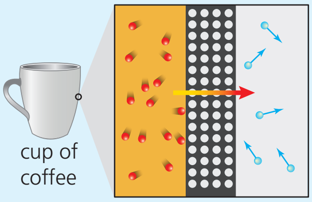
19 Explain why a metal door handle will often feel cooler than a plastic handle at the same temperature.

20 Discuss whether you would describe carbon as a conductor, or an insulator.

21 Explain the choice of materials in the manufacture of the frying pan shown in Figure B1.11.

■ **Figure B1.11** Frying pan

22 Outline the transfers of thermal energy represented in Figure B1.12.



■ **Figure B1.12** Transfer of thermal energy

23 Wetsuits are made from neoprene foam rubber (see Figure B1.13). Suggest how this can keep the surfer from getting too cold.

Figure B1.13 Surfer wearing a wetsuit

Quantitative treatment of thermal conductivity

Consider Figure B1.14, which represents the flow of thermal energy by conduction through an isolated system of a specimen of a single substance, which has an area A, and a thickness Δx. The symbol Q will be used for thermal energy.

A flow of thermal energy occurs because the left-hand side is at a higher temperature than the right-hand side: $T_H > T_C$.

The rate of thermal energy flow, $\Delta Q / \Delta t$, will be proportional to the temperature difference, ΔT, and the area, A, but inversely proportional to the thickness, Δx. It also obviously depends on the thermal properties of the substance involved. In summary:

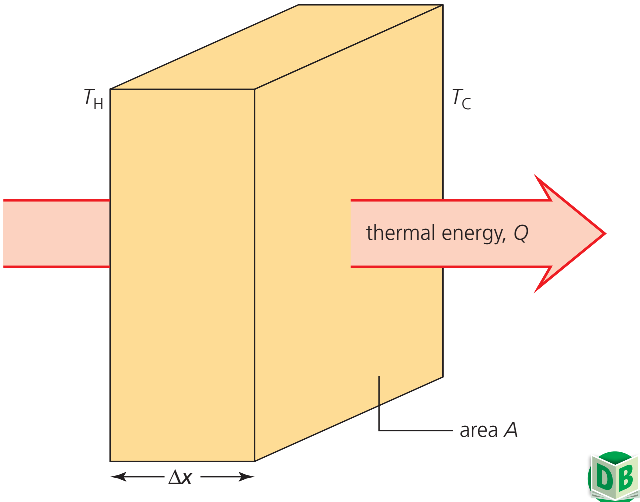


Figure B1.14 Thermal energy flowing through a block

Rate of transfer of thermal energy by conduction:

$$\frac{\Delta Q}{\Delta t} = kA\frac{\Delta T}{\Delta x}$$

◆ Thermal conductivity, *k*
Constant that represents the ability of a substance to conduct thermal energy.

k is a constant, different for each substance. It is called the **thermal conductivity** of the substance (as shown in Table B1.4). Unit: W m^{-1} K^{-1}.

$\frac{\Delta Q}{\Delta t}$ is a flow of energy per second (a power) so it is measured in watts.

WORKED EXAMPLE B1.3

The outside brick wall (single layer) of a house is 4.85 m × 2.88 m. It contains a closed glass window of dimensions 1.67 m × 1.23 m. On a hot afternoon the outside air temperature is 34.0 °C, while it is 27.0 °C inside the room.

Use the equation above to calculate the flow of thermal energy through:

a the wall of thickness 25 cm

b the window of thickness 4.5 mm.

Use data from Table B1.4.

Answer

a $$\frac{\Delta Q}{\Delta t} = kA\frac{\Delta T}{\Delta x}$$

$$= 0.72 \times \big((4.85 \times 2.88) - (1.67 \times 1.23)\big) \times \frac{7.0}{0.25}$$

$= 2.4 \times 10^2$ W into the room.

b $$\frac{\Delta Q}{\Delta t} = kA\frac{\Delta T}{\Delta x}$$

$$= 0.86 \times (1.67 \times 1.23) \times \frac{7.0}{0.0045}$$

$= 2.7 \times 10^3$ W into the room.

The thermal conductivities of brick and glass are similar. Much more thermal energy flows through each cm^2 of the glass because it is significantly thinner.

It should be noted that these calculations *considerably overestimate* the magnitude of thermal energy flows. This is because the surface temperatures of the glass and brick cannot be assumed to be the same as the surrounding air temperatures (as was done in answering the question).

The best insulator for limiting thermal energy flowing out of, or into, homes is air. See Table B1.4. However, if the air can move, thermal energy can also be transferred by *convection currents* (see next section). Various kinds of foam consist mainly of air, but the foam limits the movement of that air. Figure B1.15 shows polyurethane foam between the outer and inner walls of the outside of a house. Similar insulation can be used under the roof and below the ground floor.

Parallel sheets of glass (known as *double glazing*), as seen in Figure B1.16, can be used to trap air and limit thermal energy flow through a window. Obviously, no foam can be put between the sheets of glass, but convection is limited by keeping the separation small. Double glazing has the added benefit of reducing the transfer of sound.

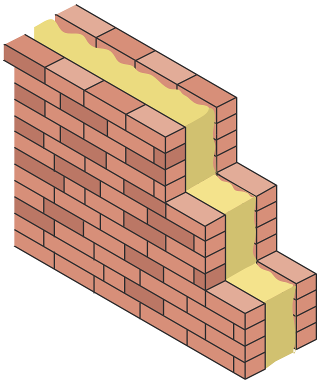


Figure B1.15 Foam insulation in a 'cavity wall'.

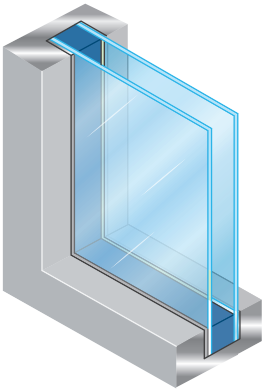


Figure B1.16 Double glazing

24 In an experiment to measure the thermal conductivity of a disc of wood, a sample of area 65 cm^2 was used, with a thickness of 5.2 mm. When the surfaces of the wooden disc were kept at 0 °C and 100 °C, the flow of thermal energy through the wood was determined to be 16 W.
Use this data to calculate a value for the thermal conductivity of the wood.

25 In 10 minutes, a total of 275 J of thermal energy was conducted through a block of material of area 12.5 cm^2 when it had a temperature gradient of 4.2 °C cm^{-1} across it.
  - **a** Determine a value for the thermal conductivity of the material.
  - **b** Would you describe this material as a conductor or an insulator?
  - **c** Suggest a material it might have been.

26 **a** If 5.6 W of thermal energy was flowing through each square metre of the insulating polyurethane foam seen in Figure B1.15, calculate the temperature difference between its surfaces if the foam had a thickness of 7.8 cm.
  - **b** Determine a value for the outside temperature if the thickness of the brick walls was 10.9 cm and the inside temperature of the outer brick wall was 5.4 °C.
  - **c** Determine the inside temperature of the interior wall.
  - **d** Sketch the arrangement and annotate your drawing with all the known data.

Thermal convection

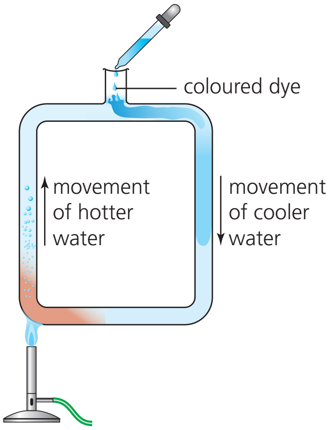
SYLLABUS CONTENT

▶ Qualitative description of thermal energy transferred by convection due to a fluid density difference.

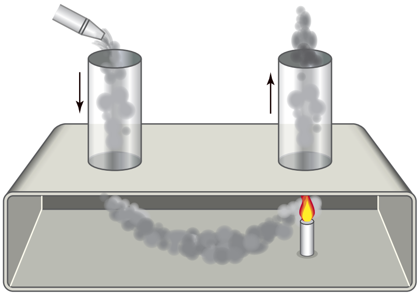
◆ **Convection** Passage of thermal energy through liquids and gases due to the movement of the substance because of differences in density.

When part of a fluid (gas or liquid) is heated, there will be a localized decrease in density. Because of increased buoyancy (see Topic A.2), the warmer part of the fluid will then rise and flow above the cooler part of the fluid, which has a slightly greater density. This movement of thermal energy in a fluid is called thermal **convection**. It is common for convection to produce currents and a circulation of a gas or liquid. Figures B1.17 and B1.18 show two common laboratory demonstrations of convection.

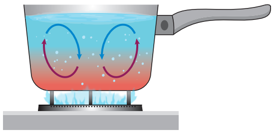
Thermal convection is the transfer of thermal energy because of the movement of a fluid due to changes in density.



■ **Figure B1.17** Demonstrating convection in water



■ **Figure B1.18** Demonstrating convection in air



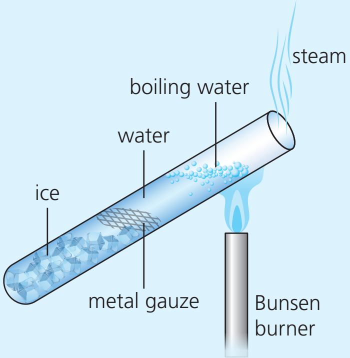
■ **Figure B1.19** Convection of water in a saucepan

There are large number of examples of convection currents, including:

- The water in heaters, saucepans, kettles, and so on, is supplied with energy (by thermal conduction) near the bottom of the container. The heated water rises, to be replaced by cooler water, which in turn will be heated. Convection currents ensure that the thermal energy spreads evenly throughout the water. See Figure B1.19.
- Room heaters are placed near to the floor, but air-conditioners are near the ceiling.
- The coolest part of a refrigerator is near the bottom.
- Water is mixed in the oceans and lakes by convection currents.
- Molten material in the Earth's core circulates because of convection.
- Convection currents occur in the very hot cores of stars, including the Sun.
- Formation and movement of clouds and storms depend on convection.
- The Earth's climate and weather patterns are controlled by convection.
- The direction of winds near coasts depends on convection.
- Smoke usually rises because of convection, but a lack of convection can make air pollution problems worse.

27 Outline how the experiment shown in Figure B1.18 is demonstrating convection in air.

28 Figure B1.20 shows a tube of water being heated near to the water surface. A metal gauze is keeping some ice at the bottom of the tube.



■ **Figure B1.20** Tube of ice being heated

a Explain what this demonstration shows us about the transfer of thermal energy in water.

b Predict and explain how the observations will change if the gauze is removed, allowing the ice to rise, and the water is heated at the bottom of the tube.

29 Convection currents in the air often flow from the sea towards the land. This is because, in the daytime, the land changes temperature quicker, and gets warmer, than the sea. Sketch an annotated diagram to help to explain this phenomenon.

30 Discuss what features of the clothing of the Antarctic explorer seen in Figure B1.21 keep the explorer warm.

■ **Figure B1.21** Antarctic explorer

31 Outline the cooking process (in terms of thermal energy transfers) for the pizza seen in Figure B1.22.

■ **Figure B1.22** Pizza oven

ATL B1A: Thinking skills

Being curious about the natural world

Under certain weather conditions, the normal convection currents which rise from the Earth's surface can be greatly reduced. This can result in trapping air pollutants that would usually disperse. Use the search term 'temperature inversion' to research into this phenomenon and write a 200–300 word summary.

Figure B1.23 The polluting effects of a temperature inversion over Almaty in Kazakhstan

Top tip!

This section on thermal radiation requires some understanding of waves, radiation and spectra, all of which are covered in Theme C. If you have not been introduced to these topics before, it may be better to delay the study of this section (thermal radiation) until after Topics C.2 and C.3 have been studied.

◆ **Emit** To send out from a source.

◆ **Thermal radiation** Electromagnetic radiation emitted because of the movement of charged particles in the atoms of all matter at all temperatures. Most commonly, infrared.

◆ **Infrared** Electromagnetic radiation emitted by all objects (depending on temperature) with wavelengths longer than visible light.

Thermal radiation

SYLLABUS CONTENT

- ▶ Quantitative description of energy transferred by radiation as a result of the emission of electromagnetic waves from the surface of a body, which in the case of a black body can be modelled by the Stefan–Boltzmann law as given by: $L = \sigma A T^4$, where L is the luminosity, A is the surface area and T is the absolute temperature of the body.
- ▶ The emission spectrum of a black body and the determination of the temperature of the body using Wien's law: $\lambda_{\max} T = 2.9 \times 10^{-3}\,\mathrm{mK}$, where $\lambda_{\max}$ is the peak wavelength.

All matter / objects **emit** electromagnetic waves because of the movement of charged particles within their atoms. (There is no need to understand this process.) This is called **thermal radiation**. Electromagnetic waves are explained in Topic C.2. Most commonly this radiation is called **infrared**, but if the temperature is hot enough, visible light is also emitted.

A flame (Figure B1.24) is an obvious example, producing significant amounts of electromagnetic radiation: we can detect the infrared by holding a hand near the flame, and detect the light with our eyes. Figure B1.4 showed the infrared emitted by a saucepan.

Although it is true to say that thermal radiation is emitted continuously by all matter at all temperatures, we tend to only notice it coming from hot objects. The power of the emitted radiation from *any* surface depends on:

1. **Surface temperature** The radiated power is proportional to the *fourth* power of the surface temperature (in kelvin), T^4. This means, for example, a metal bar at 600 K (323 °C) will emit $2^4 = 16$ times as much radiation as the same bar at 300 K (23 °C).
2. **Surface area** The radiated power is proportional to the area, A.
3. **Nature of the surface** See next section.

Note that the emitted power is *not* dependent on the chemical nature of the material.

Figure B1.24 Thermal radiation from a flame

■ **Figure B1.25** Sydney Opera House.

Good absorbers and good emitters of thermal radiation

Dark surfaces, especially black, are the best absorbers of thermal radiation. White and shiny surfaces reflect and scatter radiation well, so that they are poor absorbers. (Scattering is explained later in this topic.) The Sydney Opera House (Figure B1.25) is a poor absorber of thermal radiation. (The word **absorption** describes something taking in something else, a sponge absorbing water, for example.)

Any surface which is a good absorber of radiation will also be a good emitter. Black surfaces emit and absorb radiation well; white surfaces are poor at absorbing and emitting.

◆ **Absorption** When the energy of incident particles or radiation is transferred to other forms within a material.

◆ **Black body** An idealized object that absorbs all the electromagnetic radiation that falls upon it. A perfect black body also emits the maximum possible radiation.

◆ **Black-body radiation (spectrum)** Radiation emitted from a 'perfect' emitter. The characteristic ranges of different radiations emitted at different temperatures are commonly shown in graphs of intensity against wavelength.

Black bodies

A perfect **black body** is the term we use to describe an object which has a surface which absorbs *all* of the infrared and light (and other electromagnetic radiation) that falls on it.

No light is reflected, so we are unable to see a black body, except in outline (unless it is also hot enough to emit visible radiation: light). This is easily defined and understood; however, a perfect black body is also a perfect *emitter* of thermal radiation, but what exactly does that mean? Obviously, it cannot mean that all of the energy in the surface is emitted instantaneously!

All surfaces emit a range of different wavelengths with different powers, and this varies with temperature. A perfect black body emits the maximum possible thermal radiation, and this is best described graphically by a **black-body emission spectrum**, as shown in Figure B1.26, for three different high temperatures. A curve for 300 K (27 °C) would be too small to show on the scale of this graph and it would have its maximum value at a wavelength of about 10×10^{-6} m, which is well off the horizontal scale to the right.

We can see from the graph that, as temperatures increase, more power is emitted and the wavelength at which the maximum power is emitted, λ_{max}, becomes smaller.

At 3000 K, only a small proportion of the emitted radiation is visible light. This proportion increases with temperature and, if the surface it hot enough, some ultraviolet radiation will also be emitted.

If an object is heated (without chemical reactions occurring), a metal bar for example, it will begin to emit visible light (the red end of the spectrum) at about 850 K. If the temperature rises, other colours will be emitted, combining to give the overall effects seen in Figure B1.27.

Note that 'perfect' emitters are called black bodies, but that does not mean that they will always appear black. The Sun has a surface temperature of about 5800 K and it is a good example of a black body, so we may assume that it absorbs all the radiation falling on it; however, it is so hot that it emits enormous quantities of visible light. We are familiar with the visible spectrum, but its full black-body spectrum extends into the infrared and ultraviolet.

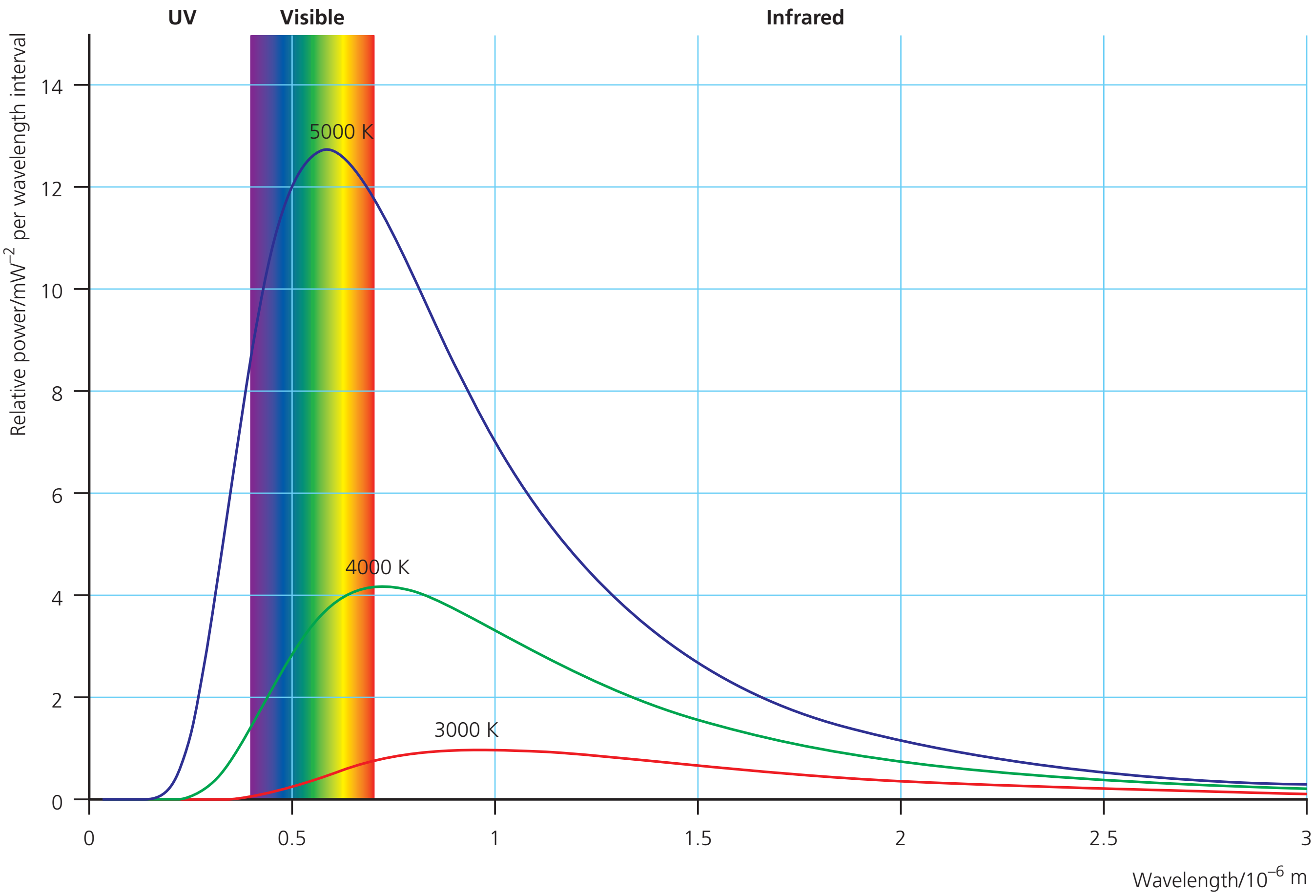


Figure B1.26 Black-body emission spectra at three different temperatures

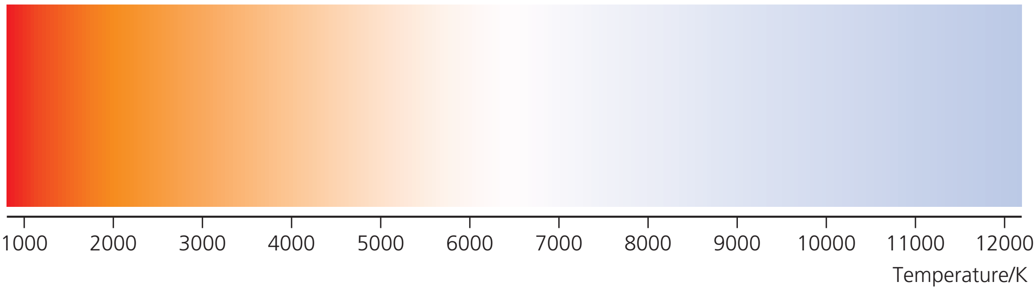


Figure B1.27 Colours of hot surfaces

◆ Stefan–Boltzmann law
An equation that can be used to calculate the total power radiated from the surface of a black body, $P = \sigma AT^4$. σ is known as the **Stefan–Boltzmann constant**.

Since everything emits thermal radiation, all bodies are continuously emitting *and* receiving radiation. In practice, we may assume that one of these is insignificant compared to the other. For example, the radiation absorbed by the Sun is insignificant compared to the energy it emits. However, this is not true when we consider the Earth. See next Topic B.2: Greenhouse effect.

The total power, P, emitted (across all wavelengths) from a perfect black body of surface area A can be calculated from the **Stefan–Boltzmann law**:

power emitted from a black body, $P = \sigma AT^4$

σ is known as the **Stefan–Boltzmann constant**. It has the value of **5.67 × 10^{-8} W m^{-2} K^{-4}**

◆ **Celestial objects** Any naturally occurring objects that can be observed in space.

When referring to **celestial objects** (stars, for example), the emitted power is usually called **luminosity**, as discussed below.

WORKED EXAMPLE B1.4

A metal wire is heated (by an electric current) to a uniform 632 °C. If its length, l, is 80 cm and its radius, r, is 1.6 mm, calculate the total power it radiates into its surroundings.

Assume that it acts as a perfect black body (which is almost true).

Answer

Surface area, $A = 2\pi rl$

$= 2\pi \times 0.0016 \times 0.80 = 8.04 \times 10^{-3}\,\text{m}^2$

$P = A\sigma T^4 = (8.04 \times 10^{-3}) \times (5.67 \times 10^{-8}) \times (632 + 273)^4 = 3.1 \times 10^2\,\text{W}$

Top tip!

We can also use the same equation for the energy *absorbed* by an object from its surroundings:

$$\frac{P}{A} = \sigma T^4$$

If the surrounding temperature is T_s, then the overall (net) radiant energy flow per second from, or to, a black body of area A is: $P = \sigma AT^4 - \sigma AT_s^4$.

For example, using this equation, we can calculate that a black-body surface at 100 °C radiates thermal energy at a rate of 1.1 kW m^{-2}.

At the same time, if the surrounding temperature is 20 °C, it will be receiving energy at a rate of 0.42 kW m^{-2}.

Tool 2: Technology

Use sensors

Infrared scanners and hand-held thermometers (Figure B1.28) have become commonplace in recent times. They detect the thermal radiation emitted by our skins and other surfaces. Their advantages are obvious: they are quick and easy to use, and they do not involve any physical contact. But they have their limitations.

Infrared scanners assume that all skin behaves as a perfect black body. That is, the results from skins of different colours or textures are approximately the same. The radiation coming from the skin is focused onto a detector which effectively determines the power and calculates the corresponding temperature of the emitting surface.

If the distance between the skin and the detector increases, the detector may receive less radiation from each square millimetre but may receive from a greater overall area: it depends on the geometry of the situation.

■ **Figure B1.28** Infrared thermometer

Wien's displacement law

There is a straightforward inverse relationship between surface temperature, T, and the wavelength at which the maximum power is received, λ_{max}.

$$T \propto \frac{1}{\lambda_{max}}$$

This is known as **Wien's displacement law**:

$$\lambda_{max} T = 2.9 \times 10^{-3}\,\text{mK}$$

◆ **Wien's displacement law** Relationship between absolute temperature and the wavelength emitted with maximum power by a black body at that temperature.

WORKED EXAMPLE B1.5

Determine the temperature corresponding to a surface which emits its maximum power with a wavelength of 5.8×10^{-7} m.

Answer

$\lambda_{max} T = 2.9 \times 10^{-3}$

$5.8 \times 10^{-7} \times T = 2.9 \times 10^{-3}$

$T = 5.0 \times 10^{3}$ K This is consistent with Figure B1.26.

■ **Figure B1.29** The Great Lakes in North America appearing dark / black from Space

Although a 'perfect' black body is an idealized concept, the following may approximate to the ideal:

- very hot objects
- dark and dull surfaces
- water
- human skin
- ice
- soil
- vegetation

In Topic B.2, we will introduce the numerical concept of *emissivity*: the ratio of the power radiated per unit area by a surface compared to that of an ideal black surface at the same temperature.

32 Give an everyday example of:
  a a dark surface being good at absorbing thermal energy
  b a dark surface being good at emitting thermal energy
  c a white or shiny surface being good at reflecting thermal energy
  d a white or shiny surface being poor at emitting thermal energy.

33 a Calculate the maximum thermal power radiated away from each square centimetre of a coffee cup which has a surface temperature of 40 °C.
  b Explain why your answer will be an overestimate of the actual power emitted.

34 An object's surface is at 25 °C. Determine the temperature (°C) to which it would have to be heated in order to double the thermal radiation that it radiates.

35 A water storage tank is in sunlight most of the day and its surface reaches a constant temperature of 36 °C. At night the surroundings cool to 23 °C. Estimate the net flow of radiant thermal energy from each square metre of the tank's surface assuming that the surface temperature remains constant.

36 An (unclothed) human body has an average skin temperature of about 35 °C.

- **a** Estimate a value for the area of the skin of a typical adult.
- **b** **i** Show that the power emitted from this area is about 1 kW.
  - **ii** What assumption did you need to make?
- **c** 1 kW is a large radiated power, outline why the body does not cool down quickly. What assumption do you need to make?

37 At what wavelength is the maximum infrared power emitted from your skin?

38 Explain how emergency 'survival blankets', as seen in Figure B1.30, can protect people against dangerous loss of thermal energy. They are made of thin plastic sheets with reflective coatings.

Figure B1.30 Survival blankets for long distance runners

Thermal radiation and stars

SYLLABUS CONTENT

- ▶ The concept of apparent brightness, b.
- ▶ The luminosity of a body as given by: $b = \frac{L}{4\pi d^2}$.

◆ **Luminosity (stellar)** Total power of electromagnetic radiation emitted by a star (SI unit: W).

Luminosity

As explained previously, the power emitted by a celestial body (across all wavelengths) is called its luminosity. The Stefan–Boltzman law can be restated with respect to stars as:

DB

luminosity of a star (or other body), $\boldsymbol{L = \sigma A T^4}$

WORKED EXAMPLE B1.6

The Pole Star (north), Polaris, has a surface area of 8.5×10^{21} m^2 and a surface temperature of 6.0×10^3 K.

- **a** Determine an approximate value for its luminosity.
- **b** Compare its luminosity to that of the Sun (3.8×10^{26} W).

Answer

a $L = \sigma A T^4 \approx (5.67 \times 10^{-8}) \times (8.5 \times 10^{21}) \times (6.0 \times 10^3)^4 \approx 6.2 \times 10^{29}$ W

b $\frac{6.2 \times 10^{29}}{3.8 \times 10^{26}} \approx 1600$

Polaris has a luminosity about 1600 times greater than the Sun.

We can write Polaris's luminosity as $L = 1600\ L_{\odot}$

Luminosity of stars is often given in terms of the luminosity of the Sun, $L_{\odot}$

$L_{\odot} = 3.8 \times 10^{26}$ W

There is no conduction or convection across the near-vacuum of space and the stars can be considered to be perfect black bodies. There is also no significant absorption of thermal energy in space (except where the distances are enormous). All this means that developing a basic understanding of thermal energy transfers across space is fairly straightforward.

All stars (except the Sun) appear as *point sources* of light. The only differences we can see with telescopes are their brightness and slight differences in colour. More than 2000 stars can be seen in Figure B1.31. Some may *appear* larger than others in the picture, but this effect is only because they are brighter.

Figure B1.31 Sagittarius Star Cloud taken from the Hubble telescope

◆ **Star map** Two-dimensional representation of the relative positions of stars as seen from Earth.

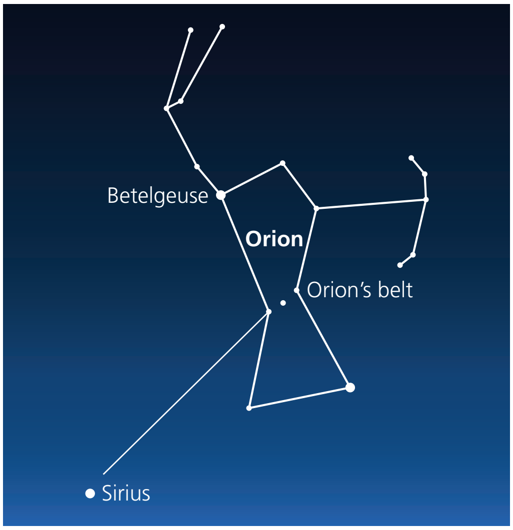


Figure B1.32 Locating *Sirius* in the night sky

There are two possible reasons why one star may appear brighter than another. The brighter star may be emitting more power (more luminous), and/or it may be closer to Earth.

Differences in the colours of stars seen in Figure B1.31 may be attributed to differences in surface temperatures, as explained above.

Sirius is the brightest star in the night sky. Observations of its spectrum show that $\lambda_{max} = 2.92 \times 10^{-7}$ m. Using Wien's law, astronomers can determine that it has a surface temperature of 9930 K. Figure B1.27 confirms that Sirius will appear slightly blue in colour. To locate Sirius in the night sky we can use a **star map**, which highlights groups of stars (constellations). Sirius can be seen close to the constellation of Orion, with its well-known three stars apparently in a line – Orion's belt. See Figure B1.32.

◆ **Apparent brightness, b** Intensity (power / area) of radiation received on Earth from a star (SI unit: $\mathrm{W\,m^{-2}}$).

◆ **Intensity, I** Wave power / area: $I = P/A$ (SI unit: $\mathrm{W\,m^{-2}}$).

Apparent brightness and intensity

We could describe the luminosity of a star as its *actual brightness*, but that is different from what we detect here on Earth, which is called a star's **apparent brightness**. Stars of equal luminosity will have different apparent brightnesses on Earth if they are different distances away. Consider Figure B1.33, which represents three stars of *equal* luminosity at different distances from Earth. If the apparent brightness of star A is b, then apparent brightness of star B is $b/2^2$, while star C has an apparent brightness of $b/3^2$.

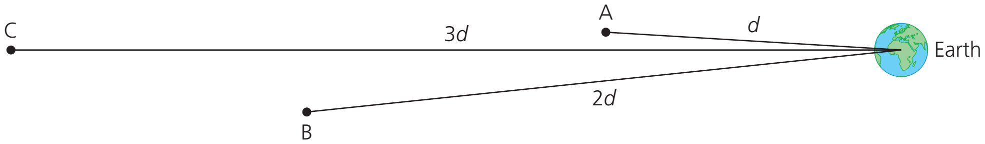


Figure B1.33 Comparing apparent brightnesses

Apparent brightness is a measure of the **intensity** of the radiation from the star which reaches Earth. SI unit: $\mathrm{W\,m^{-2}}$

The intensity, I, of radiation (or waves) is the power, P, transferred through unit area (perpendicular to the direction of energy transfer).

$$\text{intensity} = \frac{\text{power}}{\text{area}} \qquad I = \frac{P}{A} \qquad \text{SI unit: W m}^{-2}$$

WORKED EXAMPLE B1.7

A solar panel close to the Earth's surface has dimensions 3.2 m × 1.6 m. How much total power is received when the intensity of radiation falling perpendicularly on the panel is $739\,\mathrm{W\,m^{-2}}$?

Answer

$$I = \frac{P}{A}$$

$$739 = \frac{P}{(3.2 \times 1.6)} \Rightarrow P = 3.8 \times 10^3\,\mathrm{W}$$

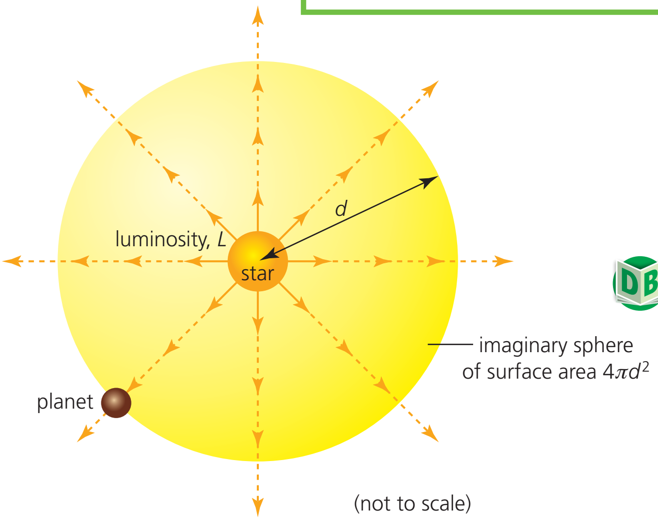


Figure B1.34 A star's radiation spreading out

If we assume that thermal radiation from a star like the Sun spreads out equally in all directions without absorption, then at a distance d from the star, the same total power, L, is passing through an area $4\pi d^2$ (the surface area of an imaginary sphere), as shown in Figure B1.34.

$$\text{apparent brightness: } \boldsymbol{b} = \frac{\boldsymbol{L}}{\boldsymbol{4\pi d^2}}$$

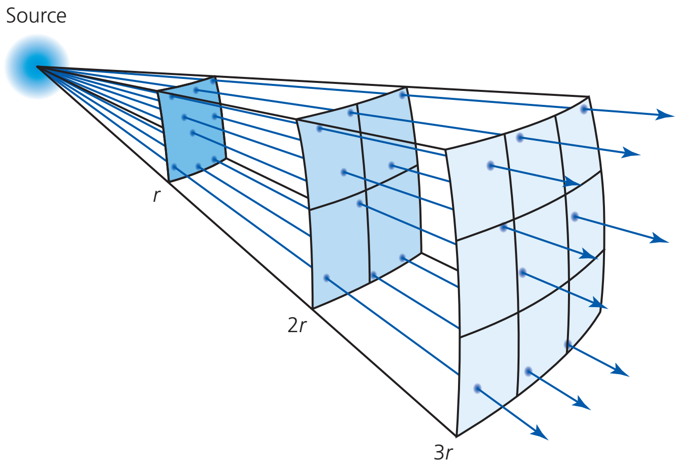
The apparent brightness of the Sun is important information in Topic B.2.

This equation is an example of an **inverse square law** (the apparent brightness is inversely proportional to the distance squared), of which there are several in this course.

◆ **Inverse square law** For waves / energy / particles / fields spreading equally in all directions from a point source without absorption or scattering, the intensity is inversely proportional to the distance squared, $I \propto 1/x^2$ (Ix^2 = constant).

Figure B1.35 shows an alternative visual representation: radiation is spreading out from a point source equally in all directions, *but without absorption*. The same power passes through greater areas as it travels away from the source.

An *inverse* square law represents the fact that a physical quantity is *divided* by 2^2 if the distance from a source is doubled and divided by 3^2 if the distance from a source is trebled, and so on.



■ **Figure B1.35** Radiation spreads to cover four times the area at twice the distance (2*r*) and nine times the area at three times the distance (3*r*)

More generally, intensity: $I \propto \frac{1}{r^2}$ or Ir^2 = constant

Tool 3: Mathematics

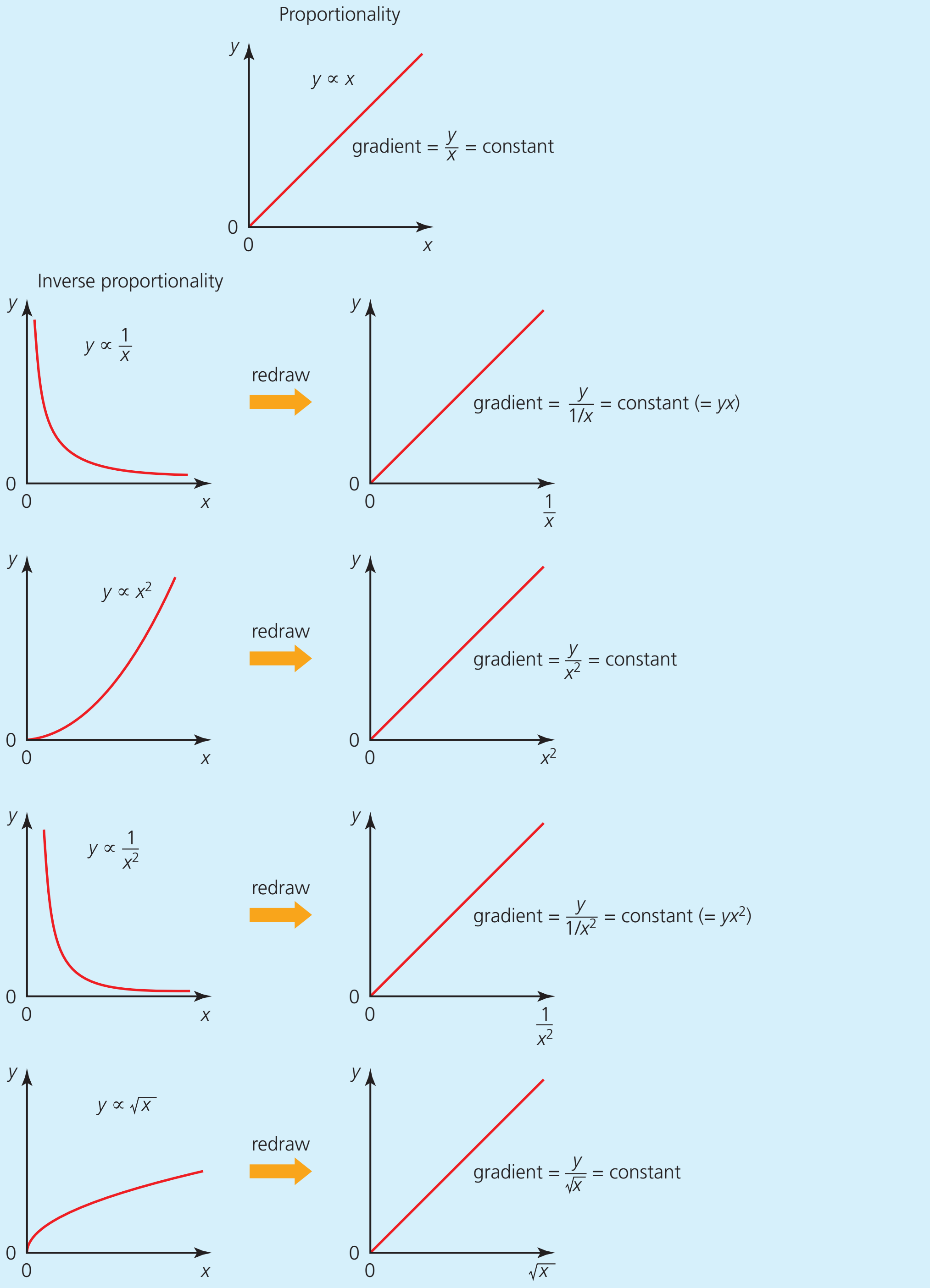
Linearize graphs

Straight lines are much easier to understand and analyse than curved lines, but when raw experimental data are plotted against each other (x and y, for example), the lines are often curved rather than linear.

Data that give an x–y curve can be used to draw other graphs to check different possible relationships. For example:

- A graph of y against x^2 could be drawn to see if a straight line through the origin is obtained, which would confirm that y was proportional to x^2.
- A graph of y against $\frac{1}{x}$ that passed through the origin would confirm that y was proportional to $\frac{1}{x}$. In which case x and y are said to be inversely proportional to each other.
- A graph of y against $\frac{1}{x^2}$ passing through the origin would represent an inverse square relationship.

Figure B1.36 shows graphs of the most common relationships.



■ **Figure B1.36** Some common graphical relationships, showing how curves can be replotted to produce straight lines

LINKING QUESTION

- Where do inverse square relationships appear in other areas of physics?

Forces around point sources in gravitational and electric fields follow inverse square laws (Topics D.1 and D.2).

WORKED EXAMPLE B1.8

The star Betelgeuse has a luminosity of 126 000 $L_\odot$ and an apparent brightness of 1.4×10^{-7} W m^{-2}. Determine its distance from Earth.

Answer

$$b = \frac{L}{4\pi d^2}$$

$$1.4 \times 10^{-7} = \frac{(126\,000 \times 3.8 \times 10^{26})}{4\pi d^2} \Rightarrow d = 5.2 \times 10^{18}\,\text{m}$$

Determining astronomical distances

The distance to 'nearby' stars can be determined from geometrical calculations made with measurements of the apparent locations of the stars at different times of the year (explained in Topic E.5), But this method is not possible with most stars because they are so far away that there is no detectable movement in their apparent locations: most stars remain in *exactly* the same positions as seen on a map of the stars. See Figure B1.32 for an example of part of a star map.

In principle, the equation $b = \frac{L}{4\pi d^2}$ can be used to determine the distance, d, to any star if we measure its apparent brightness, b, but *only if* we know its luminosity, L. However, for most stars we have no direct way of knowing their luminosities.

Fortunately, astronomers have identified a few '**standard candles**'. These are stars which have known luminosities, including a type of supernova and Cepheid variables, as explained below.

◆ **Standard candles**
Term used by astronomers to describe the fact that the distance to a galaxy can be estimated from a knowledge of the luminosity of a certain kind of star within it.

Nature of science: Patterns and trends

Cepheid variable stars, a type of 'standard candle'

In 1908, Henrietta Swan Leavitt (see Figure B1.37) discovered that a certain type of star, called a *Cepheid variable*, had a variable luminosity, the maximum value of which could be determined from the time period of its variation (typically a few days).

Figure B1.37 Henrietta Swan Leavitt

Figure B1.38 shows the graphical relationship between the periods of Cepheid variable stars and their luminosity as multiples of the Sun's luminosity, $L_\odot$.

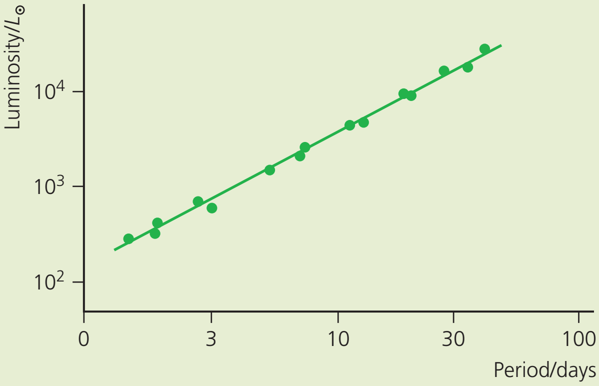


Figure B1.38 Cepheids' variable luminosities

From the graph, for a period of 30 days, the luminosity, $L \approx (1.0 \times 10^4) \times (3.83 \times 10^{26}) \approx 3.8 \times 10^{30}$ W. If a Cepheid variable star has an apparent brightness of 8.4×10^{-8},

$$b = \frac{L}{4\pi d^2}$$

$$8.4 \times 10^{-8} = \frac{3.8 \times 10^{30}}{4\pi d^2} \Rightarrow d \approx 1.9 \times 10^{18}\,\text{m}$$

This distance is approximately 200 light years. A *light year* is the distance travelled by light in one year.

Tool 3: Mathematics

Use units where appropriate: Light year

Astronomical distances are huge. The nearest star to Earth, other than our Sun, (*Alpha Proxima*), is 4.02×10^{16} m away. It becomes convenient to use larger units than metres and kilometres in astronomy. The following are non-SI units.

The light year (ly) is the *distance* travelled by light in one year:

$$\mathbf{1\,ly} = (3.00 \times 10^{8}) \times 365 \times 24 \times 360 = \mathbf{9.46 \times 10^{15}\,m}$$

(A light year is defined to be exactly a distance of 9 460 730 472 580 800 m.)

In light years, the distance to *Alpha Proxima* is 4.25 ly.

The *parsec* is another widely used unit for distance in astronomy (see Topic E.5).

39 The surface area of the Sun is $6.1 \times 10^{18}\,\text{m}^2$ and it has a surface temperature of 5780 K.

- **a** Determine the total thermal power that it emits (luminosity).
- **b** Describe the colour of the visible light emitted by the Sun.

40 The star Betelgeuse, seen in Figure B1.32, has a surface temperature of 3500 K.

- **a** Describe its colour.
- **b** At what wavelength does it emit radiation at the greatest rate?

41 If a star has a luminosity which is 1000 times greater than the Sun and a surface temperature of 15 000 K, predict its surface area compared to the Sun.

42 An LED lamp emits light energy equally in all directions with a total power of 4.1 W.

- **a** Calculate the intensity falling on a book which is 2.34 m away.
- **b** Determine the distance from the lamp where the intensity is $0.40\,\text{W m}^{-2}$.

43 The star Antares is 550 light years from Earth and it has a luminosity of 2.9×10^{31} W.

- **a** Calculate its apparent brightness as viewed from Earth. (Antares is the 15th brightest star in the sky.)
- **b** Its radius is approximately 700× the radius of the Sun (7.0×10^{8} m). Calculate its surface temperature.
- **c** Suggest why this star is described as a 'red (super) giant'. Figure B1.39 compares the size of Antares to the Sun.

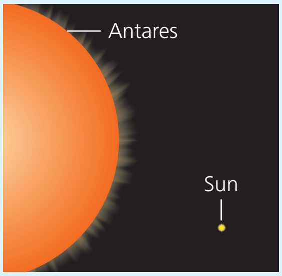


Figure B1.39 The size of Antares compared to the Sun

44 Outline the concept of a 'standard candle', as used to determine the distance to galaxies.

45 A certain type of supernova had a luminosity of 1.4×10^{36} W. If its apparent brightness was $1.9 \times 10^{-6}\,\text{W m}^{-2}$, determine its distance from Earth in light-years.

46 Calculating a distance in answers to questions similar to question 43 assumes that there is no absorption of thermal energy as it travels through space. However, there will certainly be *some* absorption over distances as large as these. Discuss how this could affect the calculated answer to question 43.

Nature of science: Observations

Understanding the Universe

Astronomers have developed an impressive understanding of the Universe, especially with recent technological advances in the detection of remote sources of radiation. Amazingly, all this knowledge has been deduced from thermal and electromagnetic radiation received on Earth, or satellites in orbit.

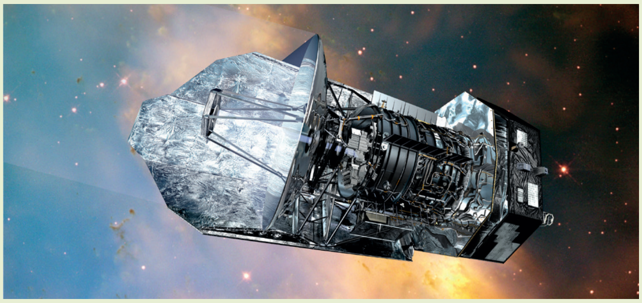


Figure B1.40 Herschel Space Telescope (ESA: 2009–2013)

Heading and cooling

SYLLABUS CONTENT

- Quantitative analysis of thermal energy transfers, Q, with the use of specific heat capacity, c, using: $Q = mc\Delta T$.

The dotted blue straight line in Figure B1.41 shows how the temperature of an object, heated at a constant rate, would change with time under the idealized circumstances of no thermal energy losses to the surroundings. The best way of supplying a constant power is with an electrical heater. The temperature rises by equal amounts in equal times. However, thermal energy losses to the surroundings are unavoidable, so the curved red line represents a more realistic situation. The curve shows that the rate of temperature rise decreases as the object gets hotter. This is because thermal energy losses are higher with larger temperature differences. If energy continues to be supplied, the object will eventually reach a constant temperature when the input power and rate of thermal energy loss to the surroundings are equal (assuming that there are no chemical or physical changes).

When something is left to cool naturally, the rate at which thermal energy is transferred away decreases with time because it also depends on the temperature difference between the object and its surroundings. See Figure B1.42.

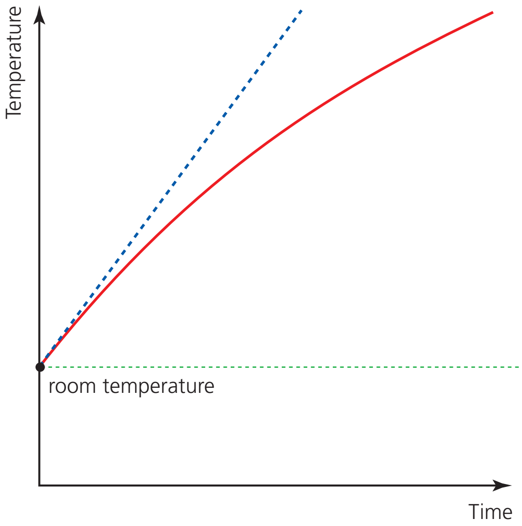


Figure B1.41 A typical graph of temperature against time for heating at a constant rate

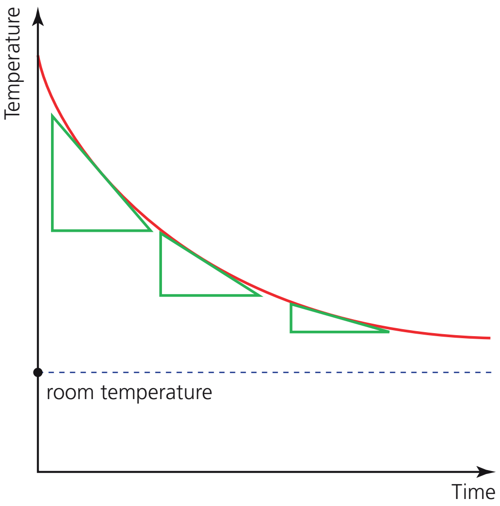


Figure B1.42 A typical graph of temperature against time for an object cooling down naturally to room temperature. Note how the gradient decreases with time

Tool 3: Mathematics

Interpret features of graphs: gradient

Figure B1.42 shows another example of determining rates of change from gradients of a graph. In this case, as can be seen on the figure, the gradients and rates of change are negative and decrease in magnitude with time. Three examples are shown in Figure B1.42, but ideally, larger triangles should be used.

Specific heat capacity

In order to compare how different substances respond to heating, we need to know how much thermal energy will increase the temperature of the same mass (1 kg) of each substance by the same amount (1 K, or 1 °C). This is called the *specific heat capacity*, *c*, of the substance. (The word 'specific' is used here simply to mean that the heat capacity is related to a specified amount of the material, namely 1 kg.)

◆ **Specific heat capacity, *c*** The amount of energy needed to raise the temperature of 1 kg of a substance by 1 K.

The **specific heat capacity** of a substance is the amount of energy needed to raise the temperature of 1 kg of the substance by 1 K. (SI Unit: $\mathrm{J\,kg^{-1}\,K^{-1}}$, but $^\circ\mathrm{C}^{-1}$ can be used instead of K^{-1}.)

The values of specific heat capacity for some common materials are given in Table B1.5.

Table B1.5 Specific heat capacities of some common materials

Material	Specific heat capacity / $\mathrm{J\,kg^{-1}\,K^{-1}}$
copper	390
aluminium	910
water	4180
air	1000
dry earth	1250
glass (typical)	800
concrete (typical)	800
steel	420

Common mistake

The unit for specific heat capacity is very often written incorrectly by students.

Substances with high specific heat capacities heat up slowly compared with equal masses of substances with lower specific heat capacities (given the same power input). Similarly, substances with high specific heat capacities will cool down more slowly. It should be noted that water has an unusually large specific heat capacity. This is why it takes the transfer of a large amount of energy to change the temperature of water and the reason why water is used widely to transfer energy in heating and cooling systems.

If a quantity of thermal energy, Q, was supplied to a mass, m, and produced a temperature rise of ΔT, we could calculate the specific heat capacity from the equation:

$$c = \frac{Q}{m\Delta T}$$

This equation is more usually written as follows:

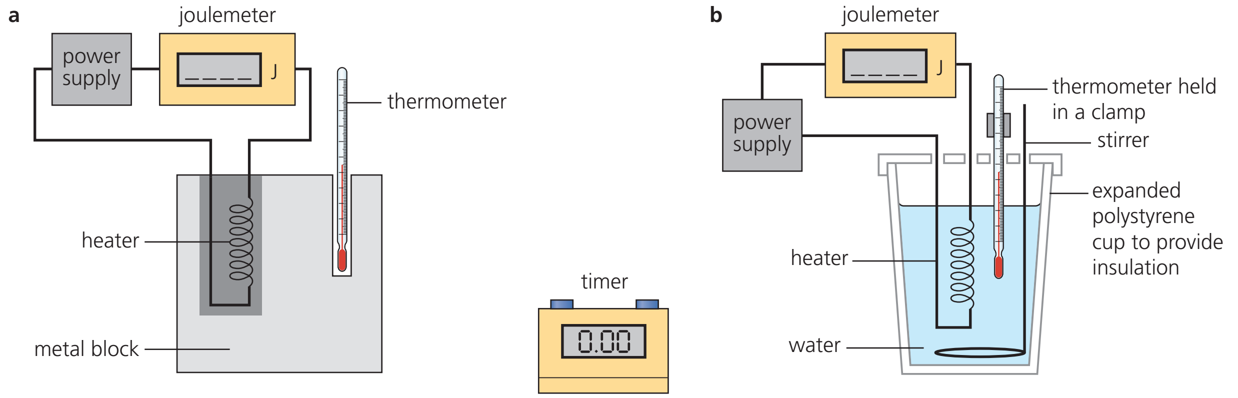
thermal energy transferred, $\boldsymbol{Q = mc\Delta T}$

When a substance cools, the thermal energy transferred away can be calculated using the same equation.

◆ **Immersion heater** Heater placed inside a liquid or object.

Figure B1.43 shows two laboratory experiments to determine specific heat capacities of **a** a metal, and **b** water, or another liquid. The energy is supplied by **immersion heaters** at a constant rate electrically and can be measured directly by a 'joulemeter'. (Alternatively, the energy can be calculated from voltage × current × time. See Topic B.5.)

To use the equation shown above to determine specific heat capacity it is necessary to determine the amount of thermal energy that was transferred to raise the temperature of a known mass by a known amount.



■ **Figure B1.43** Determining the specific heat capacity of **a** a metal, **b** water

WORKED EXAMPLE B1.9

Suppose that in an experiment similar to that shown in Figure B1.43a, a metal block of mass 1500 g was heated for exactly 5 minutes with an 18 W heater. If the temperature of the block rose from 18.0 °C to 27.5 °C, calculate its specific heat capacity, assuming that no energy was transferred to the surroundings.

Answer

$c = \dfrac{Q}{m\Delta t}$ and $Q = Pt$

$$c = \frac{18 \times (5 \times 60)}{[1.5 \times (27.5 - 18.0)]} = 3.8 \times 10^2\,\text{J kg}^{-1}\,°\text{C}^{-1}$$

In making such calculations, it has to be assumed that all of the substance was at same temperature and that the thermometer recorded that temperature accurately at the relevant times. (The thermometer is measuring a temperature that is changing.) In practice, both of these assumptions might lead to significant inaccuracies in the calculated value. Furthermore, in any experiment involving thermal energy transfers and changes in temperature, there will be unavoidable losses (or gains) from the surroundings. If accurate results are required, it will be necessary to use insulation to limit these energy transfers, which in this example would have led to an overestimate of the substance's specific heat capacity (because some of the energy input went to the surroundings rather than into the substance). The process of insulating something usually involves surrounding it with a material that traps air (a poor conductor) and is often called **lagging**.

◆ **Lagging** Thermal insulation.

Inquiry 2: Collecting and processing data

Interpreting results

In an experiment similar to that seen in Figure B1.43b, the temperature of 210 g of water rose from 22.4 °C to 30.7 °C when 7880 J were supplied.

1. Calculate a value for the specific heat of water.
2. Is this an accurate experiment? (Calculate percentage difference from the accepted value.)
3. Estimate the uncertainty in the raw data and then calculate the absolute uncertainty in the processed result.
4. Compare your answers to 2 and 3 and suggest why they are different.

Thermal capacity

◆ **Thermal capacity** The amount of energy needed to raise the temperature of a particular object by 1 kelvin.

Many everyday objects are not made of only one substance, so that referring to a specific amount (a kilogram) of such objects is not useful. In such cases we refer to the **thermal capacity** of the whole object. For example, we might want to know the thermal capacity of a room and its contents when choosing a suitable heater or air conditioner. The thermal capacity of an object is the amount of energy needed to raise its temperature by 1 K. (Unit: J K^{-1} or J °C^{-1})

$$\text{thermal capacity} = \frac{Q}{\Delta T}$$

WORKED EXAMPLE B1.10

How much thermal energy is needed to increase the temperature of a kettle and the water inside it from 23 °C to 77 °C if its thermal capacity is $4800\,\text{J K}^{-1}$?

Answer

$Q = \text{thermal capacity} \times \Delta T = 4800 \times (77 - 23) = 2.6 \times 10^5\,\text{J}$

When answering these questions, assume that no energy was transferred to, or from, the surroundings.

47 Calculate how much energy is needed to raise the temperature of a block of metal of mass 3.87 kg by 54 °C if the metal has a specific heat capacity of $456\,\text{J kg}^{-1}\,\text{K}^{-1}$.

48 Determine the specific heat capacity of a liquid that requires 3840 J to raise the temperature of a mass of 156 g by 18.0 K.

49 A drink of mass 500 g has been poured into a glass of mass 250 g (of specific heat capacity $850\,\text{J kg}^{-1}\,°\text{C}^{-1}$) in a refrigerator. Calculate how much energy must be removed to cool the drink and the glass from 25 °C to 4 °C. (Assume the drink has the same specific heat capacity as water.)

50 A 20 W immersion heater is placed in a 2.0 kg iron block at 24 °C for 12 minutes. Calculate the final temperature. (Specific heat capacity of iron = $444\,\text{J kg}^{-1}\,°\text{C}^{-1}$.)

51 An air conditioner has a cooling power of 1200 W and is located in a room containing 100 kg of air (specific heat capacity $1000\,\text{J kg}^{-1}\,°\text{C}^{-1}$) at 30 °C. Determine the minimum possible temperature after the air conditioner has been switched on for 10 minutes.

52 A water heater for a shower is rated at 9.0 kW. Water at 15 °C flows through it at a rate of 15 kg every 3 minutes.
Predict the temperature of the water in the shower.

53 A burner on a gas cooker raises the temperature of 500 g of water from 24 °C to 80 °C in exactly 2 minutes.
What is the effective average power of the burner?

54 If the thermal capacity of a room and its contents were $3.5 \times 10^5\,\text{J K}^{-1}$, estimate how long it would take a 2.5 kW heater to raise the temperature from 9 °C to 22 °C.

Exchanges of thermal energy

Figure B1.5 showed the temperature–time graphs of two objects, originally at different temperatures, placed in good thermal contact so that thermal energy can be transferred relatively quickly, assuming that the system is insulated from its surroundings. Under these circumstances the thermal energy given out by one object is equal to the thermal energy absorbed by the other object. Exchanges of thermal energy can be used as an alternative means of determining a specific heat capacity, or in the determination of the energy that can be transferred from a food or a fuel.

◆ **Calorimeter** Apparatus designed for (**calorimetry**) experiments investigating thermal energy transfers.

Calorimetry is the name used to describe experiments that try to accurately measure the temperature changes produced by various physical or chemical processes. Energy transfers can then be calculated if the masses and specific heat capacities are known. Calorimetric techniques may involve specially designed pieces of apparatus, called **calorimeters**, which are designed to limit thermal energy transfer to, or from, the surroundings.

WORKED EXAMPLE B1.11

A large metal bolt of mass 26.5 g was heated in an oven until it reached a constant temperature of 312 °C. (See Figure B1.44.)

It was then quickly transferred into 294 g of water initially at 22.1 °C.

The water was stirred and it reached a maximum temperature of 24.7 °C.

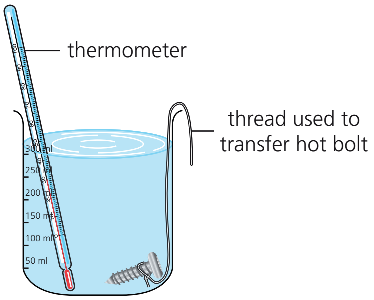


Figure B1.44 A hot metal bolt placed in cold water

a Explain why it was necessary to stir the water.

b Calculate how much thermal energy was transferred from the bolt to the water. (Assume no energy went to the surroundings.)

c Determine a value for the specific heat capacity of the metal from which the bolt was made.

d Suggest which metal the bolt was made from.

Answer

a To make sure that all the water was at the same temperature.

b $Q = mc\Delta T = 0.294 \times 4180 \times (24.7 - 22.1) = 3.20 \times 10^3\,\text{J}$

c Thermal energy transferred to the water = thermal energy transferred from the bolt

$3.20 \times 10^3 = (mc\Delta T)_{\text{bolt}} = 0.0265 \times c \times (312 - 24.7)$

$c = 4.20 \times 10^2\,\text{J kg}^{-1}\,°\text{C}^{-1}$

d Steel (see Table B1.5)

When answering these questions, assume that no energy was transferred to, or from, the surroundings.

55 What mass of water with a temperature of 18 °C has to be mixed with 1.5 kg of water at 83 °C to produce a combined temperature of 55 °C?

56 5 coins, each of mass 8.8 g, were left in some boiling water for a few minutes. They were then very quickly transferred to 98 g of water at 20.7 °C. The water was stirred and its temperature rose to a maximum of 23.6 °C.

a Calculate the specific heat capacity of the metal alloy used in the coins.

b Explain why the coins were transferred quickly.

57 When 5.6 g of wood was completely burned in a calorimeter the temperature of 480 g of water rose from 22.7 °C to 64.6 °C.

a How much thermal energy was transferred to the water from the burning wood?

b Calculate a value for how much energy can be obtained from the combustion of a kilogramme of this wood.

Changes of phase

SYLLABUS CONTENT

- A phase change represents a change in particle behaviour arising from a change in energy at constant temperature.
- Quantitative analysis of thermal energy transfers, Q, with the use of specific latent heat of fusion and vaporization of substances, L, using: $Q = mL$.

A **phase of matter** is a definite region of space in which all the physical and chemical properties of the substance contained in that space are the same. For example, a bottle containing water and oil has two different (chemical) phases of matter. A bottle containing water and ice has two different physical phases of matter.

(The word phase has another, totally different, meaning in physics: as explained in Topic C.1, the phase of an oscillation (or a wave) describes the fraction of an oscillation that has occurred since an agreed reference point.)

Water, ice and steam are commonly described as examples of the three **states of matter**. Generally, in physics, the terms 'phase' and 'state' tend to be used interchangeably.

When thermal energy is transferred to a solid it will usually get hotter. However, for many solid substances, once they reach a certain temperature they will begin to **melt** (change from a solid to a liquid), and while they are melting the temperature does not change (Figure B1.45), even as energy continues to be supplied. This temperature is called the **melting point** of the substance, and it has a fixed value at a particular air pressure (Table B1.6). Melting is an example of a **phase change**. Another word for melting is **fusion**.

◆ **Phase (of matter)** A substance in which all the physical and chemical properties are uniform. In physics, the term **phase change** is used to describe changes between solids, liquids and gases of the same substance.

◆ **States of matter** Solid, liquid or gas (or plasma).

◆ **Melting** Change from a solid to a liquid. Usually at a specific temperature (**melting point**).

◆ **Fusion (thermal)** Melting.

◆ **Freeze** Change from a liquid to a solid. Also called solidify.

◆ **Evaporation** The change from a liquid to a gas (vapour) at any temperature below the boiling point of the liquid. Occurs only at the liquid surface.

◆ **Vaporization** Change from a liquid to a vapour (gas) by boiling or evaporation. A **vapour** is a gas which can be condensed by pressure.

◆ **Boiling** Change from a liquid to a gas / vapour throughout the liquid at a precise temperature.

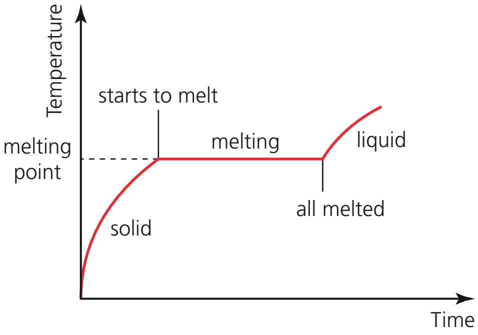


Figure B1.45 Temperature changes as a solid is heated and melted (note that the lines are curved only because of energy transferred to the surroundings)

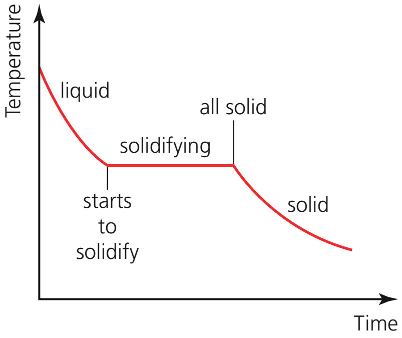


Figure B1.46 Temperature changes when a liquid cools and freezes (solidifies)

Similarly, when a liquid cools, its temperature will be constant at its melting point while it changes phase from a liquid to a solid (Figure B1.46). This process is known as solidifying or **freezing**. But be careful – the word 'freezing' suggests that this happens at a low temperature, but this is not necessarily true (unless we are referring to turning water into ice, for example). The phase changes of water are such common events in everyday life that we all tend to think of them as the obvious examples but, of course, many other substances can melt and freeze. For example, the freezing (melting) point of chocolate is variable but is approximately 30 °C to 40 °C, so many chocolates, but not all, will melt when in contact with skin, see Figure B1.47.

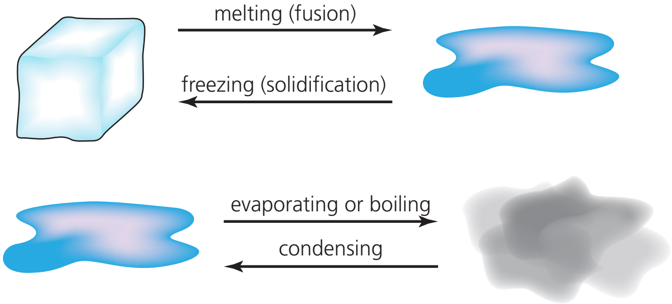
A change of phase also occurs when a liquid becomes a gas (or vapour), or when a gas (or vapour) becomes a liquid (Figure B1.48). A **vapour** is any gas at a temperature such that it can be condensed by pressure alone. The change of phase from a gas (or vapour) to a liquid can be by **boiling** or **evaporation** (it may also be called **vaporization**). Changing from a gas (or vapour) to a

◆ **Condense** Change from a gas or vapour to a liquid.

liquid is called **condensation**. The temperature at which boiling occurs is called the boiling point of the substance, and it has a fixed value for a particular air pressure (see Table B1.6).

The shape of graph showing the temperature change of a liquid being heated to boiling will look very similar to that for a solid melting (Figure B1.45), while the graph for a gas being cooled will look very similar to that for a liquid freezing (Figure B1.46).

■ **Figure B1.47** Melting chocolate



■ **Figure B1.48** Changes of phase

■ **Table B1.6** Melting points and boiling points of some substances (at normal atmospheric pressure)

	Melting point		Boiling point	
Substance	**°C**	**K**	**°C**	**K**
water	0	273	100	373
mercury	−39	234	357	630
alcohol (ethanol)	−117	156	78	351
oxygen	−219	54	−183	90
copper	1083	1356	2580	2853
iron	1538	1811	2750	3023

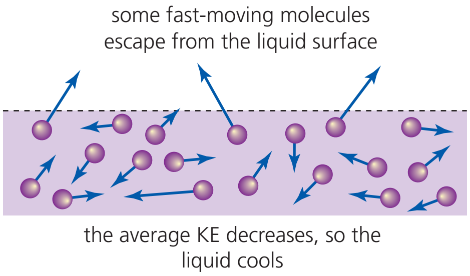
LINKING QUESTION

- How can the phase change of water be used in the process of electricity generation?

This question links to understandings in Topic D.4.

Melting, freezing, boiling and condensing are known as **phase changes**.

Boiling and evaporation



■ **Figure B1.49** Molecules leaving a surface during evaporation

In a liquid the molecules will always have a range of different random kinetic energies that are continuously transferred in interactions / collisions between them. This means there will always be some molecules near the surface that have enough energy to overcome the attractive forces that hold the molecules together in the liquid. Such molecules can escape from the surface and this effect is called *evaporation*. See Figure B1.49.

Evaporation occurs only from the surface of a liquid and can occur at any temperature, although the rate of evaporation increases significantly with rising temperature (between the melting and boiling points).

Boiling occurs at a precise temperature – the temperature at which the molecules have enough kinetic energy to form bubbles *inside* the liquid. Boiling points can vary considerably with different surrounding air pressures.

Evaporation occurs from the surface of a liquid over a range of temperatures. Boiling occurs throughout a liquid at a precise temperature.

LINKING QUESTION

- What role does the molecular model play in understanding other areas of physics?

This question links to understandings in Topics B.3 and B.4.

The loss of the most energetic molecules during evaporation means that the average kinetic energy of the molecules remaining in the liquid must decrease (until thermal energy flows in from the surroundings). This microscopic effect explains the macroscopic fall in temperature (cooling) that always accompanies evaporation from a liquid.

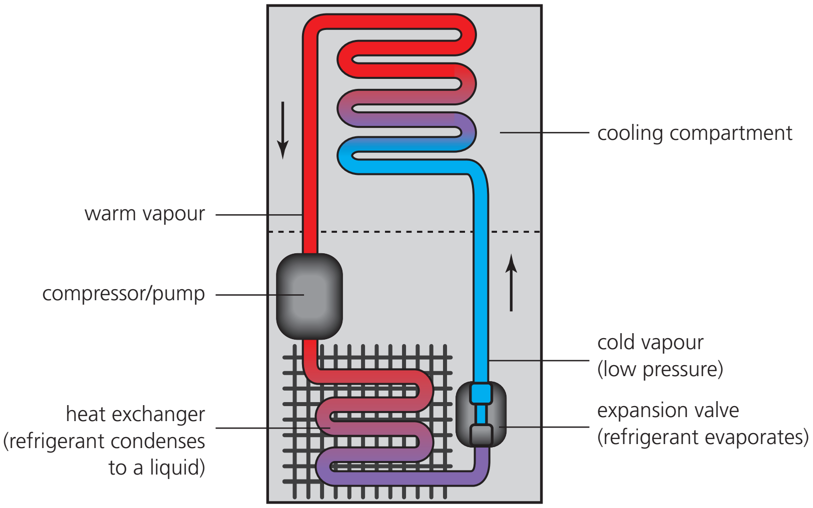
Using the cooling effect of evaporation

Sweating – Our bodies produce tiny droplets of water on our skin. See Figure B1.52. When the water evaporates it removes thermal energy from our bodies and helps to keep us cool.

Cooling buildings – The cooling effect produced by water evaporating has been used for thousands of years to keep people and buildings cool. For example, in central Asia open towers in buildings encouraged air flow over open pools of water, increasing the rate of evaporation and the transfer of thermal energy up the tower by convection currents. The flow of air past people in buildings also encourages the human body's natural process of cooling by sweating.

◆ **Refrigerant** Fluid used in the refrigeration cycle of refrigerators, air conditioners and heat pumps.

Refrigerators – Modern refrigerators rely on the cooling produced when a liquid evaporates. The liquid/gas used is called the **refrigerant**. Ideally it should take a large amount of thermal energy to turn the refrigerant from a liquid into a dense gas at a little below the desired temperature. In a refrigerator, for example, after the refrigerant has removed thermal energy from the food compartment, it will have become a gas and be hotter. In order to re-use it and turn it back into a cooler liquid again, it must be compressed and its temperature reduced. To help achieve this thermal energy is transferred from the hot, gaseous refrigerant to the outside of the refrigerator (Figure B1.50).



■ **Figure B1.50** Schematic diagram of a refrigerator

Air-conditioners use the same principle as refrigerators.

Outdoor misting systems – Outdoor misting systems (see Figure B1.51) are becoming increasingly popular. Tiny water droplets are sprayed out of nozzles and quickly evaporate, cooling the air, or people on whom the droplets fall.

◆ **Humidity** A measure of the amount of water vapour present in air.

■ **Figure B1.51** Cooling mists in a restaurant

ATL B1B: Communication skills

Clearly communicate complex ideas in response to open-ended questions

Humidity and fans

The cooling effect of sweating on the human body is greatly affected by the **humidity** of the surrounding air.

Air contains unseen water **vapour** that has evaporated from plants and various water surfaces. At 20 °C each cubic metre of air can contain up to a maximum of 17 g of invisible water vapour. As the temperature increases, each cubic meter of air can contain more water. For example, at 30 °C the maximum is 30 g.

Humidity is a measure of the amount of water vapour in air compared to the maximum possible. For example, at 20 °C if there is 17 g m^{-3}, the humidity is said to be 100%, but, if there is 8.5 g m^{-3}, the humidity is 50%, which is often reported to be about the most comfortable humidity for people.

With greater humidity in the surrounding air, it is more difficult for water to be evaporated from the skin in the process of sweating, and so the cooling effects are reduced.

■ **Figure B1.52** Sweat cools us by evaporation

Fans can be very useful in helping to keep people cool, but they do not directly reduce temperatures.

Research and explain in your own words how a fan might help to keep someone cool.

58 Suggest why the boiling point of a liquid depends on the surrounding air pressure.

59 Some spaghetti is being cooked in boiling water in an open pan on a gas cooker. Discuss what happens if the gas flow is increased so that more thermal energy is transferred to the water.

60 Explain why wet clothes will dry more quickly outside on a windy day.

61 Why will water spilt on the floor dry more quickly if it is spread out?

Latent heat

◆ **Latent heat** Thermal energy that is transferred at constant temperature during any change of physical phase.

To melt a solid, or boil a liquid, it is necessary to transfer thermal energy to the substance. However, as we have seen, melting and boiling occur at constant temperatures, so that the energy supplied is not being used to increase molecular kinetic energies (which change with temperature). Because there is no change of temperature, the thermal energy transferred during a phase change is called **latent heat** (*latent* means hidden).

The latent heat supplied is used to produce the molecular re-arrangements that characterize the differences between solids and liquids, and liquids and gases. Latent heat is used to overcome intermolecular forces and to increase molecular separations. This will increase molecular potential energies. In the case of melting, some forces are overcome and there is a slight increase in average separation, but in the case of boiling all the remaining forces are overcome as the molecules move much further apart.

When a liquid freezes (solidifies), the same amount of energy per kilogram is emitted as was needed to melt it (without a change in temperature). Similarly, boiling and condensing involve equal energy transfers.

The concept of *specific* latent heat brings a mathematical treatment to this subject:

The specific latent heat of a substance, L, is the amount of energy transferred when 1 kilogram of the substance changes phase at a constant temperature. $\boldsymbol{Q = mL}$ SI units: $\mathrm{J\,kg^{-1}}$

◆ **Specific latent heat, L_f or L_v** The amount of energy needed to melt (fusion) or vaporize 1 kg of a substance at constant temperature.

The latent heat associated with melting or freezing is called **specific latent heat of fusion**, L_f. The latent heat associated with boiling or condensing is known as **specific latent heat of vaporization**, L_v.

As an example, the specific latent heat of fusion of lead is $2.45 \times 10^4\,\mathrm{J\,kg^{-1}}$ and its melting point is 327 °C. This means that 2.45×10^4 J is needed to melt 1 kg of lead at a constant temperature of 327 °C.

Experiments to determine specific latent heats (water is often used as a convenient example) have many similarities with specific heat capacity experiments. Usually, an electric heater of known power is used to melt or boil a substance – Question 65 overleaf describes such an experiment.

WORKED EXAMPLE B1.13

The latent heats of vaporization of water and ethanol are $2.27 \times 10^6\,\mathrm{J\,kg^{-1}}$ and $8.55 \times 10^5\,\mathrm{J\,kg^{-1}}$.

a State and explain which one is 'easier' to boil (at the same pressure).

b Calculate how much thermal energy is needed to turn 50 g of ethanol into a gas at its boiling point of 78 °C.

Answer

a It is 'easier' to boil ethanol because less energy is needed to turn each kilogram into a gas.

b $Q = mL = 0.050 \times (8.55 \times 10^5) = 4.3 \times 10^4\,\mathrm{J}$

62 Some water was heated in an open 2250 W electric kettle. When it reached 100 °C the water boiled and in the next 180 s the mass of water reduced from 987 g to 829 g.

a Use these figures to estimate the latent heat of vaporization of water.

b Explain why your answer is only an estimate. Is it an underestimate, or an overestimate?

63 The latent heat of fusion of a certain kind of chocolate is $1.6 \times 10^5\,\text{J kg}^{-1}$. Predict how much thermal energy is removed from you when a 10 g bar of chocolate melts in your mouth.

64 Outline why you would expect that the latent heats of vaporization of substances are larger than their latent heats of fusion.

65 The apparatus shown in Figure B1.53 was used to determine the specific latent heat of fusion of ice. Two identical 50 W immersion heaters were placed in some ice in two separate funnels. The heater above beaker A was switched on, but the heater above B was left off. After 5 minutes it was noted that the mass of melted ice in beaker A was 54.7 g, while the mass in beaker B was 16.8 g.

a Explain the reason for having ice in two funnels.

b Use these figures to estimate the latent heat of fusion of ice.

c Suggest a reason why this experiment does not provide an accurate result.

d Describe one change to the experiment that would improve its accuracy.

66 0.53 g of steam at 100 °C condensed and then the water rapidly cooled to 35 °C.

a How much thermal energy was transferred from the steam:

i when it condensed

ii when the water cooled down?

b Suggest why a burn received from steam is much worse than from water at the same temperature (100 °C).

67 120 g of water at 23.5 °C was poured into a plastic tray for making ice cubes. If the tray was already at 0 °C, calculate the thermal energy that has to be removed from the water to turn it to ice at 0 °C.
(The latent heat of fusion of water is $3.35 \times 10^5\,\text{J kg}^{-1}$.)

68 Clouds are condensed droplets of water and sometimes they freeze to become ice particles. Suppose a typical cloud had a mass of 24 000 kg:

a Determine how much thermal energy would be released if it all turned to ice at 0 °C.

b Discuss how your answer compares to a typical value of 5×10^9 J of energy released in a single lightning strike.

69 Some water and a glass container are both at a temperature of 23 °C and they have a combined thermal capacity of $1500\,\text{J K}^{-1}$. If a 48 g lump of ice at −8.5 °C is placed in the water and the mixture is stirred until all the ice has melted, determine the final temperature.
(The specific heat capacity of ice is $2.1 \times 10^3\,\text{J kg}^{-1}\,\text{K}^{-1}$.
The latent heat of fusion of water is $3.35 \times 10^5\,\text{J kg}^{-1}$.)

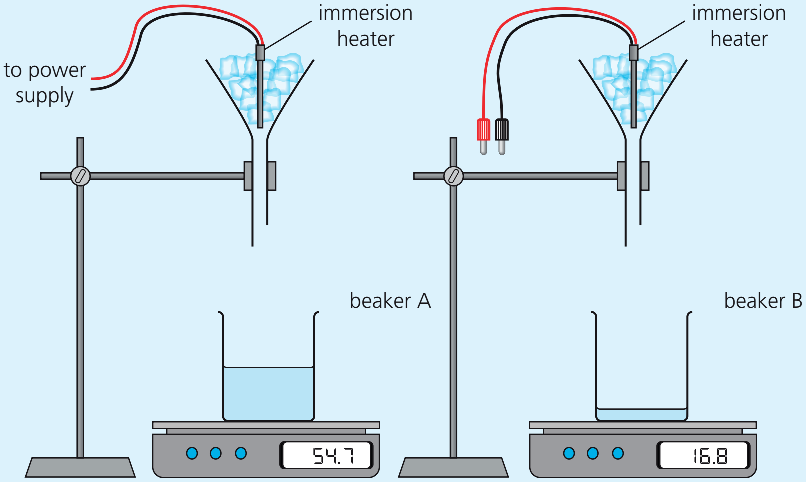


Figure B1.53 An experiment to determine the latent heat of fusion of ice

B.2 Greenhouse effect

Guiding questions

- How does the greenhouse effect help to maintain life on Earth and how does human activity enhance this effect?
- How is the atmosphere as a system modelled to quantify the Earth–atmosphere energy balance?

We use the term **greenhouse effect** to describe the fact that the Earth's atmosphere keeps the planet warmer than it would be without the atmosphere. The effect of the Earth's atmosphere is similar in some ways to how the glass in the walls and roof of a greenhouse keep the plants warmer than if they were left in the open air.

From the beginning, it is important to understand that the basic greenhouse effect is essential for life on Earth. However, human activity has changed, and continues to change, the atmosphere in ways that are making the Earth warmer. This is known as the ***enhanced* greenhouse effect**, or **anthropogenic climate change**.

◆ **Greenhouse effect** The natural effect that a planet's atmosphere has on reducing the amount of radiation emitted into space, resulting in a planet warmer than it would be without an atmosphere.

◆ **Greenhouse effect (enhanced)** The reduction in radiation emitted into space from Earth due to an increasing concentration of *greenhouse gases* in the atmosphere (especially carbon dioxide) caused by human activities; believed by most scientists to be the cause of global warming.

◆ **Anthropogenic climate change** Changes in the climate due to human activities. Also called global warming.

◆ **Solar System** The Sun and all the objects that orbit around it.

A planet's energy balance: an introduction

SYLLABUS CONTENT

- Conservation of energy.

The planets of the **Solar System**, including the Earth, have existed for billions of years, so it is reasonable to assume that they should each have reached a steady (average) temperature, over human timescales at least. This means that each planet should be receiving and emitting thermal energy at the same rate, assuming that there are no significant internal energy sources of its own.

Since the only thermal energy which can travel across Space is radiation, for a planet at constant average temperature,

radiant thermal energy received by a planet (or moon) = radiant thermal energy emitted by a planet (or moon) in the same time interval.

Figure B2.1 represents this energy balance for the Earth.

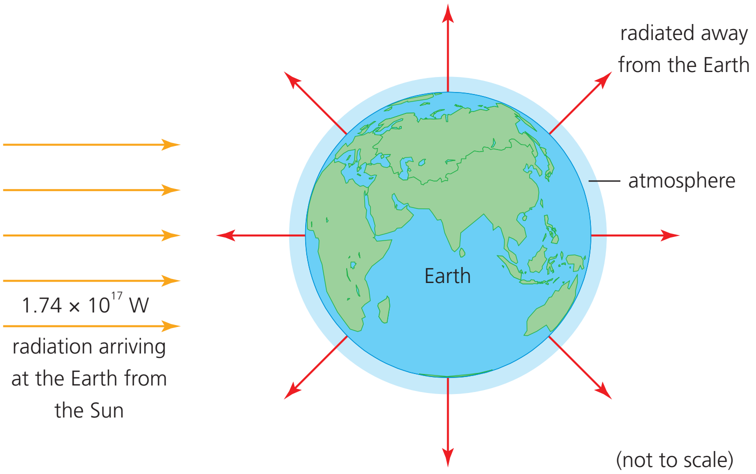


Figure B2.1 Earth receiving and emitting radiation

◆ **Simulation** Simplified visualization (imitation) of a real physical system and how it changes with time. Usually part of a computer modelling process.

In the rest of this topic, we will discuss the various factors which affect this energy balance, and how human activity is upsetting this balance on planet Earth.

We will begin by developing an understanding of the thermal energy (including light) arriving at the Earth and other planets from the Sun.

Tool 2: Technology

Identify and extract data from databases

A database is information (often extensive) which is stored electronically in a structured and organized way. A database will usually be continually updated with new data. There are many different types and sizes of database, which may be accessed in different ways, which might be only available to certain people (information within a school, for example).

Google's 'Bigtable' is an example of an enormous database which helps to run internet searches, Google Maps and so on. At the other extreme, you may wish to set up and monitor your own physical fitness database on a *Microsoft Excel* spreadsheet.

There are a few areas of study in this course for which enormous quantities of data are readily available, including climate change, energy resources and astronomy.

Generate data from models and simulations

Many situations in physics can be reduced to simplified mathematical models. These provide the essential basis for understanding, but they can later be expanded to include more details.

Many computer **simulations** are available to visually represent these models. For example, the movement of a mass bouncing up and down on the end of a spring (Topic C.1). These simulations can be very useful in the learning process, especially when you can investigate the effects of changing the variables (for example, mass on spring, stiffness of spring, air resistance and so on).

While virtual experiments like these should not replace actual experimental work, they are guaranteed to quickly produce results which are consistent with the physics theory and they enable a wider range of tests to be carried out than would usually be done in a laboratory.

As we shall see, the word equation highlighted on page 215 is the starting assumption for a mathematical analysis of the Earth's surface temperature. Without too much difficulty, we will be able to predict the average surface temperatures of planets and moons with reasonable accuracy. But, as we are all now aware, relatively small changes in the Earth's temperature can have disastrous effects. A simple model is inadequate for making predictions about how the Earth's temperature and climate may change.

The factors affecting climate change are numerous, complicated and interconnected. Computer models are needed in order to cope with this complexity and the vast amount of data available.

Luminosity and apparent brightness of the Sun: the solar constant

SYLLABUS CONTENT

- The solar constant, *S*.
- The incoming radiative power is dependent on the projected surface of a planet along the direction of the path of the rays, resulting in a mean value of the incoming intensity being *S*/4.

◆ **Solar constant** Intensity of the Sun's radiation arriving perpendicularly on an area above the Earth's atmosphere.

Using Wien's law (Topic B.1), we can determine that the surface temperature of the Sun is 5780 K. Geometrical measurements made from Earth inform us that the Sun is an average distance of 1.50×10^{11} m away, and it has a surface area of $6.05 \times 10^{18}\,\text{m}^2$. With this information we can determine its luminosity, $L_\odot$ (as defined in Topic B.1), assuming that its acts as a black body:

$$L_\odot = \sigma A T^4 = (5.67 \times 10^{-8}) \times (6.05 \times 10^{18}) \times 5780^4 = 3.83 \times 10^{26}\,\text{W}$$

Knowing its luminosity, we can then calculate the apparent brightness, *b*, of the Sun just above the Earth's atmosphere. See Figure B2.2.

$$b = \frac{L_\odot}{4\pi d^2} = \frac{3.83 \times 10^{26}}{4\pi \times (1.50 \times 10^{11})^2} = 1.35 \times 10^3\,\text{W m}^{-2}$$

The value of the intensity of thermal radiation (including light) from the Sun which passes perpendicularly through an area just above the Earth's atmosphere, is called the **solar constant**, *S*.

The accepted average value for the solar constant of the Earth is $\mathbf{1.36 \times 10^3\,W\,m^{-2}}$

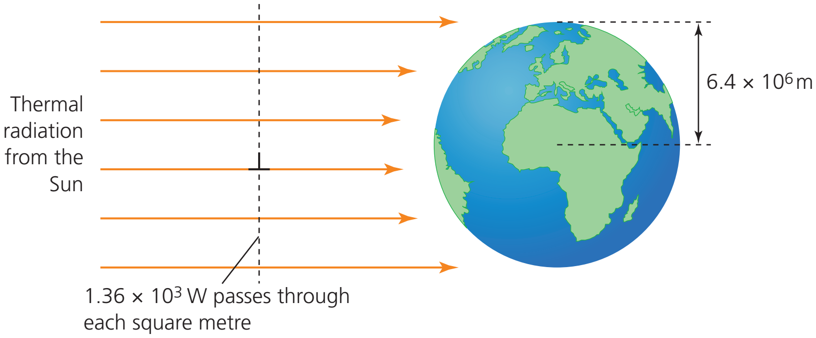


Figure B2.2 Solar constant

Although *S* is called the solar *constant*, it is well-known to vary very slightly, by about 0.1% every 11 years. However, it is not believed that this has a significant effect on the Earth's climate.

Different planets will have different solar constants. For example, Venus has a higher solar constant, $2.6 \times 10^3\,\text{W m}^{-2}$, because it is closer to the Sun. Solar constant values change slightly with periodic changes in the behaviour of the Sun and variations in the distances of planets from the Sun.

This constant will play an important part in calculations concerning the greenhouse effect, later in this topic.

If a planet has a radius *r*, the incoming radiative intensity extends over a cross-sectional area of πr^2, but the whole planet has a surface area of $4\pi r^2$ (the surface area of a sphere). This means that:

the mean value of the radiative intensity directed at the surface of the planet is: $S \times \left(\frac{\pi r^2}{4\pi r^2}\right) = \frac{S}{4}$

In the case of the Earth:

$$\frac{S}{4} = \frac{1360}{4} = 340\,\mathrm{W\,m^{-2}}$$

This would be the average intensity of thermal radiation reaching the Earth's surface if it did not have an atmosphere.

Non-black bodies: albedo and emissivity

SYLLABUS CONTENT

- Albedo as a measure of the average energy reflected off a macroscopic system as given by:
  $$\text{albedo} = \frac{\text{total scattered power}}{\text{total incident power}}$$
- Earth's albedo varies daily and is dependent on cloud formations and latitude.
- Emissivity as the ratio of the power radiated per unit area by a surface compared to that of an ideal black surface at the same temperature as given by: $\text{emissivity} = \frac{\text{power radiated per unit area}}{\sigma T^4}$

Before we can return to a discussion of the energy balances of planets (and moons), we have to consider how the surfaces of these bodies compare to the idealized concept of the surfaces of black bodies.

Although we can assume that the Sun, with its high surface temperature, behaves as a perfect black body, we cannot make the same assumptions for the Earth, or other planets and moons. We need to quantify the ability of non-black-body surfaces to:

- absorb, reflect and scatter thermal radiation; and
- emit thermal radiation.

To do this, we will introduce the two concepts of **albedo** and **emissivity**.

◆ **Albedo** The total scattered or reflected power/total incident power (on part of a planet's surface). Albedo depends on the nature of the surface and inclination of the radiation to the surface.

◆ **Incident wave, or ray** Wave (or ray) arriving at an object or a boundary.

Albedo

The surface of any planet or moon will reflect / scatter some of the thermal radiation that is **incident** upon it (arrives at the surface). We use the term albedo to quantify this:

$$\textbf{albedo} = \frac{\textbf{total scattered power}}{\textbf{total incident power}} \quad \text{(A ratio, so no units.)}$$

Scattering can be considered to be unpredictable and small-scale random reflections. For example, a plane mirror reflects light so that images may be seen in it, but a mirror broken into many small pieces scatters light. A very smooth surface may reflect light, but a rough surface scatters light.

◆ **Scattering** Irregular reflections of waves or particles from their original path by interactions with matter.

A perfect black body would have an albedo of 0: all incident thermal energy would be absorbed. An albedo of 1 would represent a surface which scatters all of the incident thermal radiation.

Table B2.1 lists approximate values for the albedos of the surfaces of some common materials, but there is a lot of variation from the surfaces of each material. The albedo of a surface also varies with the angle at which the incident radiation strikes the surface: albedo increases as the radiation is incident at greater angles to the perpendicular. This means that, at any particular location, there are variations during each day and at different times of the year. For the same reason, albedo will vary with different latitudes around the Earth.

Table B2.1 Approximate albedo values

Material	Albedo
(ocean) water	0.1
forest	0.1
road surface	0.1
soil	0.2
grass	0.3
desert sand	0.4
clouds (very variable)	0.5
ocean ice	0.6
snow	0.8

Figure B2.3 Snow has high albedo, but water has low albedo

The Earth (including its atmosphere) has an average albedo of 0.315. That is, about 68.5% of the incident thermal radiation is absorbed by the Earth and its atmosphere.

WORKED EXAMPLE B2.1

Radiation of intensity 610 W m^{-2} was incident perpendicularly on the surface of a lake at midday.

a If the water had an albedo of 0.18 at that time, calculate how much energy was absorbed every second by each square metre.

b Describe how your answer would change much later in the day

Answer

a 610 × (1.0 − 0.18) = 500 W (110 W was reflected)

b Later in the day the radiation will be incident at a greater angle to the perpendicular. The albedo will increase and more radiation is reflected. The incident intensity will also decrease because the radiation passes through a greater length of the atmosphere. The answer to **a** will decrease.

◆ **Emissivity** The power radiated by an object divided by the power radiated from a black body of the same surface area and temperature.

Emissivity

The concept of **emissivity** compares the power of the thermal radiation emitted by a surface (from unit area) to that of a perfect black body at the same temperature (σT^4). A surface with a greater emissivity emits thermal energy with more power, under the same conditions.

$$\textbf{emissivity} = \frac{\textbf{power radiated per unit area}}{\sigma T^4}$$

Emissivity is a ratio, so it has no unit. The symbol e (or ε) is sometimes used for emissivity.

Table B2.2 lists approximate values for the emissivity of the surfaces of the same materials as seen in Table B2.1.

The average emissivity of the Earth and its atmosphere is estimated to be 0.61.

A black body has an emissivity of one.

Table B2.2 Typical values for the emissivity of materials also seen in Table B2.1

Material	Emissivity
(ocean) water	0.99
forest	0.97
road surface	0.96
soil	0.95
grass	0.91
desert sand	0.90
clouds (variable)	0.55
ocean ice	0.97
snow	0.94

WORKED EXAMPLE B2.2

Determine the average emissivity of a planet if it has an average surface temperature of −63 °C, a surface area of $1.4 \times 10^{14}\,\text{m}^2$ and it emits thermal energy at a rate of 1.3×10^{16} W.

Answer

$$\text{emissivity} = \frac{\text{power radiated per unit area}}{\sigma T^4}$$

$$= \frac{(1.3 \times 10^{16}/1.4 \times 10^{14})}{5.67 \times 10^{-8} \times (273 - 63)^4}$$

$$= 0.84$$

1 580 W of thermal radiation is incident perpendicularly on photovoltaic solar panels of dimensions 1.86 × 2.12 m, similar to those seen in Figure B2.4

- **a** Calculate the intensity of the radiation.
- **b** If the surface of the panels has an albedo of 0.21, determine the total rate at which the panels absorb thermal radiation.
- **c** Describe how the albedo of the panels is minimized.
- **d** State two reasons why it is best if the panels are perpendicular to the incident radiation.

Figure B2.4 Photovoltaic panels

2
- **a** State why it would be reasonable to expect that, if the temperature of the Earth's surface were to rise, snow and ice would melt quicker.
- **b** Conversely, explain why it would also be reasonable to expect that, if large amounts of snow and ice were to melt, the temperature of the Earth's surface would rise.

3 Explain why the average albedo of ocean water will tend to increase

- **a** in winter
- **b** closer to the poles.

4 Discuss why the emissivity of clouds make them an important factor affecting the average emissivity of Earth.

5 A brick wall is 2.34 m high and 3.80 m long. The bricks have an emissivity of 0.72. At what rate is thermal energy radiated away from the wall if its temperature is 18 °C?

6 Show that the total power radiated away from the Earth's surface is approximately 1×10^{17} W. (Assume surface temperature is 288 K. Radius of the Earth is 6.4×10^6 m.)

Modelling a planet's energy balance

Having established an understanding of emissivity and albedo, we can now return to the subject of the energy balances of planets and moons.

We will take planet Earth as our first and most obvious example, but similar calculations are possible for other planets and moons.

Accurately modelling the Earth's temperature is a very complex process, which, because of its enormous consequences, has preoccupied some of the best scientific minds and fastest computers, using enormous quantities of data, for decades. However, we can use the physics already discussed to make a broad prediction:

We have the following relevant data:

- Solar constant = $1.36 \times 10^3\,\mathrm{W\,m^{-2}}$
- Radius of the Earth = $6.4 \times 10^6\,\mathrm{m}$
- Average emissivity of the Earth (including its atmosphere) = 0.61
- Average albedo of the Earth (including its atmosphere) = 0.315

As explained near the start of this topic:

Radiant thermal energy received by a planet = radiant thermal energy emitted by a planet (in the same time interval).

We can now be more detailed:

solar constant × cross-sectional area of planet × (1 – albedo)
= emissivity × σ × surface area of planet × T^4

$$1.36 \times 10^3 \times \cancel{\pi r^2} \times (1 - 0.315) = 0.61 \times (5.67 \times 10^{-8}) \times 4 \times \cancel{\pi r^2} \times T^4$$

$$T = 286\ \mathrm{K} \text{ or } 13\,°\mathrm{C}$$

The current mean temperature of the Earth's surface is 288 K (15 °C), so it would appear that our basic model is reasonably accurate.

Three things should be very clear:

- The values of emissivity and albedo have a fundamental effect on a planet's temperature. They will have different values for a planet without an atmosphere.
- The values of emissivity and albedo on Earth are being changed by human activity.
- Small changes in the Earth's temperature may produce large changes in the Earth's climate, and such changes are mostly harmful to our lives.

WORKED EXAMPLE B2.3

Calculate a value for the average emissivity of the Moon. Assume that it has an average albedo of 0.12, and an average surface temperature of 274 K (but it should be noted that the Moon's surface temperature is *very* variable and an average is not really accurately defined).

Figure B2.5 The Moon

Answer

solar constant × cross-sectional area of Moon × (1 – albedo)
= emissivity × σ × surface area of Moon × T^4

$$1.36 \times 10^3 \times \cancel{\pi r^2} \times (1 - 0.12) = \text{emissivity} \times (5.67 \times 10^{-8}) \times 4 \times \cancel{\pi r^2} \times 274^4$$

Mean emissivity = 0.94

7 To determine a value for the Earth's surface temperature if it never had an atmosphere, we need to estimate values for emissivity and albedo under those conditions. Use estimated values of 0.9 (emissivity) and 0.3 (albedo) to make the calculation.

8 Determine a value for the surface temperature of the Earth if it never had an atmosphere and behaved as a perfect black body.

9 Calculate the average surface temperature of Mars, assuming that it has an average emissivity of 0.95 and an average albedo of 0.21. Its solar constant is 590 W m^{-2}.

10 Suppose that climate change resulted in a 5% change in both the emissivity and the albedo of the Earth. Predict a new value for the maximum average surface temperature.

11 State the factors which will affect the surface temperatures of planets orbiting a distant star.

Nature of science: Models

Using computers to expand human knowledge

Trying to predict the future, or to answer the question 'what would happen if …' has always been a common and enjoyable human activity. But it seems that our predictions are usually much more likely to be wrong than right. This is partly because, in all but the simplest of examples, there are just too many variables and unknown factors. Of course, the inconsistencies of human nature play an important part when dealing with people's behaviour, but accurately predicting events governed mostly by the laws of physics – such as next week's weather – can also be difficult.

Mathematical modelling is a powerful tool to understand a situation that can be represented by equations and numbers. But even in the simplest situations, there are nearly always simplifications and assumptions that result in uncertainty in predictions. When dealing with complex situations – such as predicting next month's weather, the value of a financial stock next year, or the climate in 50 years' time – even the most able people in the world will struggle with the complexity and amount of data. The rapid increase in computing power in recent years has changed this.

Modern computers have computing power and memory far in advance of human beings. They are able to handle masses of data and make enormous numbers of calculations that would never be possible without them. They are ideal for making predictions about the future climate, but that does not necessarily mean that the predictions will turn out to be correct. Computer predictions are limited by the input data provided to them and, more particularly, by the specific tasks that human beings have asked them to perform. To check the accuracy of predictions, computer models can be used to model known complex situations from the past to see if they are able to predict what actually happened next. But predicting the past is always much easier than predicting the future.

Effect of the Earth's atmosphere: greenhouse effect

So far, we have modelled the Earth's energy balance by treating the planet and its atmosphere as one system. In order to fully understand what is happening, we now need to consider transfers of energy within that system: between the Earth and its atmosphere. This is where a knowledge of the greenhouse effect becomes important.

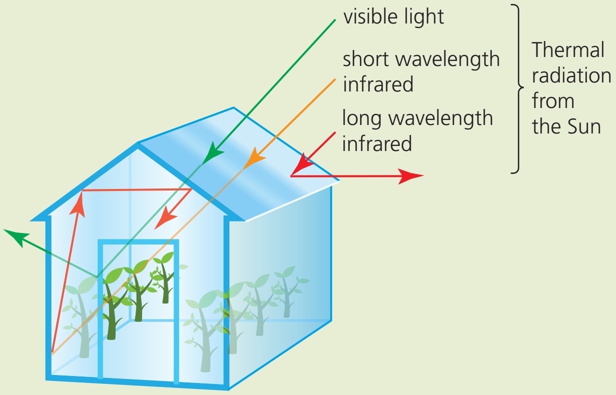
Nature of science: Models

Greenhouses and the Earth's atmosphere

A **greenhouse** is a building for controlling the temperature of growing plants. The walls and roof are made of glass, or transparent plastic. A significant number of large, impressive, ornate structures have been constructed over the last two centuries in colder countries, where there was much enthusiasm for growing plants, especially fruit and vegetables, native to hotter climates. See Figure B2.6.

■ **Figure B2.6** The Palm House at Schönbrunn Palace Park in Vienna

The major advantage of a greenhouse is that the temperature inside is hotter than outside. To understand the main reason why, we need to compare the radiation arriving at the greenhouse from the Sun to the radiation emitted from the contents of the greenhouse. See Figure B2.7.



■ **Figure B2.7** Thermal radiation in a greenhouse

We have already discussed the black-body spectrum from the Sun's surface (at 5780 K) and most of this thermal radiation falling on the glass of the greenhouse (light and infrared) will be transmitted through into the inside, where it will raise the temperature of the contents. Some longer wavelengths will be reflected back by the glass. Much of the light radiation entering the greenhouse will pass back out of the glass after reflection from surfaces in the interior.

The ground, other surfaces, pots and plants inside the greenhouse will get warmer because of the thermal radiation (infrared) that they absorb. Then they transfer thermal energy to the air by conduction. Convection currents move the warmer air upwards, but it mostly remains within the greenhouse.

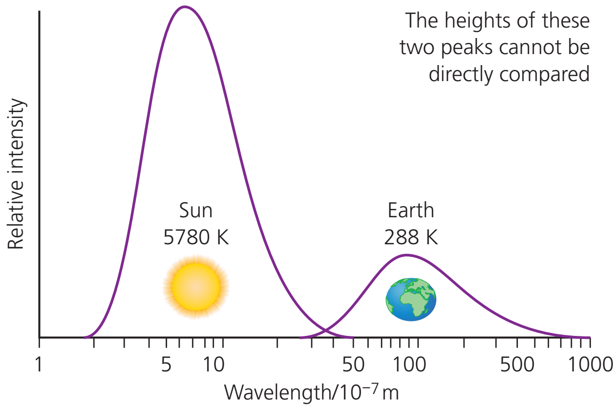
The thermal radiation emitted from the warmed contents of the greenhouse comes from much cooler surfaces (≈ 300 K) than the Sun (infrared, but no light). The spectrum of this radiation is very different from that of the Sun. Most importantly, the infrared wavelengths are much longer and they are not able to pass back out through the glass. In this way we may describe thermal energy as being 'trapped' inside the greenhouse.

When you have completed this topic, discuss in pairs: how valid is the comparison between the physics of greenhouses described here and the physics of the Earth's atmosphere? Is 'greenhouse effect' a useful term for these effects? To what extent might it mislead?

◆ **Greenhouse** Structure made mostly from a transparent material (usually glass) used for controlling plant growth.

Comparing the radiation emitted by the Sun to the radiation emitted by the Earth

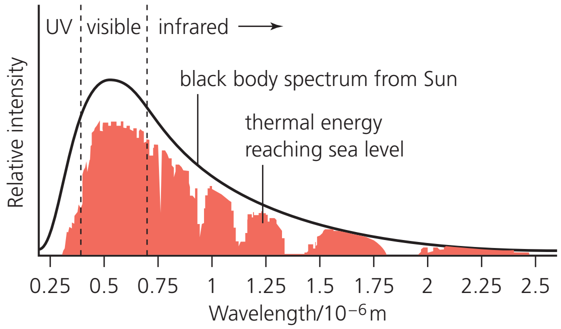
The spectrum of the black-body thermal radiation received from the Sun is very different from the thermal radiation spectrum radiated from the Earth's much cooler surface, as can be seen in Figure B2.8. Note that the horizontal scale is not linear (it is logarithmic) and the relative height of the peak for the Earth's spectrum has been *greatly* exaggerated for clarity.



■ **Figure B2.8** Comparing the radiation emitted by the Sun to the radiation emitted by the Earth

Compared to radiation from the Sun, the thermal radiation from the Earth is at a much lower power, with much longer wavelengths.

Most of the radiation from the Sun passes through the Earth's atmosphere, as can be interpreted from Figure B2.9.



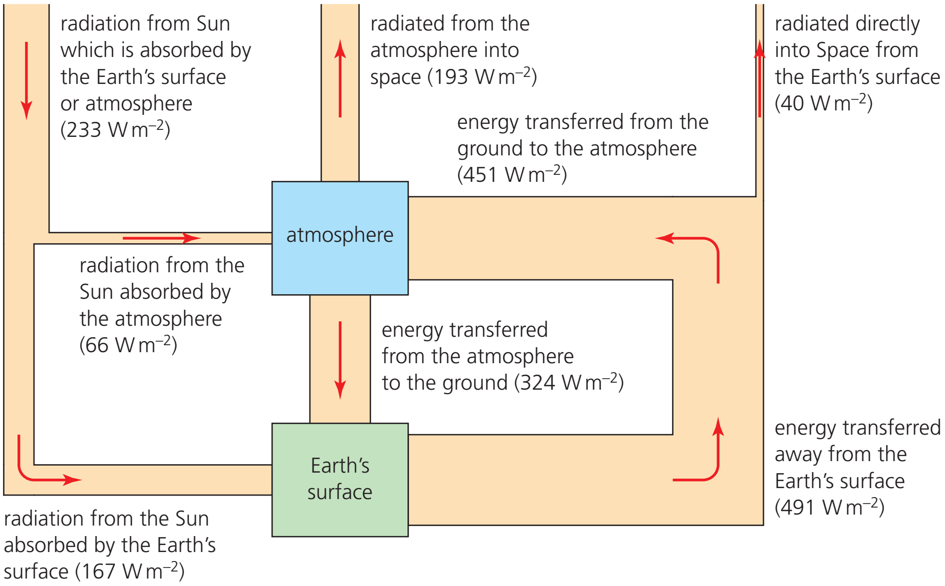
■ **Figure B2.9** The effect of the atmosphere on radiation arriving at the Earth's surface from the Sun

Because of its different wavelength range, a smaller proportion of the infrared from the Earth's surface is able to pass back out through the same atmosphere. However, over time, the temperature of the Earth's surface adjusts so that the *total* thermal power received by the Earth (including its atmosphere) = total thermal power radiated by the Earth (including its atmosphere).

Energy flow through the Earth's atmosphere

We have already seen that the mean intensity of thermal radiation arriving at the Earth from the Sun, averaged over the whole planet, over an extended period of time, is 340 W m^{-2}. We have also noted that the mean albedo of the Earth and its atmosphere is 0.315, meaning that only 68.5% of that 340 (= 233 W m^{-2}) is absorbed in the atmosphere or the Earth's surface.

Figure B2.10 summarizes what happens to that 233 W m^{-2}. You will find that numerical data varies slightly, depending on your source of information and its date. Note that data shows that the whole Earth, its surface and its atmosphere, are each receiving and emitting thermal energy at the same rates. Note that thermal energy transfer between the Earth's surface and its atmosphere is not just by radiation. Convection is also very important.



■ **Figure B2.10** Energy flow through the Earth's atmosphere

The importance of the atmosphere should be clear from studying Figure B2.9. We will now consider *how* the atmosphere absorbs and radiates energy.

ATL B2A: Research skills

Comparing, contrasting and validating information

Use the internet to research into data which either supports or conflicts with that shown in Figure B2.10. Is there much disagreement? Which sources do you think are the most reliable and/or up-to-date? Do you think that different websites get their information from the same original sources?

Greenhouse gases

SYLLABUS CONTENT

- Methane, CH_4, water vapour, H_2O, carbon dioxide, CO_2, and nitrous oxide, N_2O are the main greenhouse gases and each of these has origins that are both natural and created by human activity.
- The absorption of infrared radiation by the main greenhouse gases in terms of the molecular energy levels and the subsequent emission of radiation in all directions.
- The greenhouse effect can be explained in terms of either a resonance model or molecular energy levels.

◆ **Greenhouse gases** Gases in the Earth's atmosphere that absorb and re-emit infrared radiation, thereby affecting the temperature of the Earth. The principal greenhouse gases are water vapour, carbon dioxide, methane and nitrous oxide. Atmospheric concentrations of the last three of these have been increasing significantly in recent years.

The Earth's atmosphere has been formed over millions of years by naturally occurring volcanic and biological processes and from collisions with comets and asteroids. The air in the atmosphere contains approximately (by volume) 78% nitrogen, 21% oxygen and 0.9% argon. There are also naturally occurring traces of many other gases, including carbon dioxide and water vapour. Some of these trace gases are called **greenhouse gases** because they play a very important part in controlling the temperature of the Earth in the greenhouse effect. Greenhouse gases absorb (and then re-emit) infrared radiation.

There are many greenhouse gases but the four most important, in decreasing order of their contribution to the greenhouse effect, are:

- water vapour, H_2O (but see Top tip!)
- carbon dioxide, CO_2
- methane, CH_4
- nitrous oxide (dinitrogen monoxide), N_2O.

Top tip!

Water vapour is by far the most abundant of the greenhouse gases. However, for a variety of reasons, scientists do not believe that any future *changes* in water vapour concentrations in the atmosphere will significantly affect the Earth's climate.

Nitrogen, oxygen and argon have no greenhouse effect (because they have non-polar molecules or atoms).

The relative importance of these gases in causing the greenhouse effect depends on their abundance in the atmosphere as well as their ability to absorb infrared radiation. Each of the gases has natural as well as human origins.

Carbon dioxide contributes most to the overall greenhouse effect. Methane and nitrous oxide absorb infrared radiation more strongly than carbon dioxide, but their concentrations in the atmosphere are much lower.

Molecules of the greenhouse gases absorb some of the thermal radiation (infrared) emitted by the Earth's surface. Without these gases, the radiation would continue to travel away from the planet. The molecules will re-radiate the same energy a short time later, but in random directions, which means that about half is directed back towards the Earth's surface, keeping it warmer than it would be without the greenhouse gases.

LINKING QUESTIONS

- What relevance do simple harmonic motion and resonance have to climate change?
- What limitations are there in using a resonance model to explain the greenhouse effect?

These questions link to understandings in Topics C.1 and C.4.

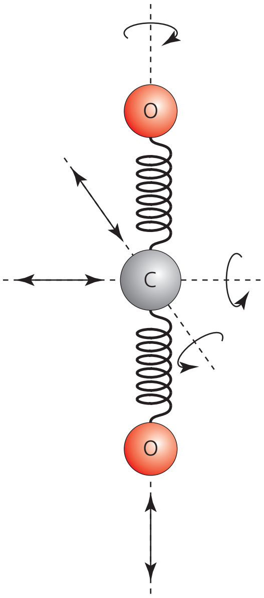


Figure B2.11 A few possible molecular vibrations in carbon dioxide

We will explain the absorption of infrared radiation by molecules of greenhouse gases by using carbon dioxide, CO_2, as an example. See Figure B2.11.

Molecules of greenhouse gases absorb infrared radiation because the atoms within their molecule are not at rest – they vibrate at high frequencies. (See Topic C.1: they oscillate with *simple harmonic motion*, like masses connected by springs.) Figure B2.11 shows a simplified example of possible ways in which a carbon dioxide molecule can vibrate.

If the atoms in a molecule vibrate at the same frequency as the infrared radiation passing through the greenhouse gas, then energy can be absorbed (an example of an effect known as *resonance* – see Topic C.4), raising the molecule to a higher energy level. The energy is quickly released again as the molecule returns to its lower energy level, but the released energy is radiated in random directions.

Since most of the radiation from the Sun is at higher frequencies it is much less likely to be absorbed than the mostly lower frequencies of radiation emitted from the cooler surface of the Earth.

As we have said before, the greenhouse effect, as described above, is essential for life on Earth. Next, we will discuss how the situation is changing.

12 Write a word equation representing the energy balance of a greenhouse at constant temperature. Assume that the inside is hotter than the outside and that there are no open windows.

13 Consider Figure B2.9. Outline the reasons for the decreases of intensity at some wavelengths (because of water vapour in the atmosphere).

14 One natural frequency of vibration of a carbon dioxide molecule is 2.0×10^{13} Hz.

- **a** Determine the wavelength of thermal radiation which this molecule will absorb.
- **b** Explain why carbon dioxide can absorb radiation from the Earth but not from the Sun.

15 Use data from Figure B2.10 to show that it is representing an atmosphere in thermal equilibrium.

16 Many people are concerned about the increasing levels of methane in the atmosphere. Use the internet to research the reasons for these concerns.

◆ **Global warming** Increasing average temperatures of the Earth's surface, atmosphere and oceans.

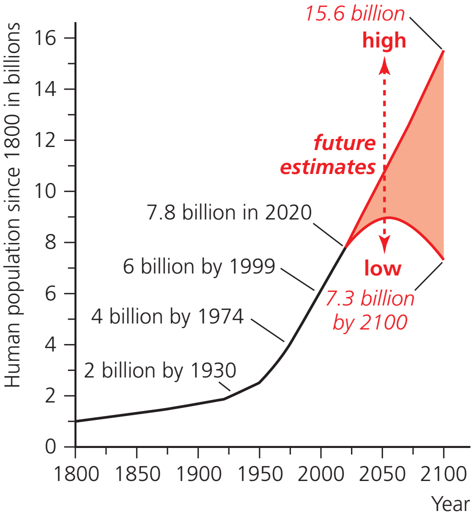
◆ **Combustion (of fuels)** Burning. Release of thermal energy from a chemical reaction between the fuel and oxygen in the air.

◆ **Natural gas** Naturally occurring fossil fuel: mixture of gases (mainly methane).

Enhanced greenhouse effect: global warming

SYLLABUS CONTENT

- ▶ The augmentation of the greenhouse effect due to human activities is known as the enhanced greenhouse effect.



■ **Figure B2.12** Prediction of World population growth (*source of data: UN World Population Prospects 2019*)

The population of the Earth passed eight billion towards the end of 2022, having quadrupled in less than a century. See Figure B2.12.

Understandably, everyone wants easy, enjoyable lifestyles, to be well fed, to be protected from heat and cold, to travel ... and so much more. All of these human activities involve transfers of energy. That energy has to be provided from somewhere and, after our activities, the energy is mostly dissipated into the surroundings and cannot be recovered.

For the last 300 years, we have been using the vast store of chemical potential energy available from the **combustion** (burning) of coal, oil and **natural gas** to power various types of machines and engines. These energy sources are called **fossil fuels** because they are made over the course of millions of years under the surface of the ground by the decomposition of once living material in the absence of oxygen. The enormous benefits to society from the use of fossil fuels (for example in the generation of electricity and in transport) are undeniable.

◆ **Fossil fuels** (A **fuel** is a store of chemical or nuclear energy that can be used to do useful work.) Naturally occurring fuels that have been produced by the effects of high pressure and temperature on dead organisms (in the absence of oxygen) over a period of millions of years. Coal, oil and natural gas are all fossil fuels.

However, we have become increasingly aware of the considerable disadvantages of the continued use of fossil fuels. Most notably, the release of increased amounts of greenhouse gases (principally carbon dioxide) into the atmosphere is responsible for **global warming**.

Increased concentrations of greenhouse gases in the atmosphere results in more of the thermal energy that is radiated away from the Earth's surface being absorbed in the atmosphere, some of which is re-radiated randomly back to the surface, increasing its average temperature.

This is known as the enhanced greenhouse effect. (Enhanced means increased.) It is currently believed that the enhanced greenhouse effect has resulted in increasing the average temperature of the Earth's surface by just over 1 °C during the last 60 years. Further rises are considered to be inevitable.

◆ Non-renewable energy sources Energy sources that take a very long time to form and which are being rapidly used up (*depleted*). Oil, natural gas and coal.

The disadvantages of using fossil fuels have been well understood and discussed among scientists for well over 50 years. To begin with, attention was mainly on the polluting effects on their immediate environments and when accidents occurred, and the fact that they were **non-renewable sources**: there was a limited supply (some predictions were made that supplies would be running low by now). In more recent times, the world's attention has shifted to their effect on the global climate.

The burning of fossil fuels is almost certainly the greatest cause of the enhanced greenhouse effect.

With enormous quantities of relevant data available and the use of super-computers, nearly all scientists agree with the last two highlighted statements and their consequence: global warming. Curiously, some of the general public and some politicians have been less easy to convince. It is difficult to understand why.

Nature of science: Models

Simple and complex

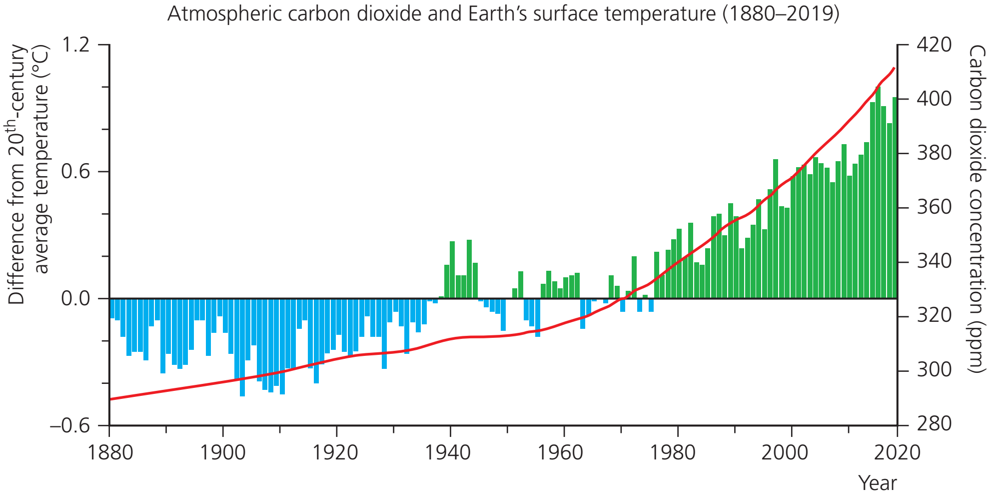
The kinetic theory of gases is a model that can be applied successfully to relatively small amounts of gases in closed containers. The Earth's atmosphere is also a gaseous system, but one which is vastly larger and more complex. Meteorologists use computer modelling to predict the weather in a particular location with reasonable accuracy up to about 10 days in advance. But predicting the climate of an entire planet with any certainty for many years in the future is a near impossible task. However, climate modelling is a problem which has understandably attracted an enormous amount of scientific attention in recent years and, with the availability of better data and faster processing, together with international **collaboration**, long-term **climate models** are believed to have become more consistent and reliable. We will have to wait to see how accurate they are.

◆ Collaboration (scientific) Two or more scientists sharing information or working together on the same project.

◆ Climate model A complex computerized model that attempts to predict the future climate of the planet, especially how it will be affected by global warming.

◆ Correlation There is a correlation between two sets of varying data if they show similarities that would not be expected to occur because of chance alone.

Figure B2.13 compares the levels of carbon dioxide in the atmosphere (in parts per million, ppm) with the Earth's average surface temperature, which is represented by the difference of the annual average with the average over the hundred years of the twentieth century. The **correlation** is easy to see, and it is believed by scientists that increased carbon dioxide levels caused the increasing temperatures, although the data in this chart is not conclusive by itself. Variations in carbon dioxide levels correlate well with variations in temperature over tens of thousands of years, long before humans began burning fossil fuels in vast quantities. See Figure B2.14. However, one important aspect of Figure B2.13 is the unusually short timescales involved.



■ **Figure B2.13** Correlating the concentration of carbon dioxide in the atmosphere with the Earth's temperature (*source of data: NOAA Climate.gov; ESRL / ETHZ / NCEI*)

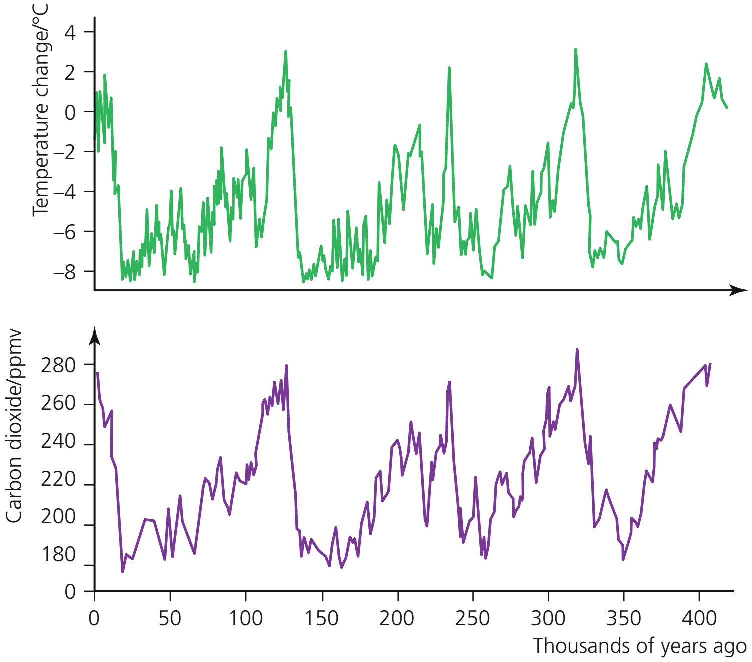


Figure B2.14 There is a strong correlation between average global temperatures and the amount of carbon dioxide in the atmosphere (ppmv = parts per million by volume)

The correlation between increasing concentrations of greenhouse gases and rising global temperatures is well established and accepted by (almost) everybody. However, that does not mean that we can be 100% sure that global warming is caused by the release of more greenhouse gases. Obviously, controlled experiments to test such a theory cannot be carried out and we must rely on statistical evidence, computer modelling and scientific reasoning. In such cases, 100% certainty is never possible and individuals and societies must make informed judgements based on the best possible scientific evidence. Of course, some people will always choose to disagree with, or ignore, the opinions of the majority.

Tool 3: Mathematics

Determine the effect of changes to variables on other variables in a relationship

When the similarities between two sets of data are as clear as that seen in Figures B2.13 and B2.14, we describe it as a *correlation*. Without any further evidence, it is easy to believe that one effect (A) caused the other (B), which we would describe as *cause and effect*. However, it may be possible that effect A was caused by effect B, or maybe they are inter-dependent in some way. Two other possibilities need to be considered: both effects are a consequence of another cause (C), or maybe the correlation is just an unlikely random phenomenon, which has no known explanation.

It is also possible that A did cause B, but only because there was some other effect involved, that has not been identified. Greater certainty can only be gained by gathering further data.

Consequences of the enhanced greenhouse effect

These are well documented elsewhere and there is no intention to go into detail here. These interconnected consequences, which are expected to get worse, are mostly detrimental to the lives of people and animals on the Earth, although some places will be more affected than others, and richer countries will be better able to cope with the changes.

- Increasing temperatures of the oceans, land and atmosphere
- Climate change, melting snow and ice
- More frequent extreme weather conditions (storms, floods, drought, fires and so on)
- Rising sea levels with increased acid levels.

Figure B2.15 Wildfires in Greece 2021

Figure B2.16 Flooding in Germany 2021

Common mistake

Many people wrongly believe that increasing global temperatures is a consequence of the basic greenhouse effect. More correctly, we should say that human activities are changing global temperatures because of an *enhanced* greenhouse effect.

TOK

Knowledge and the knower

- What criteria can we use to distinguish between knowledge, belief and opinion? Is the truth what the majority of people accept?
- Can probability become certainty?

The four bullet points listed above are empirical facts, based on very extensive measurements. But do we know *for certain* that they are a consequence of burning fossil fuels? Clearly, the vast majority of scientists and the general public now believe so. Ten or twenty years ago more people were doubtful, but the increasing weight of evidence is convincing.

At what point does something like this become accepted knowledge, or will there always be some uncertainty? It seems likely that, under any circumstances, some people will continue to believe that climate change is unconnected to burning fossil fuels.

What are the possible solutions to climate change?

There is a wide range of actions that individuals and governments can take. These include:

- Support governments who will take appropriate and prompt action. (Including an *increasing use of renewable energy sources* – see below, discouraging energy-intense activities and investing in scientific research and development, for example, into 'carbon capture').
- Accept higher taxation as a method of discouraging energy intense activities and the use of fossil fuels.
- Invest in renewable energy sources for individual homes, such as solar panels on the roof.
- Use energy-efficient devices and use them less frequently.
- Improve home insulation against both hot and cold weather. See Figure B2.17.
- In colder countries: reduce the temperatures of homes and the hot water used in showers and washing machines.
- In hotter countries: increase the temperature settings on air-conditioners.

Figure B2.17 Improving insulation under the roof

- Waste less (of everything, but especially food). Re-use items and resist the temptation to keep buying new things.
- Reduce methane emissions.
- Many populations could eat less, especially meat.
- Make public transport cheaper and more plentiful.
- Support businesses which are taking action against climate change.
- Use electric vehicles. Use smaller, fewer and less powerful vehicles.
- Limit travel (especially for pleasure).
- Recycle.
- Change lifestyles and expectations.
- Reforestation.

LINKING QUESTION

- How do different methods of electricity production affect the energy balance of the atmosphere?

This question links to understandings in Topics B.2 and D.4.

ATL B2B: Thinking skills

Evaluating and defending ethical positions

Do people in rich and developed countries have any ethical authority to complain about the way in which fossils fuels are used in poorer countries?

Do we have individual responsibilities to take action against climate change? Is it acceptable to do nothing, or assume that the government will act on our behalf?

Would you be happy if the country where you lived took strong measures to try to limit climate change, while other countries did much less, or nothing?

◆ **Renewable energy** Energy from sources that will continue to be available for our use for a very long time. They cannot be used up (depleted), except in billions of years, when the Sun reaches the end of its lifetime.

World use of energy resources

Using more **renewable energy** sources should reduce our dependence on fossil fuels. As their name suggests, renewable energy sources will continue to be available to us in the future, because they are continuously being renewed by the energy arriving from the Sun. Their other major advantage is that they do not contribute to climate change.

Figure B2.18 shows recent information (2020) about the world's energy resources. Other sources of information will show some variations from these figures, but the trends are generally agreed.

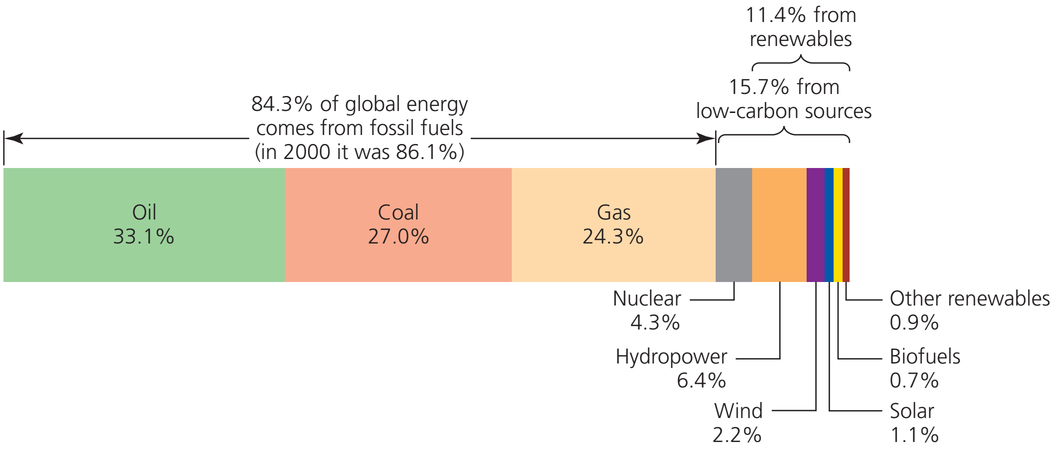
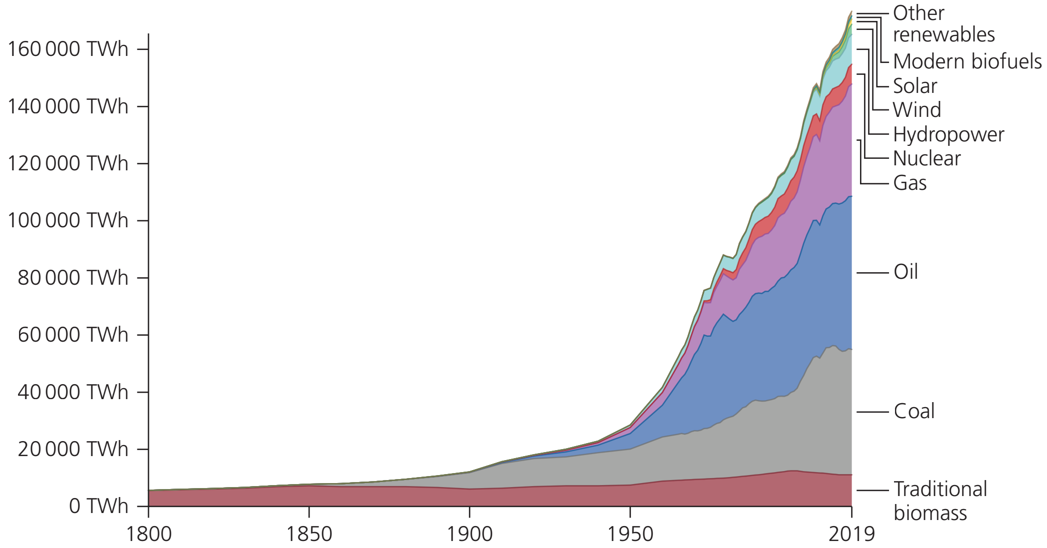


Figure B2.18 BP's 2020 Statistical Review of World Energy (*source of data: BP Statistical Review of World Energy*)

Figure B2.19 shows how the uses of various energy sources has changed over the last 120 years. (1 TWh = 3.6×10^{15} J). There are justified high hopes for the continued rapid increases in the use of wind power and solar power, but currently they still amount to less than 4% of our global needs.



🟥 **Figure B2.19** Changes in energy consumption over 120 years (*source of data: Vaclav Smil (2017) and BP Statistical Review of World Energy*)

Perhaps the most important information that can be seen in Figure B2.19 is the increase in the planet's *overall* demand for energy. Although the Covid pandemic has had a temporary effect in reducing demand, the overall trend is likely to continue upwards because the populations of poorer countries will understandably hope to match the living standards of those in richer countries.

Most of the energy for this increased demand has come from fossil fuels. Over the last 20 years, fossil fuels have continued to supply about 85% of the world's increasing energy needs. This means that we are burning about 50% more fossil fuels now than in the year 2000. All this is despite the demand for a greater use of renewable energy sources. Fossil fuels remain plentiful and relatively cheap, millions of people around the world are employed in the fossil-fuel businesses, and we already have the systems and infrastructures in place to continue their use.

ATL B2C: Communication skills

Reflecting on the needs of the audience

Prepare a 5 minute presentation on the enhanced greenhouse effect which could be understood by 10–12 year-old students. What simplifications will you need to make? How can you present the scientific information in an engaging and accessible way?

17 Use Figure B2.13 to estimate the percentage by which the concentration of carbon dioxide in the atmosphere increased in the 50 years between 1970 and 2020.

18 Sketch Sankey diagrams which show all the principal energy transfers that have resulted in:

a a wind generator producing an electric current

b an oil-fired power station producing an electric current.

19 Explain why hydroelectric power is considered to be a renewable energy source.

🟥 **Figure B2.20** What a 2 °C rise in ocean temperature could do to the Hard Rock 2020 Super Bowl Stadium at Miami Beach, Florida

20 Use the internet to determine

a two reasons why increasing temperatures result in increasing sea levels

b the latest predictions for future sea levels.

21 a Discuss whether nuclear energy is renewable or non-renewable.

b Figure B2.21 shows how use of nuclear power has changed since the 1960s. Suggest some reasons why it is not used more widely.

c State what major incidents happened in 1986 and 2011 which affected people's opinions about nuclear power.

22 Use the internet to find the primary sources of energy used to generate electricity in the country where you live.

23 Would you like the government of the country where you live to take more action in an attempt to limit climate change? If not, why not? If yes, what changes would you recommend?

24 Use the internet to gain information and data about the risks associated with the generation of electricity from non-nuclear sources.

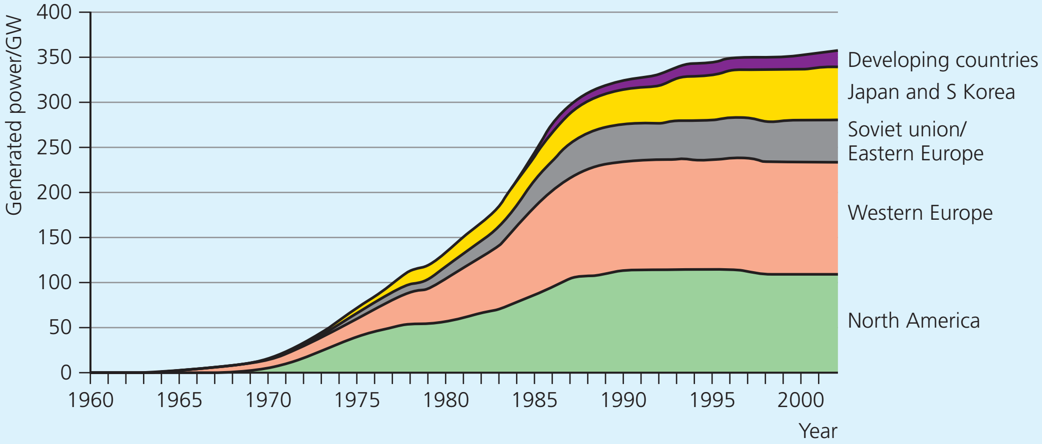


Figure B2.21 Use of nuclear power

Nature of science: Global impact of science

United Nations Climate Change Conferences

After years of relative inaction, at the time of writing, the urgent need for significant and widespread action on climate change finally seems to have become widely accepted, especially among younger people.

These important annual COP meetings involve detailed discussions about how the effects of climate change could be reduced. They involve scientists and representatives from the governments of most of the countries of the world.

Undoubtedly, important agreements are reached. Research online to find out what these were at the latest COP meeting.

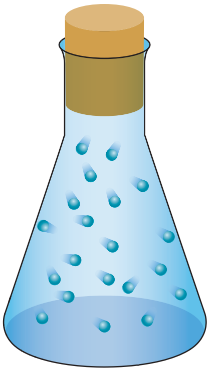
However, there are many voices raised in opposition: Did the proposed changes go far enough? Was it all just talk? Is keeping temperature rises to 1.5 °C already impossible? Why were some important countries apparently not fully involved? Will countries do in the future what they have promised to do? We will see.

Figure B2.22 Activists at COP 26

B.3 Gas laws

Guiding questions

- How are macroscopic characteristics of a gas related to the behaviour of individual molecules?
- What assumptions and observations lead to universal gas laws?
- How can models be used to help explain observed phenomena?



■ **Figure B3.1** Model of gas molecules in a container

◆ **Gas laws** Laws of physics relating the temperature, volume and pressure of a fixed amount of a gas: Boyle's law, Charles' law and the pressure law.

◆ **Pressure, *P*** Force acting normally per unit area: pressure = force / area (SI unit: pascal, Pa). 1 Pa = 1 N m^{-2}.

◆ **Amount of gas** The quantity of gas in a container, expressed in term of the number of particles it contains.

At the beginning of Topic B.1 we introduced the kinetic theory of matter and explained that all gases contain particles (usually molecules) moving in random directions, usually at high speeds, as represented in Figure B3.1.

Because we can usually assume that there are no forces acting between the molecules (except in collisions), the physical properties of gases are much easier to study and understand than solids and liquids. As we will see, using an idealized model of the motion of molecules in a gas, we can use knowledge of dynamics (from Topic A.2) to predict the physical behaviour of gases. It will be important to understand this link between microscopic motions of molecules and the macroscopic properties of gases.

Under most conditions, all gases show similar physical behaviour. The (universal) **gas laws** is the name that scientists give to the straightforward relationships that describe the physical behaviour of all gases.

We can determine the following four physical properties of any gas in any container:

- volume, V
- temperature, T
- **pressure**, P (explained below)
- **amount of gas**, n, in terms of the number of gas molecules it contains (explained below).

In the kinetic theory of gases, the number of molecules in a gas is usually more useful information than their overall mass. The *density* of a gas may also be of interest but can be determined from mass / volume.

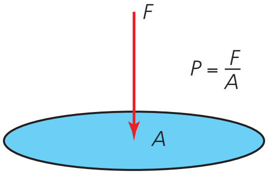
Before looking at the physical properties of gases, we need to explain more about two of the four properties in the above list: pressure and amount of a substance.

Pressure

SYLLABUS CONTENT

▶ Pressure, $P = \frac{F}{A}$ where F is the force exerted perpendicular to the surface.

Generally, the effect of a force, F, often depends on the area, A, on which it acts. For example, when the weight of a solid pushes down on a surface, the consequences usually depend on the area underneath it, as well as the magnitude of the weight (see Figure B3.2).

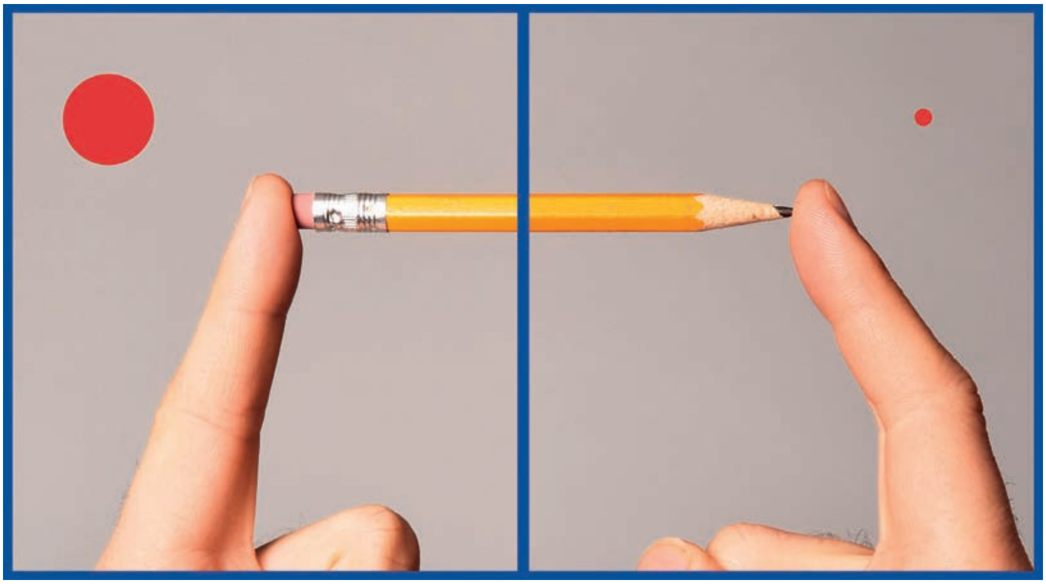


■ **Figure B3.2** Calculation of pressure from force and area

Pressure is defined as perpendicular (normal) force per unit area: $\boldsymbol{P = \frac{F}{A}}$
SI unit: **pascal**, Pa. 1 Pa = 1 N m^{-2}

◆ **Pascal** Derived SI unit for pressure. 1 Pa = 1 N m^{-2}.

The concept of pressure is needed to explain, for example, why one finger in Figure B3.3 will be less comfortable than the other. From Newton's third law, we know that the forces on both fingers are the same.



■ **Figure B3.3** Same force, different pressure

WORKED EXAMPLE B3.1

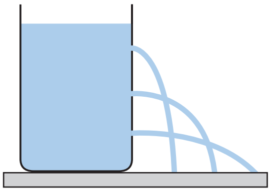
A girl of mass 51 kg stands on one leg. If the effective area of her foot in contact with the ground is 62 cm², calculate the pressure she is exerting on the ground.

Answer

$$P = \frac{F}{A} = \frac{(51 \times 9.8)}{62 \times 10^{-4}} = 8.1 \times 10^4\,\text{Pa}$$

WORKED EXAMPLE B3.2

Figure B3.4 shows some water coming out of a container with three holes. This is often shown to students as a demonstration about pressure in liquids.



■ **Figure B3.4** Pressure with depth apparatus

a What conclusion should students reach after watching this demonstration?

b If the depth of the water is 26 cm and the area of the bottom of the container is 84 cm², calculate:

- **i** the volume of water in the container
- **ii** the mass of water (density = 1000 kg m^{-3})
- **iii** the weight of the water.

c Determine the pressure of the water at the bottom of the container. (Ignore the pressure due to the air above the water.)

Answer

a The pressure in the water increases with depth.

b **i** volume = depth × area = 0.26 × (84 × 10^{-4}) = 2.2×10^{-3} m³ (2.184×10^{-3} seen on calculator display)

ii mass = volume × density = $2.184 \times 10^{-3} \times 1000$ = 2.2 kg (2.184 seen on calculator display)

iii weight = mg = 2.184 × 9.8 = 21 N (21.403... seen on calculator display)

c $P = \frac{F}{A} = \frac{21.403...}{84 \times 10^{-4}} = 2.5 \times 10^3\,\text{Pa}$

To determine the *total* pressure at the bottom of the container we would need to add the pressure due to the water to the gas pressure of the air above the water surface (1.0×10^5 Pa).

Pressure in gases

Our main concern in this topic is the pressure created by a gas in a container, but we explain gas pressure in a different way to the pressure under solids and liquids. Consider Figure B3.5.

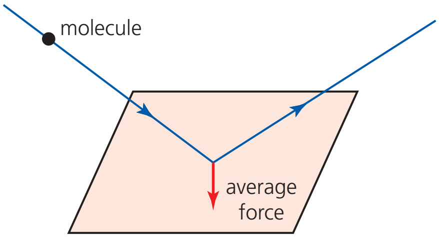


Figure B3.5 Every molecular collision with a wall creates a tiny force on the wall

Each collision of a molecule with the containing walls creates a tiny outwards force. The average forces are perpendicular (normal) to the surfaces. Because the molecular motions are random, each collision can result in a different sized force, but the enormous number of molecules colliding every second with a small area means that the average force / area will stay the same. That is, the pressure is constant. Note that molecules also collide with other molecules, but this simply re-directs their motion and has no other overall effect.

Note that gas pressure acts in *all* directions, not just downwards. This is also true for the pressure in liquids.

◆ **Atmospheric pressure** Pressure due to the motions of the gas molecules in the air. Can be considered as being due to the weight of the air above an area of 1 m^2. Acts equally in all directions.

The pressure of the air around us (**atmospheric pressure**) is 1.0×10^5 Pa at sea level. It decreases with height above the ground.

Amount of a substance

SYLLABUS CONTENT

- The amount of substance, n, as given by: $n = \frac{N}{N_A}$, where N is the number of molecules and N_A is the Avogadro constant.

The **amount of a substance** (symbol: n) is a measure of the number of atomic-scale particles it contains. The nature of the 'particles' depends on the substance being considered. For example, the gas helium, He, consists of separate atoms, but most other gases are molecular. Examples include H_2, O_2, CO_2 CH_4 and N_2.

Even very small amounts of gas can contain an enormously large number of particles (≈ 10^{19} or more). The SI unit of amount of a substance is more manageable. It is called the **mole** (mol):

One mole is the amount of a substance that contains exactly $6.022\,140\,76 \times 10^{23}$ of its particles. This number is known as the **Avogadro constant**. It is given the symbol N_A. For most calculations we can use a value of $\mathbf{6.02 \times 10^{23}}$.

◆ **Amount of substance, *n*** Measure of the number of atomic-scale particles (atoms or molecules) it contains (SI unit: mole).

◆ **Mole, mol** SI unit of amount of substance (fundamental).

◆ **Avogadro constant, N_A** The number of particles in 1 mole of a substance: 6.02×10^{23}.

Top tip!

Since 2018, the mole and the Avogadro constant have been defined only by the number shown above. Previously, the Avogadro constant was defined with reference to a standard substance: carbon. N_A was the same number of particles as there are atoms in exactly 12 g of the isotope carbon-12. (Isotopes are explained in Topic E.3.) This link has now been removed from the definition, but the number is still the same.

Figure B3.6 shows one mole of various substances. Figure B3.7 shows a ball which contains about 0.25 moles of air at a pressure of about 9×10^4 Pa.

Figure B3.6 One mole of water (in the form of ice), sugar, copper and aluminium

Figure B3.7 A football

If we know the number of particles in a substance, *N*, the number of moles, *n*, is determined by dividing by the Avogadro constant:

$$\text{amount of substance in moles} = \frac{\text{number of particles}}{\text{Avogadro constant}}$$

$$n = \frac{N}{N_A}$$

◆ **Molar mass** The mass of a substance that contains 1 mole of its defining particles.

The **molar mass** of a substance is the mass which contains one mole. Usual unit: g mol^{-1}. Table B3.1 shows some common molar masses.

Table B3.1 Molar masses

Substance	**Molar mass / g mol^{-1}**
hydrogen molecules	2.02
helium atoms	4.00
carbon-12 atoms	12.00
carbon atoms	12.01
water molecules	18.01
aluminium atoms	26.98
nitrogen molecules	28.02
oxygen molecules	32.00
carbon dioxide molecules	44.01
gold atoms	197.00

Top tip!

The molar mass of a substance depends on the number of particles (protons and neutrons) in the nucleus of each atom or molecule. (Details are provided in Topic E.3.) More massive atoms have more protons and neutrons, so that a mole of their atoms will have a greater mass. For example, carbon atoms usually have 12 particles, carbon dioxide molecules usually have 44 particles and hydrogen molecules usually have two particles.

A numerical example: an oxygen molecule has two atoms, each has 16 particles, each with a mass of 1.67×10^{-27} kg.

Total mass of one mole (6.02×10^{23} particles) $= 6.02 \times 10^{23} \times 2 \times 16 \times 1.67 \times 10^{-27} = 3.2 \times 10^{-2}$ kg (32 g)

WORKED EXAMPLE B3.3

a How many moles are there in 0.50 kg of molecular oxygen gas?

b How many molecules are there in 0.50 kg of the same gas?

Answer

a $\frac{\text{total mass}}{\text{molar mass}} = \frac{0.50}{32 \times 10^{-3}} = 16\,\text{mol}$ (15.625 seen on calculator display)

b $N = nN_A = 15.625 \times (6.02 \times 10^{23}) = 9.4 \times 10^{24}$ molecules

1 A car of weight 1500 kg has four tyres, each of which has an area of 180 cm^2 in contact with the road.
   - a Calculate the pressure under each tyre (in Pa).
   - b State the pressure of the air inside the tyre.
   - c If the driver puts more air into the tyre, what will happen to:
     - i the area in contact with the ground
     - ii the pressure on the road?

2 The air pressure at a height of 10 km above sea level is 2.6×10^4 Pa. Inside a passenger aircraft at this height, the air pressure is maintained at 80% of the pressure at sea level (1.0×10^5 Pa).
   - a What is the difference in pressure between inside and outside the aircraft?
   - b What resultant force due to the air is acting on a window which is 27 × 47 cm?

Figure B3.8 Aircraft window

3 a Using the concept of pressure, explain how it is possible for the water to remain in the upside-down glass shown in Figure B3.9. (Assume that the glass is full of water.)
   - b If the glass is only half full of water to begin with, how will the demonstration change?

Figure B3.9 Water in an upside-down glass

4 Consider the ball shown in Figure B3.7.
   - a If the molar mass of the air in the ball is approximately 29 g mol^{-1}, what is the mass of the air in the ball?
   - b Compare the pressure of the air inside the ball to the pressure of the air outside the ball.

5 a Estimate the volume of the water seen in the glass in Figure B3.9.
   - b Water has a density of 1000 kg m^{-3}. Using your answer to part **a**, estimate:
     - i the mass of water in the glass
     - ii the number of moles of water in the glass
     - iii the number of water molecules in the glass.

6 The thickness of a car tyre decreased by 5.0 mm over a distance of 30 000 km. The circumference of the tyre was 1.9 m and its surface area was 3000 cm^2.

Figure B3.10 Car tyre

- **a** Calculate the volume of the rubber in the tyre which was worn away in this distance.
- **b** Assuming the density of the rubber was 950 $kg\,m^{-3}$, what mass was worn away?
- **c** If about 90% of the mass in rubber is due to carbon atoms, estimate:
  - **i** the number of moles of carbon spread into the environment in travelling each 1000 km
  - **ii** the number of carbon atoms lost from the tyre in each rotation of the wheel.

7 Nitrogen is often put inside potato crisps / chips packets. This is to keep the crisps fresh and limit damage to them. Estimate the amount of nitrogen in a typical packet (see Figure B3.11). Assume the density of the nitrogen is 1.4 $kg\,m^{-3}$.

Figure B3.11 Nitrogen was used inside the pack to keep these crisps / chips fresh.

8 Most iron atoms contain a total of 54 protons and neutrons, each with a mass of 1.67×10^{-27} kg. Determine a value for the molar mass of iron.

Investigating the physical properties of gases

SYLLABUS CONTENT

- The empirical gas laws for constant pressure, constant volume and constant temperature.

As we have already noted, given a closed container with a gas sealed inside, there are four physical properties of the gas which can be easily measured: pressure, volume, temperature and mass (P, V, T and m). The amount of gas in a mole, n, can be calculated from its mass, as explained above.

All gases exhibit the same physical properties (under most conditions), so that any gas, or mixture of gases, can be used in the following experiments, and the same conclusions should be reached. Air is the obvious and easy choice.

The following are three classic investigations of the ‘gas laws’. Each keeps the amount of gas and one other variable constant:

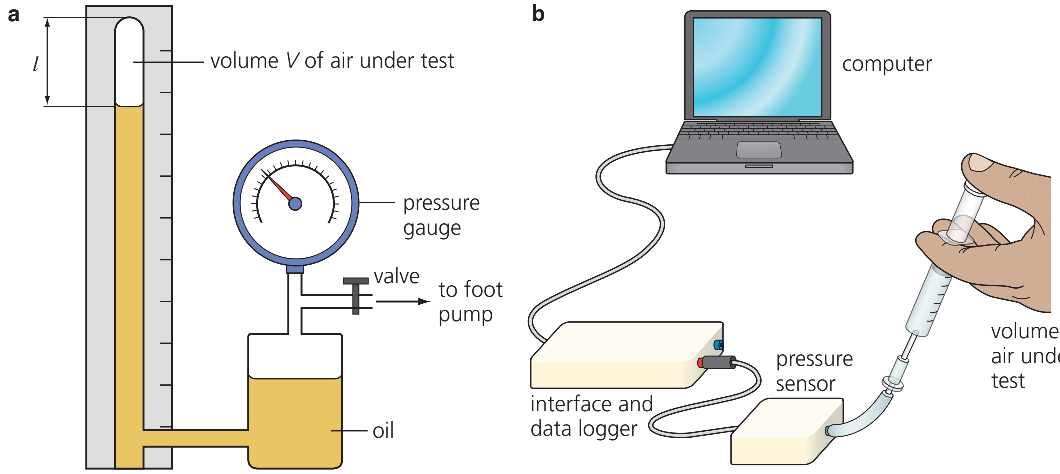
- How does the volume of a gas change with pressure (at constant temperature)?
- How does the volume of a gas change with temperature (at constant pressure)?
- How does the pressure of a gas change with temperature (at constant volume)?

We will look at each of these in more detail.

Variation of gas volume with pressure at constant temperature (Boyle’s law)

Figure B3.12 shows two sets of apparatus that could be used to investigate how changing the pressure on a fixed mass of gas affects its volume, at a constant temperature. However, note that using a force to do work and change the volume of a gas will tend to change its temperature, which could complicate the results. To minimize this unwanted effect, the changes should be made slowly.

Figure B3.12 Investigating Boyle's law



Tool 1: Experimental techniques

Recognize and address relevant safety, ethical or environmental issues in an investigation

The safety of you (and your teachers) in a science laboratory is an important concern. However, it is not a major issue if your behaviour is appropriate. Most schools have a list of behaviours expected of students in a laboratory. The few accidents which occur are usually because of inattention or carelessness.

Experimental situations requiring particular attention include the following:

- **Using electrical equipment which is connected to the mains supply (110–230 V).** Such equipment should have regular safety checks and never be used close to water. Bare wires connected to the mains should not be used for experimentation. The laboratory should have appropriate **circuit breakers** and an RCCB which can cut off the electrical supply very quickly and protect lives.
- **High voltage supplies** (for example, 5000 V) may be needed for a few teacher demonstrations, but they are provided with protection due to the very high internal resistance included deliberately in their design.
- **Electrostatic high voltage generators** always make interesting and dramatic demonstrations. They are very safe to use, but people with serious medical conditions are usually advised not to get involved.
- **Hazardous chemicals** are not usually used in physics experiments.
- **Equipment made from glass** always needs to be treated carefully to avoid breakage. Goggles should be worn.
- **Containers with a gas at high or low pressure** should have a shield around them for protection in the unlikely event of an explosion or implosion.
- **Radioactive sources** will usually be used only by a teacher, although some schools allow older students to use the sources under close supervision. They should be labelled and stored securely and used for as short a time as possible. Students should not be too close to the experiments and the sources should be shielded and never directed towards students.
- **The light from a laser** (or other very intense light source) should not be allowed to fall on anyone's eyes.
- **High temperatures** are required for a few experiments. Appropriate care is needed, especially if very hot water is involved.

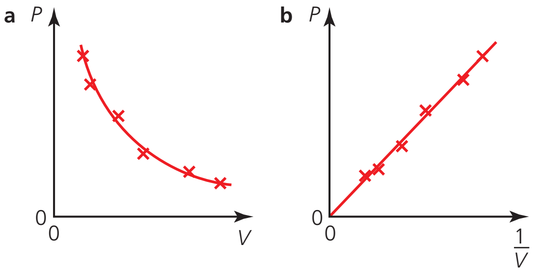
The gas pressure in the apparatus seen in Figure B3.12 can be high enough that there is small possibility that a glass tube could explode. It should be surrounded by a clear plastic protective screen.

◆ **Circuit breaker** Electromagnetic device used to disconnect an electrical circuit in the event of a fault.

◆ **Isothermal** Occurring at constant temperature.

Figure B3.13 shows graphs of typical results. The lines are called **isotherms**, which means that all points are at the same temperature.

The graph shown in Figure B3.13b represents the same data as in Figure B3.13a, but the graph has been re-drawn to produce a straight line to show that the pressure and volume are inversely proportional to each other.



■ **Figure B3.13** Two graphs showing that gas pressure is inversely proportional to volume

For a fixed amount of gas at constant temperature: $P \propto \frac{I}{V}$

This is known as **Boyle's law**.

Boyle's law can be stated as PV = constant. If the pressure and/or volume of a fixed amount of gas are changed from initial values of P_1 and V_1 to final values of P_2 and V_2, then, provided that the temperature has not changed:

$$P_1V_1 = P_2V_2$$

◆ **Boyle's law** Pressure of a fixed amount of gas is inversely proportional to volume (at constant temperature).

We can explain this relationship in terms of molecular behaviour. If the volume of a container is reduced, the molecules (travelling with the same average speed) will collide with a given area of the walls more frequently. In other words, there will be more molecular collisions with each square centimetre every second, which will increase the gas pressure.

WORKED EXAMPLE B3.4

A sample of gas has a volume of 43 cm^3 when at normal atmospheric pressure (1.0×10^5 Pa).

a If the pressure on the same amount of gas is increased to 3.7×10^5 Pa, calculate its new volume.

b State the assumption you made in answering part **a**.

Answer

a $P_1V_1 = P_2V_2$

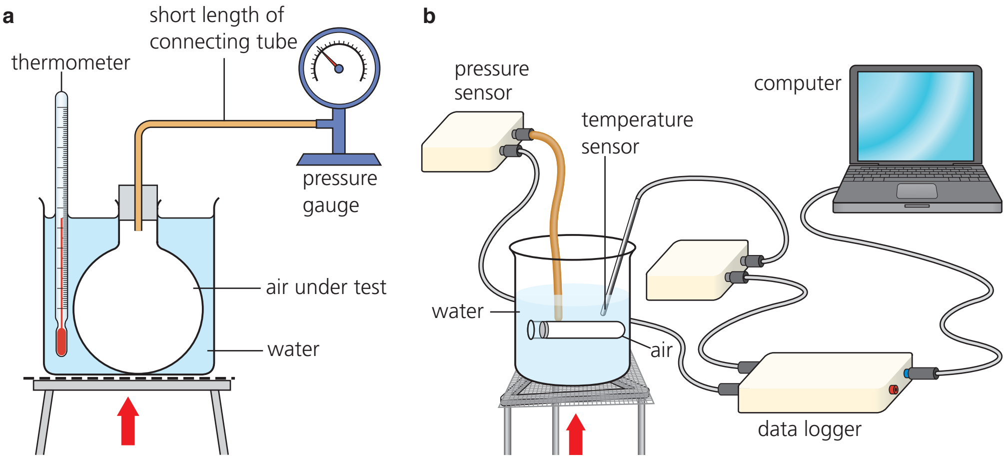
$(1.0 \times 10^5 \text{ Pa}) \times 43 = (3.7 \times 10^5) \times V_2$

$V_2 = 12 \text{ cm}^3$

b The temperature of the gas did not change.

Variation of gas pressure with temperature at constant volume (pressure law)

This relationship can be investigated with either of the two sets of apparatus shown in Figure B3.14. Measurements are usually taken for temperatures between 0 °C and 100 °C.



■ **Figure B3.14** Two sets of apparatus that can be used to investigate the pressure law

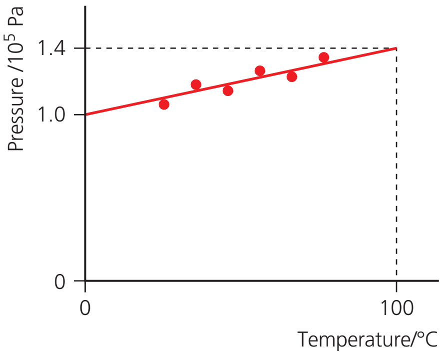
Tool 1: Experimental techniques

Recognize and address relevant safety, ethical or environmental issues in an investigation

There are several possible hazards when using the apparatus seen in Figure B3.14. There is the possibility that a larger beaker of very hot water could be knocked over, so the apparatus should not be near the edge of the table and nobody should be sitting nearby. There needs to be a good seal where the bung is pushed into the flask, but gloves should be worn when pushing them firmly together, in case the glass breaks.

Figure B3.15 shows a graph representing typical raw data.

The results represent a linear relationship between pressure and temperature in degrees Celsius.

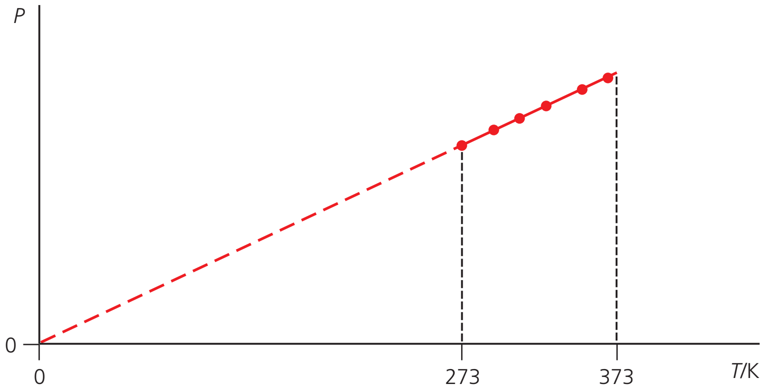


■ **Figure B3.15** Variation of gas pressure with temperature

We can explain this relationship in terms of molecular behaviour. If the temperature is reduced, the molecules move more slowly, so that they will collide with a given area of the walls less frequently, so that the pressure is reduced.

If the straight-line graph seen in Figure B3.15 is extrapolated to lower and lower temperatures and pressures, the temperature at which the pressure will be predicted to be zero is −273 °C. The same result is obtained with any gas. We can assume that this is the temperature at which (almost) all molecular motion has stopped, molecules are no longer colliding with the walls. −273 °C is called absolute zero, and it is the lower fixed point for the Kelvin temperature scale, as already explained in Topic B.1.

In practice, most gases will condense and then freeze at various low temperatures, but that does not change the concept of an absolute zero at −273 °C (0 K).



■ **Figure B3.16** Gas pressure is proportional to temperature in kelvin

Figure B3.16 shows the results of Figure B3.15 re-drawn. It represents a proportional relationship but remember that temperature must be measured in kelvin.

◆ **Pressure law** For a fixed mass of gas with a constant volume, the pressure is proportional to the kelvin temperature.

For a fixed amount of gas at constant volume: $P \propto T$

This is known as the **pressure law**.

If the pressure and/or temperature of a fixed amount of gas are changed from initial values of P_1 and T_1 to final values of P_2 and T_2, then, provided that the volume has not changed:

$$\frac{P_1}{T_1} = \frac{P_2}{T_2}$$

Common mistake

Remember that when making calculations involving temperatures (rather than temperature *changes*), you should always use kelvin.

Tool 3: Mathematics

Extrapolate and interpolate graphs

The considerable extrapolation seen in Figure B3.16 (and Figure B3.18) would not normally be recommended. However, in this case, we are not predicting the unknown behaviour of a gas all the way down to zero pressure. Rather, we are specifically asking the question: '*if* the gas continued to behave in this way, at what temperature would its pressure reduce to zero?'

Common mistake

It is widely stated that the volume of a gas increases when the temperature rises. But the volume of a gas will *only* increase if we allow it to, as in the next investigation.

WORKED EXAMPLE B3.5

Some air in a sealed container was heated from 60 °C to 92 °C. If its final pressure was 1.82×10^5 Pa, determine its pressure at 60 °C.

Answer

$$\frac{P_1}{T_1} = \frac{P_2}{T_2}$$

$$\frac{P_1}{(273 + 60)} = \frac{1.82 \times 10^5}{(273 + 92)}$$

$$P_1 = 1.7 \times 10^5 \text{ Pa}$$

Variation of gas volume with temperature at constant pressure (Charles' law)

The apparatus seen in Figure B3.17 can be used for this investigation. When the gas is warmed by thermal energy conducted from the surrounding water, it expands along the tube, keeping the pressure in the gas equal to the pressure outside from the air in the atmosphere.

Figure B3.18 shows the concluding graph from this investigation. It is similar to the pressure–temperature graph.

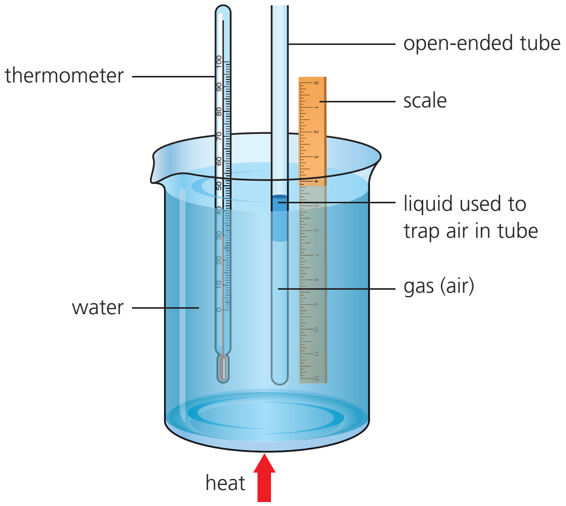


Figure B3.17 Simple apparatus for investigating Charles' law

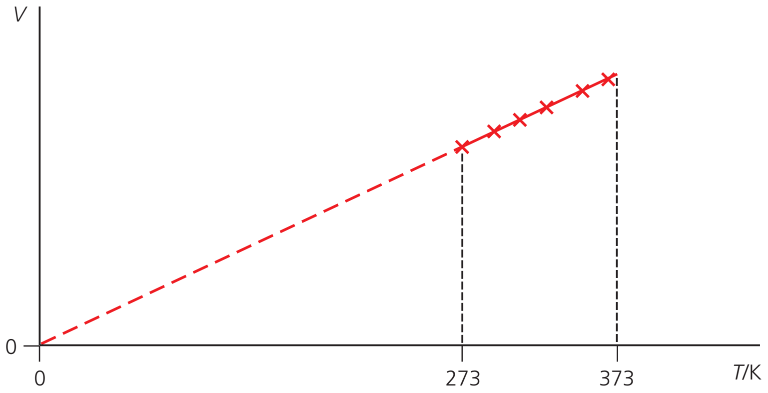


Figure B3.18 Gas volume is proportional to absolute temperature (K)

We can explain this relationship in terms of molecular behaviour. If the temperature of a gas is reduced, the molecules move more slowly, so that they would collide less frequently with the same area. But if the volume is also reduced, the rate of collision can be kept constant.

◆ **Charles' law** Volume of a fixed amount of gas is proportional to absolute temperature (at constant pressure).

For a fixed amount of gas at constant pressure: $V \propto T$ (K)

This is known as **Charles' law**.

If the pressure and/or temperature of a fixed amount of gas are changed from initial values of V_1 and T_1 to final values of V_2 and T_2, then, provided that the pressure has not changed:

$$\frac{V_1}{T_1} = \frac{V_2}{T_2}$$

WORKED EXAMPLE B3.6

In an experiment such as shown in Figure B3.17, the temperature of the gas was increased from 5.0 °C to 95.0 °C. If the initial volume was 3.1 cm³, calculate the final volume of the gas.

Answer

$$\frac{V_1}{T_1} = \frac{V_2}{T_2}$$

$$\frac{3.1}{(273 + 5.0)} = \frac{V_2}{(273 + 95)}$$

$V_2 = 4.1\text{ cm}^3$

9 After the pressure on a gas was increased from 1.0×10^5 Pa to 4.5×10^5 Pa its volume had become 280 cm³. What was its original volume, assuming that its temperature was constant?

10 The temperature of a gas was reduced from 80 °C to 20 °C, while keeping it at the same pressure. If the starting volume was 110 cm³, what was the final volume?

11 A fixed volume of gas at 310 K and a pressure of 1.2×10^5 Pa was heated in an oven. At what temperature (°C) will the pressure have risen to 1.9×10^5 Pa?

ATL B3A: Thinking skills

Applying key ideas and facts in new contexts

Using knowledge from this topic and from Topic A.2, explain how a hot air balloon can be made to rise and fall.

Hint: Consider the motion of air particles inside and outside the balloon, and the forces the particles apply to the balloon skin.

■ **Figure B3.19** Hot air balloons over Cappadocia, Turkey

Combined gas laws

SYLLABUS CONTENT

- ▶ The ideal gas law equation can be derived from the empirical gas laws for constant pressure, constant volume and constant temperature as given by: $\frac{PV}{T}$ = constant.
- ▶ The equation governing the behaviour of ideal gases as given by: $PV = nRT$.

◆ **Empirical** Based on observation or experiment.

The three separate **empirical** gas laws described above can be combined into one equation:

For a fixed amount of gas, $PV \propto T$, or:

$$\frac{PV}{T} = \textbf{constant} \quad \text{or} \quad \frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2}$$

Nature of science: Observations

Empirical science and theories

By describing the gas laws as *empirical*, we mean that they are based only on direct observation and experiment, using the human senses. The gas laws, as such, are not *theoretical*. Based *only* on the empirical results of these experiments, we can make reasonably accurate predictions about the way gases will behave under most circumstances. More detailed observations result in more accurate and improved predictions of real gas behaviour.

Clearly, empirical research is the basis of the *scientific method* and most scientific and technological advances. *Theoretical thinking* is then needed to explain the experimental results. The *ideal gas theory* later in this topic is a good example of a theory used to explain experimental results and make predictions. Theories can only be accepted if they have been widely tested by further experimentation.

In everyday language, the word *theory* is often used much more loosely, as a casual explanation, or even a guess. However, in science, a theory describes an explanation that has been extensively tested and is widely accepted.

More generally, *empiricism* is the view that all knowledge originates from our experiences.

WORKED EXAMPLE B3.7

Gas in a container at a temperature of 289 K was heated in an oven to a temperature of 423 K. If the volume expanded from 0.27 m^3 to 0.35 m^3, and the initial pressure was 1.1×10^5 Pa, determine the final pressure.

Answer

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2}$$

$$\frac{(1.1 \times 10^5) \times 0.27}{289} = P_2 \times \frac{0.35}{423}$$

$$P_2 = 1.2 \times 10^5\,\text{Pa}$$

So far, we have only discussed fixed amounts of gas. We need to expand the discussion to include *any* amount of gas.

Investigations can show that the pressure of a fixed volume of gas at constant temperature is proportional to the amount of gas. This is what we would expect from our molecular understanding: If we doubled the number of molecules in a fixed volume, then the frequency of collisions with the walls would double if they still had the same average speed.

This leads us to $PV \propto nT$, or:

$$PV = nRT$$

The constant R is known as the **universal (molar) gas constant.** It has the value **8.31 J K^{-1} mol^{-1}.**

◆ **Universal (molar) gas constant** The constant, R, that appears in the equation of state for an ideal gas ($pV = nRT$). $R = 8.31\,\text{J K}^{-1}\,\text{mol}^{-1}$.

R is the macroscopic equivalent to the *Boltzmann constant, k,* which was introduced in the microscopic interpretation of temperature in Topic B.1.

This equation is being presented here as a summary of empirical results. It can be used to accurately predict the physical behaviour of most gases under most conditions. For example, for a fixed amount of gas at constant temperature, the equation predicts an inverse proportionality between volume and pressure.

◆ **Equation of state for an ideal gas, $pV = nRT$:** Describes the macroscopic physical behaviour of ideal gases. Also called the **ideal gas law**.

◆ **Piston** A solid cylinder that fits tightly inside a hollow cylinder, trapping a fluid. Designed to move as a result of pressure differences.

Ideal gases (see below) obey this relationship perfectly, so it is called the **ideal gas law** (also called the **equation of state**).

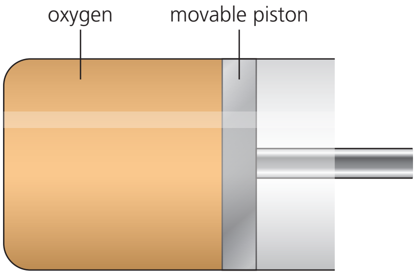
WORKED EXAMPLE B3.8

Some oxygen, O_2, is contained in the cylinder seen in Figure B3.20. No gas can move past the movable **piston**. The temperature of the gas is 298 K, it has a volume of 27 cm^3 and its pressure is equal to normal atmospheric pressure (1.0×10^5 Pa).

a Calculate the amount of gas in the cylinder.

b What was the mass of oxygen in the cylinder?

c The piston was suddenly pulled out increasing the volume to 32 cm^3. The pressure fell to 8.2×10^4 Pa. Determine the new temperature of the gas.



■ **Figure B3.20** Cylinder with movable piston

Answer

a $PV = nRT$

$(1.0 \times 10^5) \times (27 \times 10^{-6}) = n \times 8.31 \times 298$

$n = 0.0011$ mol

b $0.11 \times 32 = 3.5$ g

c $PV = nRT$

$(8.2 \times 10^4) \times (32 \times 10^{-6}) = 0.0011 \times 8.31 \times T$

$T = 287$ K

The expansion of the gas has resulted in a temperature fall of 11 K.

Alternatively, $\frac{PV}{T}$ = constant can be used to answer part **c**.

Inquiry process: Processing data

Processing data

A student carried out an experiment similar to that shown in Figure B3.12 and obtained the following results (Table B3.2). She used 6.1×10^{-4} mol of gas at a temperature of 21 °C.

Use these results to draw a straight line graph. One reading was an outlier and should not be considered when drawing the line of best fit. Which one? Use the gradient of the graph to determine a value for the molar gas constant.

■ **Table B3.2**

P / 10^5 Pa	V / cm^3
1.0	15.0
1.2	12.6
1.5	9.8
1.8	8.5
2.0	6.6
2.3	6.4
2.5	5.9
2.7	5.6

◆ ***PV* diagram** A graphical way of representing changes to the state of a gas during a thermodynamic process.

Pressure–volume diagrams

The state of a known amount of gas, as defined by values of its pressure, volume and temperature, can be identified as a point on a pressure–volume ***(PV)* diagram**.

For example, the point S in Figure B3.21 could represent one mole of a gas at a temperature of 300 K in a container of volume 1.0×10^{-2} m^3 and with a pressure of 2.5×10^5 Pa (data needed for later question).

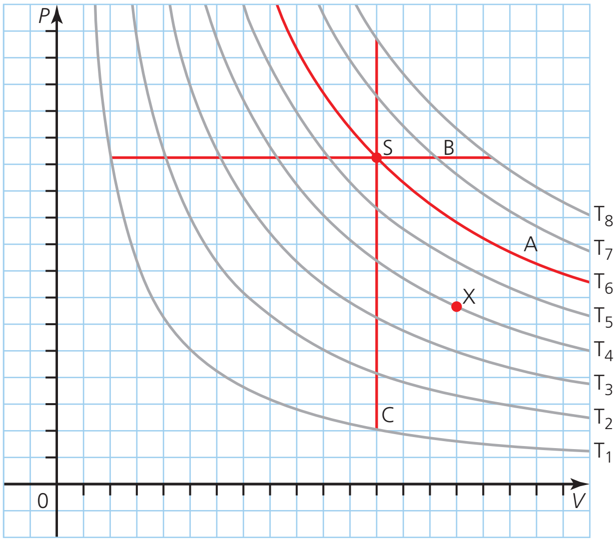


Figure B3.21 *PV* diagram

We know that there are three interconnected variables for a fixed amount of gas: pressure, volume and temperature. In order to represent these three variables on a graph with two axes, different temperatures are shown by the curved isothermal lines, labelled on Figure B3.21 by T_1 (lowest temperature) to T_8 (highest temperature).

Changes in the state of the gas can be represented by paths on *PV* diagrams. For example:

- Line A represents variations in pressure and volume at constant temperature.
- Line B represents variations in temperature and volume at constant pressure.
- Line C represents variations in pressure and temperature at constant volume.
- Any other path through point S (to point X, for example) will involve changing all three variables.

12 The volume of a scuba diving tank similar to that seen in Figure B3.22 is 11×10^{-3} m^3. The air inside is compressed to a pressure 210× atmospheric pressure.

Figure B3.22 Scuba diving

a What is that air pressure in pascals?

b What volume of air at atmospheric pressure was pumped into the tank?

c The water temperature decreases with depth below the surface. Would you expect that to significantly affect the pressure in the tank?

d As she breathes out, the diver releases air bubbles into the water. Explain why the bubbles get bigger as they rise towards the surface.

13 The volume of a gas in a cylinder such as seen in Figure B3.20 was 38 cm^3 when the temperature was 20 °C.

a The apparatus was heated. Explain why the piston moved to the right as thermal energy flowed into the gas through the cylinder.

b What was the temperature (°C) of the gas when the volume had increased to 50 cm^3?

14 Helium gas is widely used to fill balloons (Figure B3.23). If the density of the helium in a balloon is $2.0 \times 10^{-4}\,\mathrm{g\,cm^3}$, *estimate* its:

a volume

b mass

c amount in moles

d number of atoms.

■ **Figure B3.23** Helium-filled balloons

15 A container of gas has a volume of $120\,\mathrm{cm^3}$. At a temperature of 300 K the gas has a pressure of $1.5 \times 10^6\,\mathrm{Pa}$. Determine the amount of gas in the container.

16 How can you be sure that the isotherm T_5 on Figure B3.21 is representing a higher temperature than T_2?

17 Consider Figure B3.21. What changes must be made to a fixed amount of gas at point S on the diagram to move it to point X?

18 Represent the following changes to a fixed mass of gas on a *PV* diagram:

a a gas expanding at constant pressure, as its temperature increases

b a compression at constant temperature

c a gas cooled so that its volume and pressure decreased.

Modelling gas behaviour: ideal gases

SYLLABUS CONTENT

- ▶ Ideal gases are described in terms of the kinetic theory and constitute a modelled system used to approximate the behaviour of real gases.
- ▶ The change in momentum of particles due to collisions with a given surface gives rise to pressure in gases, and, from that analysis, pressure is related to the average translational speed of molecules: $P = \frac{1}{3}\rho v^2$

LINKING QUESTION

- How do the concepts of force and momentum link mechanics and thermodynamics?

This question links with understandings in Topics A.2, A.3.

◆ **Ideal gas** Gas which obeys the ideal gas law equation perfectly. The microscopic particle model of an ideal gas makes several important assumptions about the particles and their motions.

We now want to extend our microscopic kinetic theory of gases to include a mathematical treatment, which will predict macroscopic behaviour, including the equation $PV = nRT$.

First, we will provide a more detailed definition of an **ideal gas**. Although most gases are molecular, we will often use the term *particles* in order to be more general.

Top tip!

For a gas all at the same temperature, if collisions between gas particles resulted in a loss of kinetic energy, that would mean that the gas would keep getting colder. Collisions of gas particles with particles in the containing walls will result in a transfer of kinetic energy only if there is a temperature difference (thermal conduction).

◆ **Average value** Any single number used to represent a quantity which is varying.

◆ **Range (data)** Spread of data from smallest to largest values.

◆ **Anomalous** Different from the pattern of other similar observations.

◆ **Outlier** A value which is significantly different from the others in the same data set.

◆ **Mean** A certain type of average: the sum of all of the numbers divided by the number of values involved.

Assumptions about the particles in an ideal gas:

- The gas contains a very large number of identical particles.
- The volume of the particles is negligible compared with the total volume occupied by the gas.
- The particles are moving in random directions, with a wide variety of speeds.
- There are no forces between the particles, except when they collide. Because there are no forces, the particles have no (electrical) potential energy. This means that any changes of internal energy of an ideal gas are assumed to be only in the form of changes of random kinetic energy.
- The motion of the particles obeys Newton's laws of motion.
- All collisions between particles are elastic. (This means that the total kinetic energy of the particles remains constant at the same temperature.)

Tool 3: Mathematics

Calculate mean and range

We have often referred to the **average values** of particle speeds but have not really explained what that means.

Consider a more straightforward example: the following times (s) recorded for a ball to roll down a particular slope: 15, 12, 11, 21, 14, 11, 14, 13.

The **range** of this data is from 11 s to 21 s.

The value of 21 s is **anomalous** and is inconsistent with the other values. It may be described as an **outlier**. It may have been a mistake and should be excluded from calculations, unless checked and confirmed.

An *average value* is any single number that has been chosen to represent a range of values. There are several possibilities, including the central (*median*) value (13), or the most common value (11 or 14). However, in physics, average values are usually **mean** values. A *mean value* is obtained by adding all the values and dividing by the number of values. In this example, the mean is 12.9, which may be better limited to 2 significant figures (13).

In this topic, average particle speeds cannot be determined as straightforward means. As we shall see, (average speed)^2 is calculated from average kinetic energies.

Mathematical model for gas behaviour

We will start by considering just one particle in a rectangular box. See Figure B3.24. The particle has a mass m and is moving with a velocity v perpendicularly towards the end wall.

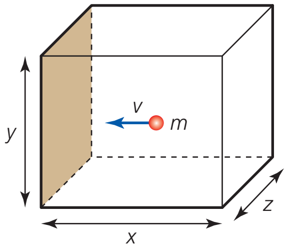


Figure B3.24 One particle in a box

After the particle has an elastic collision with the wall it will return along the same path, with velocity $-v$.

Its change of velocity is: $v - (-v) = 2v$

average force on end wall, F = change of momentum $= m \times 2v = 2mv$

time interval between collisions of the particle with the same wall, $t = \frac{2x}{v}$

$$\text{average force on end wall} = \frac{\text{change of momentum}}{\text{time between collisions}} = \frac{2mv}{\left(\frac{2x}{v}\right)} = \frac{mv^2}{x}$$

$$\text{average pressure on end wall, } P = \frac{\text{average force}}{\text{area}} = \frac{\left(\frac{mv^2}{x}\right)}{yz} = \frac{mv^2}{xyz} = \frac{mv^2}{V}$$

where V is the total volume of the box $(= xyz)$.

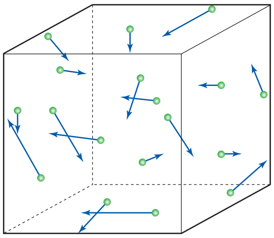


Figure B3.25 Gas molecules moving around at random in a container

However, of course, the container may not be rectagular, and all gases have a very large number of particles moving in random directions with different speeds. See Figure B3.25.

- If there are N particles in the box (all moving perpendicular to the end wall):

  average pressure on end wall becomes: $\frac{Nmv^2}{V}$
- If the particles are moving in random directions:

  average pressure on end wall becomes:

  $$\frac{\left(\frac{1}{3}Nmv^2\right)}{V}$$

  because there are three perpendicular directions, so that on average $\frac{1}{3}$ of the velocities are directed in any one of these directions.
- If the particles have different speeds:

  v^2 is the average value of their speeds-squared.
- If the container is a different shape:

  it can be shown that the shape of the container does not affect the pressure.

Common mistake

This equation contains two Vs. One upper case (volume), and one lower case (velocity). It is easy to get them confused.

So, finally we have an equation which links the macroscopic properties of a gas (pressure and volume) to its microscopic properties (the number of particles, their mass and their speed):

Pressure of an ideal gas: $P = \frac{1}{3}Nm\frac{v^2}{V}$

This long derivation need not be remembered, but it should be understood. It provides an all-important link between the theories of particle behaviour and measurements that can easily be made in a laboratory.

Since Nm is the total mass of the gas:

$$\frac{Nm}{V}$$

is the density of the gas, ρ, leading to an alternative expression:

DB

Pressure of an ideal gas: $\boldsymbol{P = \frac{1}{3}\rho v^2}$

Tool 3: Mathematics

Check an expression using dimensional analysis of units

The units of an expression on the left-hand side of an equation must be the same as the units of an expression on the right-hand side of the equation.

This can be used to check if a suggested equation could be correct. It is a simplified version of what is known as **dimensional analysis**.

To check that there is no obvious mistake in the equation $P = \frac{1}{3}\rho v^2$, we can see if the units on both sides of the equation are the same.

The units of pressure are pascals, which are the same as $\mathrm{N\,m^{-2}}$. We also know that newtons are equivalent to $\mathrm{kg\,m\,s^{-2}}$ ($F = ma$). So that, the units of pressure can be reduced to the SI base units: $\mathrm{kg\,m\,s^{-2}/m^{-2} = kg\,m^{-1}\,s^{-2}}$.

On the right-hand side of the equation: $\frac{1}{3}$ has no units, ρ has the units $\mathrm{kg\,m^{-3}}$ and v has the units $\mathrm{m\,s^{-1}}$. Combining these, we get: $\mathrm{kg\,m^{-1}\,s^{-2}}$ which is the same as on the left-hand side of the equation. This check has found no obvious mistake in the equation.

◆ **Dimensional analysis** Method of checking if an equation may be correct. The units (dimensions) of all terms should be the same.

WORKED EXAMPLE B3.9

a Determine a value for the average speed of nitrogen molecules at normal air pressure (1.0×10^5 Pa) if a mass of 1.4 g of the gas was in a container of volume 1200 cm^3.

b How many moles of nitrogen were in the container (see Table B3.1)?

c Calculate the temperature of the gas.

d Determine a value for the average molecular speed if the temperature rose to 350 K in the same container.

Answer

a $\rho = \frac{1.4 \times 10^{-3}}{1200 \times 10^{-6}} = 1.17\,\mathrm{kg\,m^{-3}}$

$P = \frac{1}{3}\rho v^2$

$1.0 \times 10^5 = \frac{1}{3} \times 1.17 \times v^2$

$v = 5.1 \times 10^2\,\mathrm{m\,s^{-1}}$

(This value is obtained from the square root of the average value of speeds². This is not exactly the same as the value of the average speed.)

b $\frac{1.4}{28} = 5.6 \times 10^2\,\mathrm{mol}$

c $PV = nRT$

$(1.0\times 10^5) \times (1200 \times 10^{-6}) = 0.050 \times 8.31 \times T$

$T = 289\,\mathrm{K}$

d We can use $\frac{P_1}{T_1} = \frac{P_2}{T_2}$ to calculate the new pressure:

$\frac{1.0 \times 10^5}{289} = \frac{P_2}{350}$

$P_2 = 1.21 \times 10^5\,\mathrm{Pa}$

Then, $P = \frac{1}{3}\rho v^2$

$1.21 \times 10^5 = \frac{1}{3} \times 1.17 \times v^2$ (density is unchanged)

$v = 560\,\mathrm{m\,s^{-1}}$

19 Make a list of all the basic physics principles used to derive the equation $P = \frac{1}{3}\rho v^2$, as shown above.

20 What pressure would be created by 5.0×10^{23} ideal gas molecules, each of mass 5.3×10^{-26} kg in a volume of 0.010 m^3 if the temperature of the gas was such that the average speed of the molecules was 500 $\mathrm{m\,s^{-1}}$?

21 An ideal gas has a density of 2.4 $\mathrm{kg\,m^{-3}}$. If it creates a pressure of 1.5×10^5 Pa on its container, determine a value for the average speed of its molecules.

22 At room temperature the average speed of oxygen molecules is 500 $\mathrm{m\,s^{-1}}$. What pressure will these molecules create if the density of oxygen was the same as in air (1.3 $\mathrm{kg\,m^{-3}}$)? Assume that oxygen behaves as an ideal gas.

23 Three particles have speeds 440 $\mathrm{m\,s^{-1}}$, 480 $\mathrm{m\,s^{-1}}$ and 520 $\mathrm{m\,s^{-1}}$.

- **a** What is their average speed?
- **b** What is the square of their average speed?
- **c** What is the average of their speeds-squared?
- **d** What is the square root of the average of their speed squared?

Nature of science: Models

Randomness

A key feature of the motions of the particles in an ideal gas is that they are 'random'. But what exactly does 'random' mean? The word has various uses throughout science and more generally, often with slight differences in meaning. For example, we might say that the result of throwing a six-sided die is random because we cannot predict what will happen, although we probably appreciate that there is a one-in-six chance of any particular number ending up on top. In this case, all outcomes should be equally likely. Another similar example could be if we were asked to 'pick a card at random' from a pack of 52.

■ **Figure B3.26** Six-sided dice

Sometimes we use the word random to suggest that something is unplanned. For example a tourist might walk randomly around the streets of a town.

Unpredictability is a key feature of random events and that certainly is a large part of what we mean when we say a gas molecule moves randomly. All possible directions of motion may be equally likely, but the same cannot be said for speeds. Some speeds are definitely more likely than others. For example, at room temperature a molecular speed of $500\,\mathrm{m\,s^{-1}}$ is much more likely than one of $50\,\mathrm{m\,s^{-1}}$. Similarly, when we refer to random kinetic energies of molecules, we mean that we cannot know or predict the energy of individual molecules, although some values are more likely than others. But there is a further meaning: we are suggesting that individual molecules behave independently and that their energies are disordered.

Perhaps surprisingly, in the kinetic theory, the random behaviour of a very large number of individual molecules on the microscopic scale leads to complete predictability in our everyday macroscopic world. Similar ideas occur in other areas of physics, notably in radioactive decay (Topic E.1), where the behaviour of an individual atom is unknowable, but the total activity of a radioactive source is predictable. Of course, insurance companies, betting companies and casinos can make good profits by understanding the statistics of probability, without being too concerned about individual events.

LINKING QUESTIONS

- How can gas particles of high kinetic energy be used to perform work?
- How does a consideration of the kinetic energy of molecules relate to the development of the gas laws?

These questions link to understandings in Topic B.4 (HL).

Internal energy of an ideal gas

SYLLABUS CONTENT

- ▶ The relation between the internal energy, U, of an ideal monatomic gas and the number of molecules, or amount of substance as given by: $U = \frac{3}{2}Nk_BT$ or $U = \frac{3}{2}RnT$.
- ▶ The equations governing the behaviour of ideal gases are given by: $PV = Nk_BT$ and $PV = nRT$.

◆ Internal energy of an ideal monatomic gas, U The sum of the random translational kinetic energies of all the molecules.

The **internal energy of an ideal (monatomic) gas** is the total random *translational* kinetic energy of its particles. Ideal monatomic gas particles do not have any potential energies, or vibrational or rotational kinetic energies.

Internal energy is given the symbol U. It can be calculated by multiplying the number of particles by their average random translational kinetic energy:

$$U = N \times \frac{1}{2}mv^2$$

Comparing $PV = nRT$ with $PV = \frac{1}{3}Nmv^2$ we see that: $nRT = \frac{1}{3}Nmv^2$.

But $\frac{1}{3}Nmv^2$ can be rewritten as $\frac{2}{3} \times \frac{1}{2}Nmv^2$ or $\frac{2}{3}U$ so: $nRT = \frac{2}{3}U$.

Or:

internal energy of an ideal (monatomic) gas: $\boldsymbol{U = \frac{3}{2}nRT}$ (using macroscopic quantities R and T.)

Most gases consist of molecules, rather than atoms. They still approximate well to the macroscopic behaviour of an ideal gas, but their internal energy is more complicated, because of the vibrational and rotational kinetic energies of their molecules.

The forces between particles in solids and liquids makes calculating their internal energies much more complicated.

WORKED EXAMPLE B3.10

Calculate the total internal energy of one mole of a monatomic ideal gas at 300 K.

Answer

$U = \frac{3}{2}nRT = 1.5 \times 1.0 \times 8.31 \times 300 = 3.7 \times 10^3$ J

If we divide both sides of the highlighted equation for U by the number of particles, N, we get the average random translational kinetic energy of a single atom in an ideal gas, $\overline{E}_k$ $(=U/N)$:

$\overline{E}_k = \frac{3}{2}\frac{nRT}{N}$, but we know that $\frac{n}{N}$ = the Avogadro constant, N_A, so that: $\overline{E}_k = \frac{3}{2}\frac{RT}{N_A}$.

$\frac{R}{N_A}$ is the molar gas constant divided by the number of particles in a mole.

It is called the Boltzmann constant, k_B (unit: J K^{-1}), which was intoduced in Topic B.1:

$$k_B = \frac{8.31}{6.02 \times 10^{23}} = 1.38 \times 10^{-23}\,\text{J K}^{-1}$$

Boltzmann constant: $k_B = \frac{R}{N_A}$

This leads us to the important relationship between temperature (K) and the average random translational kinetic energy of particles, that we first met in Topic B.1, and which is repeated here:

$$\boldsymbol{E_k = \frac{3}{2}k_BT} \left(= \frac{1}{2}mv^2\right)$$

This equation is *not* restricted to the particles in ideal gases. When particles collide / interact, translational kinetic energy is exchanged, so that the particles in all gases in good thermal contact with each other will eventually all have the same average *translational* kinetic energy. In other words:

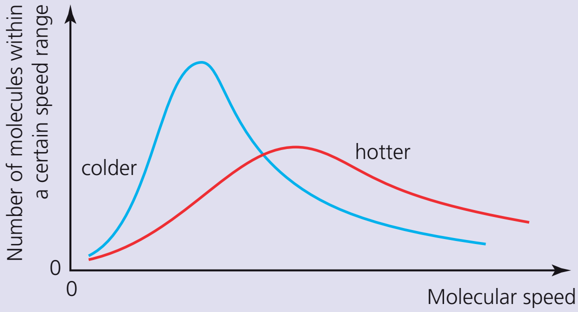
At the same temperature, all gases contain molecules with the same average random translational kinetic energy.

ATL B3B: Research skills

Providing a reasoned argument to support conclusions

Kinetic theory and the distribution of molecular speeds

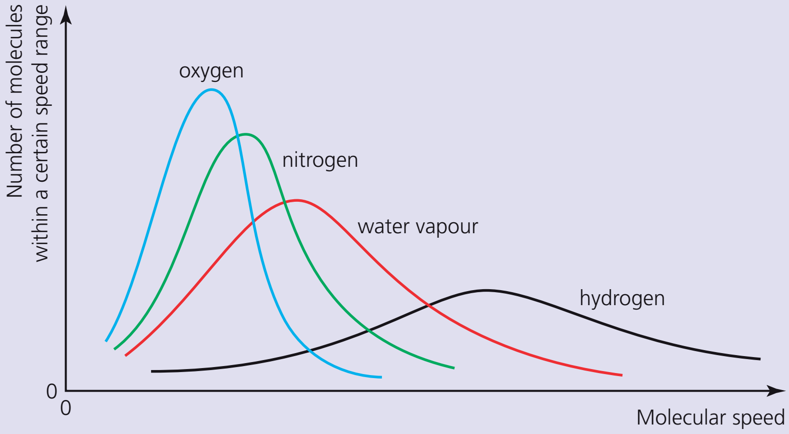
Figure B3.27 shows the range and distribution of molecular speeds in a typical gas, and how it changes as the temperature increases. This is known as the Maxwell–Boltzmann distribution.



■ **Figure B3.27** Typical distributions of molecular speeds in a gas

Note that there are no molecules with zero speed and few with very high speeds. Molecular speeds and directions (that is, molecular velocities) are continually changing as the result of intermolecular collisions. As we have seen, higher temperature means higher kinetic energies and therefore higher molecular speeds. But the range of speeds also broadens and so the peak becomes lower, keeping the area under the graph, constant.

Figure B3.28 illustrates the ranges of molecular speeds for different gases at the same temperature. It shows that, as we have noted before, at the same temperature less massive molecules have higher speeds than more massive molecules.



■ **Figure B3.28** Distribution of molecular speeds in different gases at the same temperature

For a given sample of a gas, the area under the curves in Figure B3.27 is the same. Explain why this must be true.

Understanding that $k_B = \frac{R}{N_A} = \frac{Rn}{N}$, internal energy, $U = \frac{3}{2}nRT$ can now be expressed in terms of k (rather than R) if we prefer:

Internal energy of an ideal monatomic gas: $\boldsymbol{U = \frac{3}{2}Nk_BT}$ (using the microscopic quantities N and k_B.)

We can also re-write the ideal gas equation ($PV = nRT$) in terms of k (rather than R):

$\boldsymbol{PV = Nk_BT}$

WORKED EXAMPLE B3.11

a A sample of an ideal monatomic gas in a closed container at 18 °C has a total internal energy of 380 J. Determine how many particles (atoms) it contains.

b If the container has a volume of 435 cm^3, calculate the pressure of the gas.

Answer

a $U = \frac{3}{2} N k_B T$

$380 = 1.5 \times N \times (1.38 \times 10^{-23}) \times (273 + 18)$

$N = 6.3 \times 10^{22}$

b $PV = N k_B T$

$P \times (435 \times 10^{-6}) = (6.3 \times 10^{22}) \times (1.38 \times 10^{-23}) \times (273 + 18)$

$P = 5.8 \times 10^5$ Pa

Real gases compared to ideal gases

SYLLABUS CONTENT

- Temperature, pressure and density conditions under which an ideal gas is a good approximation to a real gas.

◆ **Real gases** Modelling of gas behaviour is idealized. Real gases do not behave exactly the same as the model of an ideal gas.

An ideal gas is impossible to achieve, but we have seen that **real gases** obey the ideal gas equation ($PV = nRT$) under most circumstances. However, this will not be true if there are significant differences from the stated assumptions about the particles in an ideal gas, such as:

- At high densities and pressures the particles will be closer together than has been assumed. The forces between the particles may no longer be negligible.
- At low temperatures the forces between particles will have a greater effect because the particles are moving slower. Most real gases will also turn into liquids and solids if the temperature is low enough.

Figure B3.29 Liquid nitrogen has a boiling point of 77 K

Most real gases behave like ideal gases unless their pressure or density is very high, or the temperature is very low.

24 a Calculate the average random translational kinetic energy of oxygen (O_2) molecules at 300 K.

b Oxygen molecules each have a mass of 5.31×10^{-26} kg. Determine their average speed at 300 K.

c Explain why an average carbon dioxide (CO_2) molecule travels slower than oxygen molecules (O_2) at the same temperature.

25 a Determine how much energy is needed to raise the temperature of two moles of a monatomic gas from 20 °C to 50 °C.

b Why will more energy be needed to raise two moles of a *molecular* gas through the same temperature rise?

26 At what temperature will 1.0×10^{23} atoms of an ideal gas have a total internal energy of 1000 J?

27 Some nitrogen gas was cooled from a temperature of 300 K to 100 K.

a Estimate by what percentage its pressure was reduced.

b State two assumptions you had to make in order to answer part **a**.

c Explain why the actual final pressure was less than that predicted by theory.

28 What is the pressure of an ideal gas which has a temperature of 47 °C and contains 4.2×10^{25} particles in a volume of 0.037 m^3?

29 a Explain why the internal energy of an ideal gas can be determined from: 1.5 × pressure × volume.

b Discuss why temperature does not appear in this equation.

c What volume (cm^3) of an ideal gas has an internal enegy of 10 J and exerts a pressure of twice normal atmospheric pressure?

30 The *molar heat capacity* of a gas is a similar concept to specific heat capacity. It equals the thermal energy needed to raise the temperature of one mole of the gas by one kelvin.

a Calculate the molar heat capacity of argon. (Assume that the gas is kept in a constant volume.)

b Explain why the molar heat capacity of a *molecular* gas, oxygen for example, will be greater.

c Why will the molar heat capacity of argon be greater than your answer to part **a** if it is allowed to exand?

TOK

Knowledge and the knower

- How do our interactions with the material world shape our knowledge?

Models always have limitations

We cannot directly use our human perception to help understand the behaviour of molecules in a gas. We need a *model* to help to understand the situation. In this topic we have presented both a visual and a mathematical model of an 'ideal' gas. These models are extremely useful and accurate predictors of the behaviour of real gases, but they are not perfect.

Can modelling any system which is too big or too small to be seen ever be 'good enough' to count as true knowledge? Can we ever be sure that a model is a true representation of reality, if we can never observe events directly? Added to which, one purpose of a model is to simplify reality, to make it easier to understand.

Or will there always be some doubt? And, if the model is useful, does it really matter if the model is a 'true' representation of something that we cannot observe directly anyway? Can a model of the solar system, for example, be considered to be knowledge of a higher level (than the model of an ideal gas), because we can directly observe and record the motion of individual planets?

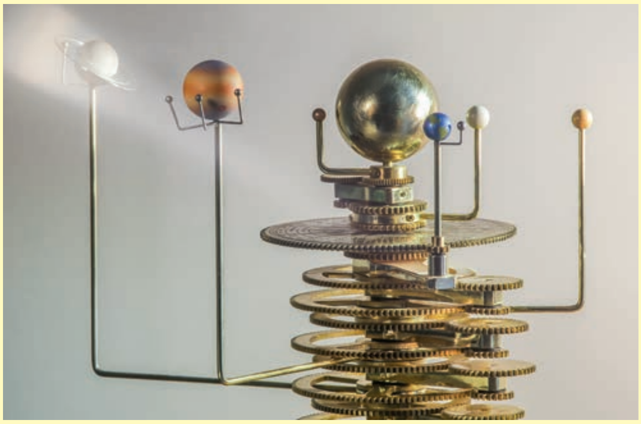


Figure B3.30 Simple model of the Solar System

B.4 Thermodynamics

◆ **Heat engine** Device that uses the flow of thermal energy to do useful work.

◆ **Thermodynamics** Branch of physics involving transfers of thermal energy to do useful work.

Guiding questions

- How can energy transfers and storage within a system be analysed?
- How can the future evolution of a system be determined?
- In what way is entropy fundamental to the evolution of the Universe?

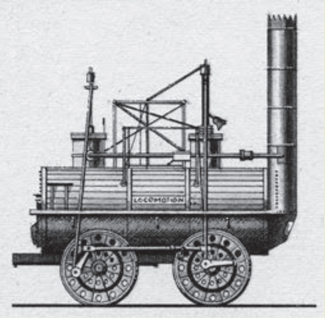
Nature of science: Global impact of science

Heat engines

The invention of devices that could continuously use the thermal energy transferred from a burning fuel to do useful mechanical work changed the world completely. No longer did people, or animals, have to do such hard work – they could get engines to do it for them, and much quicker than they could do the same work themselves.

The idea that burning a fuel to heat water to make steam, which could then be used to make something move, had been understood for a very long time. But, using this in a practical way was much more difficult, and it was not until the early eighteenth century that the first commercial steam engines were produced. It was about 100 years afterwards when George Stephenson built the engine 'Locomotion' (See Figure B4.1) for the first public steam railway, opened in the UK in 1825.

About 200 years later, things are very different. We live in a world that is dominated by **heat engines** (devices that get useful mechanical work from a flow of thermal energy). We are surrounded by all sorts of different engines – in cars, boats, trains, planes, factories and power stations producing electricity (see Figure B4.2).



■ **Figure B4.1** The 'Locomotion'

All these engines need a transfer of thermal energy from fuels in order to work. The fuels are usually fossil fuels. It is difficult to overstate the importance that these devices have had in modern life, because without them our lives would be very different. Of course, we are now also very much aware of the problems associated with the use of heat engines, as discussed in Topic B.2: limited fossil-fuel resources, inefficient devices, pollution and global warming.

This topic describes the process of using thermal energy to do useful mechanical work in heat engines. This branch of physics is known as **thermodynamics**. Although thermodynamics grew out of a need to understand heat engines, it has much wider applications. The study of thermodynamics leads to a better understanding of key scientific concepts such as internal energy, heat, temperature, work and pressure, and how they are all connected to each other and to the microscopic behaviour of particles.

■ **Figure B4.2** Using hot gases in heat engines

HL ONLY

LINKING QUESTION

- What paradigm shifts enabling changes to human society, such as harnessing the power of steam, can be attributed to advancements in physics understanding? (NOS)

◆ **Closed system** Allows the free flow of thermal energy, but not matter.

◆ **Reservoir (thermal)** Part of the surroundings of a thermodynamic system that is kept at approximately constant temperature and is used to encourage the flow of thermal energy.

◆ **Isobaric** Occurring at constant pressure. $\Delta P = 0$.

In this topic we will concentrate our attention on understanding the basic principles of the processes that involve the volume increase (expansion) of fixed masses of gases.

In the rest of this chapter, and throughout physics, there are many references to thermodynamic ‘systems’ and ‘surroundings’. Before going any further, we should make sure that these simple and widely used terms are clearly understood.

A *system* is simply the thing that we are studying or talking about. In this topic it will be a gas.

In this topic we will be discussing **closed systems**, in which energy can be transferred into or out of the system as heat, or work, but no mass can be transferred in or out. Compare this to an *isolated system*: one in which neither mass nor energy can be transferred in or out. For example, when discussing the conservation of momentum in Topic A.2, we were referring to an *isolated* system.

The *surroundings* are everything else – the gas container and the rest of the Universe. Sometimes the surroundings are called the *environment*. If we wish to suggest that a part of the surroundings was deliberately designed for thermal energy to flow into it or out of it, we may use the term **(thermal) reservoir**.

A thermodynamic system can be as complex as a rocket engine, planet Earth or a human body, but in this topic, we will develop understanding by considering the behaviour of gases in heat engines.

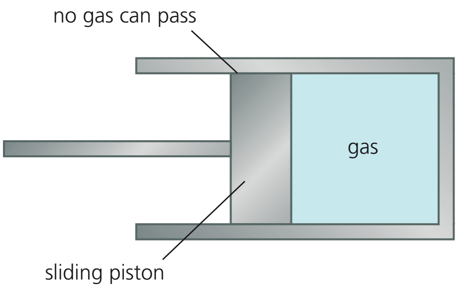
Work done when a gas expands (or is compressed)

SYLLABUS CONTENT

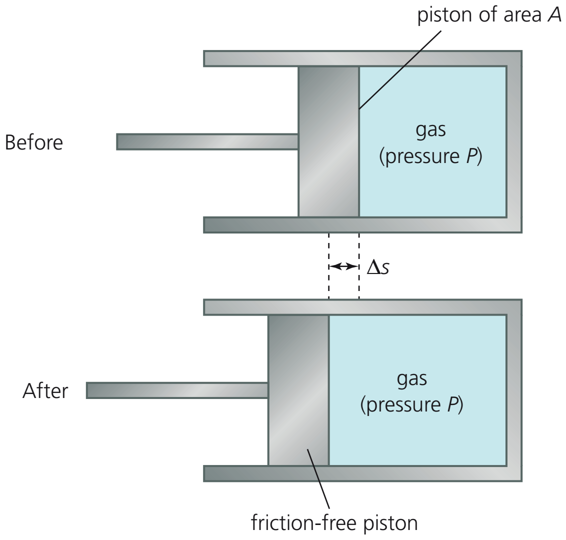
▶ The work done by or on a closed system, as given by $W = P\Delta V$, when its boundaries are changed, can be described in terms of pressure and changes of volume of the system.

For simplicity, the thermodynamic system that we are considering is often shown as a gas in a regularly shaped cylindrical container, constrained by a gas-tight piston that can move without friction (Figure B4.3).

First consider the example of a gas expanding so that there is no change in pressure. (This is called an **isobaric** change, as described later.) If the gas trapped in the cylinder in Figure B4.4 is given thermal energy it will exert a resultant force on the piston (because the pressure in the cylinder is momentarily higher than the pressure from the surroundings) and the piston will move outwards as the gas expands, keeping the pressures equal.



■ **Figure B4.3** Gas in a cylinder with a movable piston



■ **Figure B4.4** Gas expanding in a cylinder

We say that work has been done *by* the gas in pushing back the surrounding air (remember that we are assuming that there is no friction).

work done by gas = force × distance moved in direction of force, or

work done by gas = $(PA)\Delta s$ (because force = pressure × area)

Because change of volume, $\Delta V = A\Delta s$:

work done when a gas changes volume at constant pressure, $W = P\Delta V$

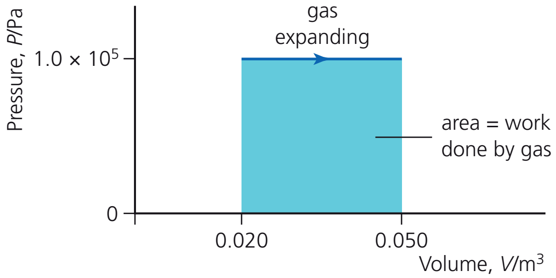


Figure B4.5 Work done during expansion of an ideal gas

If work is done *on* the gas to reduce its volume, ΔV and the work done, W, will have negative values.

We introduced PV diagrams in Topic B.3 and they are very useful in representing changes in the **state of a gas** during thermodynamic processes. An example is shown in Figure B4.5, which shows the expansion of a gas from 0.020 m³ to 0.050 m³ at a constant pressure of 1.0×10^5 Pa.

The work done *by* the gas in expansion in this example:

$W = P\Delta V = (1.0 \times 10^5) \times (0.050 - 0.020) = 3.0 \times 10^3$ J

Note that this calculation to determine the work done, $P\Delta V$, is numerically equal to calculating the area under the PV diagram. This is true for all thermodynamic processes, regardless of the shape of the graph, and this is one reason why PV diagrams are so widely used in thermodynamics to represent various processes. Figure B4.6 shows two further examples.

The work done when a gas changes pressure and/or volume can be determined from the area under a PV diagram.

◆ **Work done when a gas changes volume, W** Work is done by a gas when it expands (W is positive). Work is done on a gas when it is compressed (W is negative). At constant pressure $W = P\Delta V$. If the pressure changes, the work done can be determined from the area under a PV diagram.

◆ **State of a gas** Specified by quoting the pressure, P, temperature, T, and volume, V, of a known amount, n, of gas.

WORKED EXAMPLE B4.1

Determine values for the work done in the two changes of state represented in Figure B4.6.

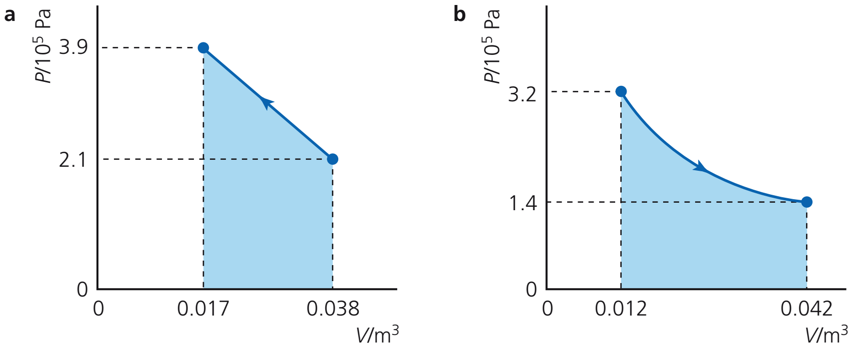


Figure B4.6 Determining areas under pressure–volume graphs

Answer

a $W = P\Delta V =$ area under graph

$$= \left[\frac{1}{2} \times (3.9 - 2.1) \times 10^5 \times (0.038 - 0.017)\right] + \left[(0.038 - 0.017) \times 2.1 \times 10^5\right] = 6.3 \times 10^3\ \text{J}$$

The work is done *on* the gas as it is compressed into a smaller volume.

b In this example, work is done *by* the gas as the volume increases. Because the graph is curved, the area underneath it must be estimated, as explained below.

Tool 3: Mathematics

Interpret features of graphs: areas under the graph

The areas under curved graphs can be estimated by drawing a rectangle that is judged by eye to have the same area, as shown by the example in Figure B4.7.

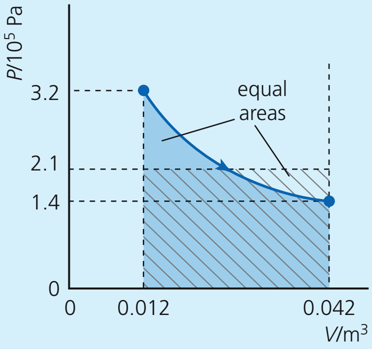


Figure B4.7 Estimating the area under a curved graph

W = area under graph ≈ $(2.1 \times 10^5) \times (0.042 - 0.012) = 6.3 \times 10^3$ J

1 The volume of a gas expanded from 43 cm³ to 51 cm³ while its pressure remained constant at 1.1×10^5 Pa.
   a Calculate the work done.
   b State whether work was done on the gas, or by the gas.

2 0.84 J of work was done on a gas of volume 7.6×10^{-5} m³. If the pressure was constant at 1.4×10^5 Pa, calculate the new volume.

3 Figure B4.8 shows how the volume of a gas changed as the pressure on it was increased. Show that the work done during the expansion was approximately 230 J.

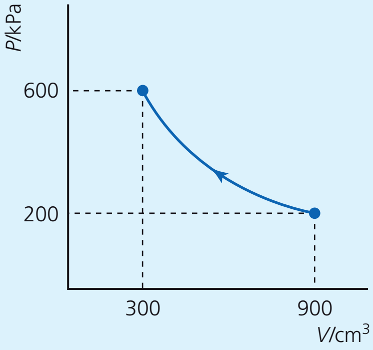


Figure B4.8 *PV* graph for a gas

First law of thermodynamics

SYLLABUS CONTENT

- The first law of thermodynamics, as given by $Q = \Delta U + W$, results from the application of conservation of energy to a closed system and relates the internal energy of a system to the transfer of energy as heat and as work.
- The change in internal energy, as given by $\Delta U = \frac{3}{2}Nk_B\Delta T$ is related to the change of its temperature.

If an amount of thermal energy, $+Q$, is transferred *into* a system, such as that seen in Figure B4.3, then, depending on the particular circumstances, the gas may gain internal energy, $+\Delta U$, and/or the gas will expand and do work *on* the surroundings, $+W$. We can use the *principle of conservation of energy* (from Topic A.2) to describe how these quantities are connected:

Thermal energy supplied to a gas:

$$Q = \Delta U + W$$

◆ **First law of thermodynamics** If an amount of thermal energy, $+Q$, is transferred into a system, then the system will gain internal energy, $+\Delta U$, and/or the system will expand and do work on the surroundings, $+W$: $Q = \Delta U + W$.

This important equation, known as the **first law of thermodynamics**, covers all the possibilities of expanding or compressing gases, and/or supplying or removing thermal energy from a system.

Common mistake

Students often get confused over the signs used in this equation. They are restated here:

- Thermal energy transferred *into* the gas will be given positive values: $+Q$; thermal energy *removed* from the gas will be considered to be negative: $-Q$.
- An *increase* in the internal energy of a gas will be given positive values: $+\Delta U$; a *decrease* in internal energy will be considered to be negative: $-\Delta U$.
- Work done *by* the gas during *expansion* will be given positive values: $+W$; work done *on* the gas during *compression* will be given negative values: $-W$.

WORKED EXAMPLE B4.2

80 J of work was done by a gas when 120 J of thermal energy was transferred to it. Determine the change in internal energy of the gas.

Answer

$$Q = \Delta U + W$$

$$(+120) = \Delta U + (+80)$$

$\Delta U = (+40)$ J. The positive sign shows that the internal energy increased.

WORKED EXAMPLE B4.3

150 J of work was done when a gas was compressed. At the same time, its internal energy increased by 50 J. Calculate how much thermal energy flowed into, or out of, the system during this process.

Answer

$Q = \Delta U + W$

$Q = (+50) + (-150) = (-100)$ J of thermal energy was transferred. The negative sign shows that the transfer was *out of* the gas.

Top tip!

When applying the first law of thermodynamics, show the energies as shown in the above worked example: in brackets with either a + or – sign.

Changes in internal energy of an ideal monatomic gas

We have seen in Topic B.3 that the internal energy of an ideal monatomic gas can be calculated from:

$$U = \frac{3}{2}Nk_BT = \frac{3}{2}nRT$$

This means that:

changes in internal energy, ΔU, of an ideal monatomic gas can be calculated from:

$$\boldsymbol{\Delta U = \frac{3}{2}Nk_B\Delta T = \frac{3}{2}nR\Delta T}$$

WORKED EXAMPLE B4.4

Consider the previous worked example. If the gas contained 3.4×10^{23} particles, determine the temperature rise. Assume the gas was an ideal monatomic gas.

Answer

$$\Delta U = \frac{3}{2}Nk_B\Delta T$$

$$50 = 1.5 \times (3.4 \times 10^{23}) \times (1.38 \times 10^{-23}) \times \Delta T$$

$$\Delta T = 7.1\text{ K}$$

4 200 J of work was done when a gas expanded at constant temperature. How much thermal energy was transferred?

5 70 J of work was done while a gas was compressed. At the same time 30 J of thermal energy was transferred out of the gas. Calculate the change in internal energy.

6 50 J of thermal energy was removed from a gas and its internal energy decreased by 10 J.
- **a** How much work was done?
- **b** Was the work done on the gas, or by the gas?

7 An ideal monatomic gas containing 3.4×10^{24} particles was heated from 297 K to 348 K.
- **a** What was the change in internal energy of the gas?
- **b** If, in this time, 5600 J of thermal energy was supplied to the gas, what was the amount of work done by, or on, the gas?

8 The internal energy of 3.2 moles of some helium gas, at an initial temperature of 302 K, rose by 540 J.
- **a** What was the final temperature of the gas?
- **b** If, at the same time, the gas expanded from a volume of 220 cm^3 to 380 cm^3 at constant pressure of 2.3×10^5 Pa, how much work was done?
- **c** How much thermal energy was transferred?

9 During an investigation of how the volume of a fixed mass of air changed with temperature at constant atmospheric pressure (1.0×10^5 Pa), the volume increased from 2.3 cm^3 to 3.1 cm^3.
- **a** Calculate how much work was done.
- **b** Was the work done on the gas, or by the gas?
- **c** The expansion of the air happened when 0.150 J of thermal energy was transferred into the system. Determine the change in internal energy of the gas.

10 1.82 J of thermal energy was supplied to a gas, increasing its pressure steadily from 1.0×10^5 Pa to 1.2×10^5 Pa. During this time the volume was increased steadily from 35 cm^3 to 45 cm^3.
- **a** Represent this process on a *PV* diagram.
- **b** Indicate the work done during this process on your drawing.
- **c** Calculate the work done.
- **d** Determine the change in internal energy of the gas.

11 Consider Figure B4.9. Work has been done to change the state of a gas from point A to point B on the *PV* diagram.

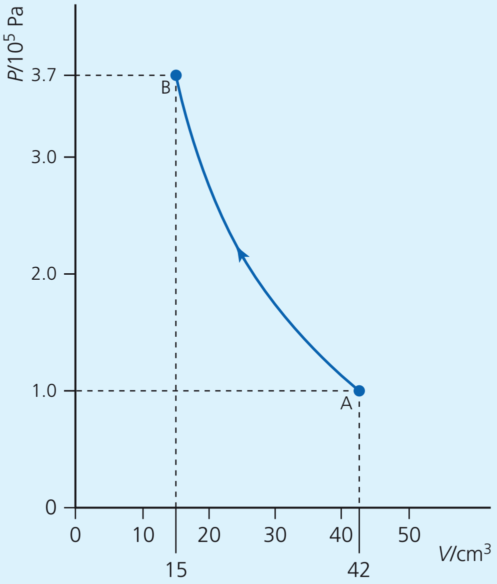


Figure B4.9 *PV* graph

- **a** Describe the process.
- **b** Did the process obey Boyle's law?
- **c** What has happened to the temperature of the gas? Explain your answer.
- **d** Estimate the work done during the process.
- **e** During the process, 1.9 J of thermal energy flowed out of the gas. Determine the change in its internal energy.

12 1.74 mol of an ideal monatomic gas expanded and during the process its temperature changed. The container was well insulated and no thermal energy was able to flow into, or out of, the gas.
- **a** If the internal energy of the gas decreased by 29.7 J, calculate the temperature change of the gas.
- **b** Determine how much work was done by the gas during this expansion.

13 Explain why a gas gets hotter when it is compressed rapidly.

Four thermodynamic processes

SYLLABUS CONTENT

- Isovolumetric, isobaric, isothermal and adiabatic processes are obtained by keeping one variable fixed.
- Adiabatic processes in monatomic ideal gases can be modelled by the equation as given by: $PV^{\frac{5}{3}} = \text{constant}$

Among all the various possible changes of state that a gas might experience, it is convenient to consider the first law of thermodynamics under four extremes:

- $\Delta U = 0$
- $W = 0$
- $Q = 0$
- $\Delta P = 0$

These are represented in Figure B4.10, in which the dotted lines are all isothermals.

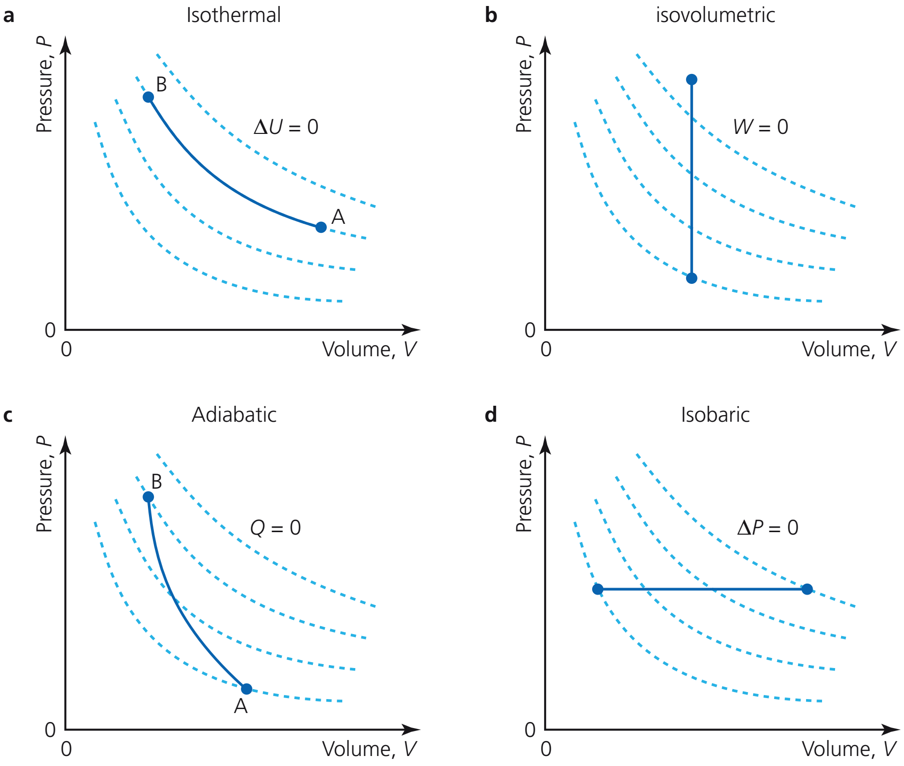


Figure B4.10 Four thermodynamic processes

$\Delta U = 0$ (isothermal process)

(The prefix iso- means equal.)

There is no change in the internal energy of the gas because its temperature is constant. Therefore,

in an isothermal change:

$$Q = 0 + W \quad \text{or } Q = W$$

In an isothermal expansion (B → A) all the work done by the gas on the surroundings is supplied by thermal energy transferred into the gas. In an isothermal compression (A → B), the work done on the gas is all transferred away from the gas as thermal energy. For a process to approximate to the ideal of being isothermal, the change must be as slow as possible. Isothermal changes obey Boyle's law (as described in Topic B.3): $PV = \text{constant}$.

$W = 0$ (isovolumetric process)

There is no work done by or on the gas because there is no change in volume. Therefore,

◆ **Isovolumetric** Occurring at constant volume.

in an **isovolumetric change**:

$$Q = \Delta U + 0 \quad \text{or } Q = \Delta U$$

In this straightforward process, if thermal energy is transferred into a gas, it simply gains internal energy and its temperature rises. If thermal energy is transferred away from a gas, its internal energy and temperature decrease.

$Q = 0$ (adiabatic process)

No thermal energy is transferred between the gas and its surroundings. Therefore

◆ **Adiabatic** Occurring without thermal energy being transferred into or out of a thermodynamic closed system.

in an **adiabatic change**:

$$0 = \Delta U + W$$

$\Delta U = -W$ for a compression and $-\Delta U = W$ for an expansion.

In an adiabatic expansion (B → A) all the work done by the gas is transferred from the internal energy within the gas, ΔU is negative and the temperature decreases. In an adiabatic compression (A → B) all the work done on the gas ($-\Delta W$) is transferred to the internal energy of the gas, which gets hotter.

When gas molecules hit the inwardly moving piston during a compression, they gain kinetic energy and the temperature rises. When gas molecules hit the outwardly moving piston during an expansion, they lose kinetic energy and the temperature falls.

For a process to approximate to the ideal of being adiabatic, the change must be as rapid as possible in a well-insulated container.

Note that adiabatic lines on PV diagrams must be steeper than isothermal lines, because in equal expansions, the temperature decreases during an adiabatic change, but is constant (by definition) during an isothermal change. This difference can be quantified by considering PV relationships:

We know that PV = constant during an isothermal change (that is, PV^1 = constant) but,

in an adiabatic change of an ideal monatomic gas:

$$\boldsymbol{PV^{\frac{5}{3}} = \text{constant}}$$

For gases other than monatomic ideal gases, V is raised to a different power, but this is not included in this course.

Comparing this equation to that for an isothermal change, we see that similar changes in volume are associated with greater changes in pressure during adiabatic changes. This is because there are also accompanying temperature changes.

$\Delta P = 0$ (isobaric process)

Any expansion or compression that occurs at constant pressure. Therefore,

in an isobaric change:

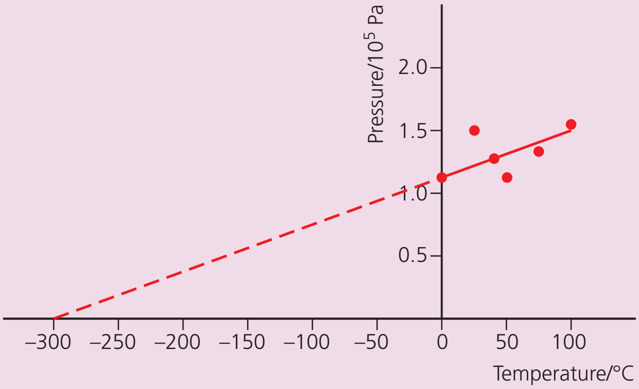
$$Q = \Delta U + W \text{ (and } W = P\Delta V\text{)}$$

Most isobaric changes occur when gases are allowed to expand or contract freely when their temperature changes, keeping their pressure the same as the surrounding pressure.

Inquiry 2: Collecting and processing data

Interpreting results

When you write an investigation report it will probably contain charts, diagrams or graphs, such as seen in Figure B4.11: a student used an isovolumetric change of a gas to estimate a value for absolute zero, using the apparatus shown in Figure B3.14, Topic B.3.



■ **Figure B4.11** Results of an isovolumetric experiment.

Graphical representations summarize the data collected and enable the reader to gain a quick impression of the results of the investigation. Your investigation report should summarize what you have learned from any such representations.

This may involve any, or all, of:

- considering the quantity and spread of measurements taken
- discussing the quality and reliability of graphs
- identifying any relationship that can be identified between the two variables
- describing and explaining any pattern or trend shown by a graph (if no precise relationship is apparent)
- quoting values for intercepts and explaining their significance
- calculating gradients and explaining their significance
- calculating areas under graphs and explaining their significance
- identifying and explaining any maxima or minima (and other turning points).

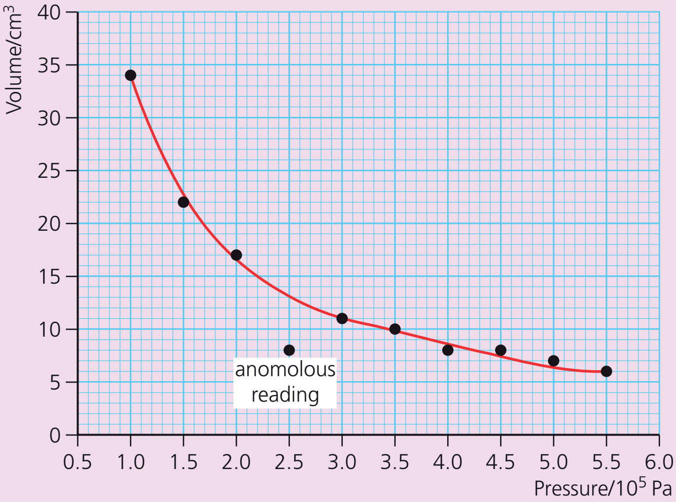
Anomalous readings (outliers) should not usually be rejected without being checked. The most common reason for an outlier is a simple error in measurement, or recording, which will be evident if it is repeated.

It is unusual that an anomalous reading is confirmed as being correct. But, if it is, you should take further measurements at slightly different values to try to determine the extent of the anomaly.

Ideally, the pattern of the data should be processed quickly at the time of the experiment, often by drawing a rough graph. This prevents the situation whereby you only notice an anomalous reading when processing the data later, when checking the measurement again is not possible. See Figure B4.12 which shows the results of a Boyle's law experiment (Topic B.3).

If not checked, you should include any outliers in your report (and note them as outliers). You need to use your judgement as to whether the outlier should affect your conclusions.

Analyse the graphs shown in the last two figures and discuss the quality of the experimentation that they represent.



■ **Figure B4.12** An anomalous reading (uncertainty bars not shown)

WORKED EXAMPLE B4.5

A gas of volume 0.080 m^3 and pressure 1.4×10^5 Pa expands to a volume of 0.11 m^3 at constant pressure when 7.4×10^3 J of thermal energy are supplied.

a Name the thermodynamic process.

b Calculate the work done by the gas.

c Determine the change in the internal energy of the gas.

Answer

a An isobaric process.

b $W = P\Delta V = (1.4 \times 10^5) \times (0.11 - 0.080) = +4.2 \times 10^3$ J

c $Q = \Delta U + W$

$(+7.4 \times 10^3) = \Delta U + (+4.2 \times 10^3)$

$\Delta U = +3.2 \times 10^3$ J (The internal energy of the gas increases.)

WORKED EXAMPLE B4.6

The volume of an ideal monatomic gas is reduced in an adiabatic compression by a factor of 8.0. Determine the factor by which the pressure in the gas changes.

Answer

$$PV^{\frac{5}{3}} = \text{constant}$$

$$P_1V_1^{\frac{5}{3}} = P_2V_2^{\frac{5}{3}}$$

$$\frac{P_2}{P_1} = \left(\frac{V_1}{V_2}\right)^{\frac{5}{3}} = 8^{\frac{5}{3}}$$

$$\log\left(\frac{P_1}{P_2}\right) = \frac{5}{3} \times \log 8 = 1.505$$

$$\frac{P_1}{P_2} = 32$$

The pressure has increased by a factor of 32.

If this had been an isothermal change the pressure would have increased by the same factor (8.0) as the volume has decreased. In this example the pressure has increased by a bigger factor because the temperature increased in an adiabatic compression.

14 Figure B4.13 represents four successive changes to the state of a gas of constant mass.

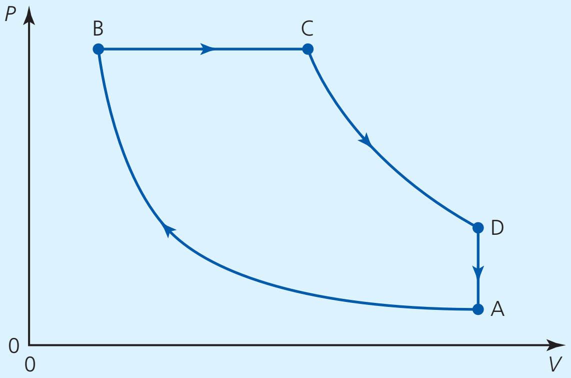


Figure B4.13 Four successive changes to the state of a gas

a Name the processes shown by:

i BC **ii** DA.

b If CD is an adiabatic process, compare the temperatures at points C and D.

c If AB is an isothermal process, compare the temperatures at points A and B.

15 Copy and complete Table B4.1 by putting 0, +, –, or ± in each box

Table B4.1

		Q	ΔU	W	ΔP	ΔV	ΔT
isothermal	expansion						
	compression						
adiabatic	expansion						
	compression						
isobaric	expansion						
	compression						
isovolumetric	pressure increase						
	pressure decrease						

16 An ideal monatomic gas expands adiabatically from a volume of 3.13×10^{-3} m^3 to 3.97×10^{-3} m^3.

a If the original pressure was 2.60×10^5 Pa, determine the final pressure.

b If the final temperature of the gas was 398 K, what was the starting temperature?

Thermodynamic cycles and *PV* diagrams

SYLLABUS CONTENT

- Cyclic gas processes are used to run heat engines.
- A heat engine can respond to different cycles and is characterized by its efficiency: $\eta = \frac{\text{useful work}}{\text{input energy}}$
- The Carnot cycle sets a limit for the efficiency of a heat engine at the temperatures of its heat reservoirs: $\eta_{\text{carnot}} = 1 - \frac{T_c}{T_h}$.

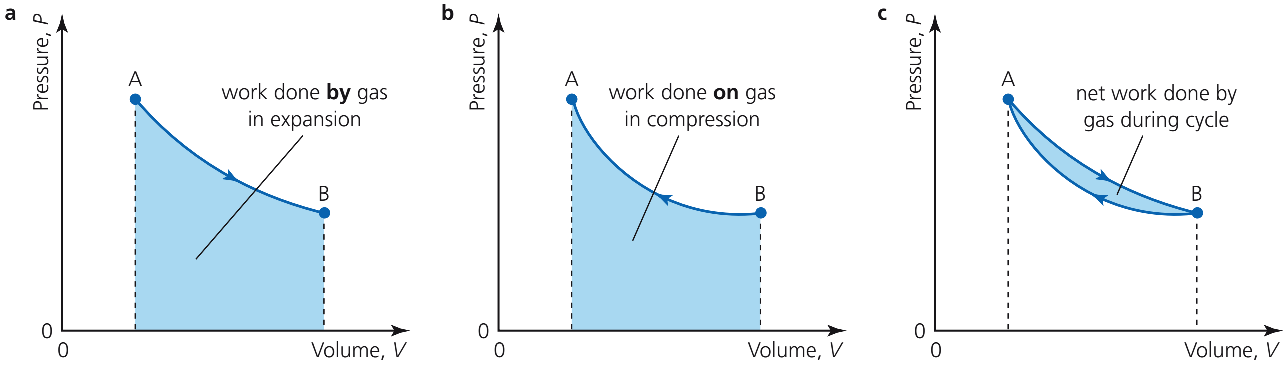
◆ **Working substance** The substance (usually a gas) used in thermodynamic processes to do useful work.

◆ **Cycle (thermodynamic)** A series of thermodynamic processes that return a system to its original state (for example, the Carnot cycle). Usually, the process repeats continuously.

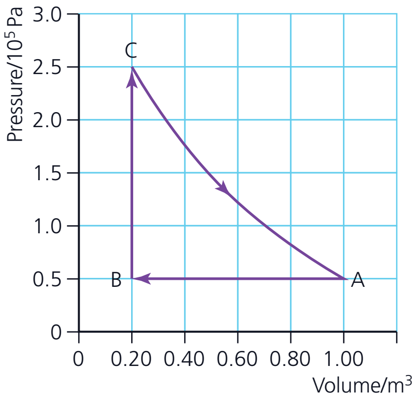
An expanding gas (sometimes called a **working substance**) can do useful work – for example, by making a piston move along a cylinder. However, a gas in a container cannot expand for ever, so any practical device transferring thermal energy to mechanical work must move in **thermodynamic cycles**, involving repeated expansions followed by compressions, followed by expansions and so on. In this section we will discuss some of the physics principles that are fundamental to cyclical processes. It is important to realize that we are not describing details of the actual mechanical processes in any particular type of engine.

> The essential process in a *heat engine* is the transfer of thermal energy to produce expansion and do useful mechanical work.

This is represented by path AB in Figure B4.14a. The shaded area under the graph represents the work done *by* the gas during the expansion.



■ **Figure B4.14** Work done in a thermodynamic cycle



■ **Figure B4.15** An idealized complete cycle

In a cyclical process, the gas has to be compressed and returned to its original state. Assume, for the sake of simplicity, that this is represented by the path shown in Figure B4.14b. The area under this graph represents the work done *on* the compressed gas. The difference in areas, shown in Figure B4.14c, is the net useful work done by the gas during one cycle. Of course, if we imagined the impossible situation that, when the gas was compressed, it returned along exactly the same path that it followed during expansion, there would be no net useful work done and no energy wasted.

Figure B4.15 shows a simplified example of an complete cycle. Of course, there are a large number of cycles that could be drawn on a *PV* diagram and, if they are to be considered as the basis of a useful engine, then it is important that they have high efficiencies.

Efficiency of heat engines

We have discussed *efficiency* in Topic A.3, and in the context of this topic it can be restated as:

$$\text{efficiency of a heat engine, } \boldsymbol{\eta} = \frac{\textbf{useful work}}{\textbf{input energy}}$$

Calculating the area under a *PV* diagram is a useful way of determining the efficiency of a thermodynamic cycle.

WORKED EXAMPLE B4.7

Determine the efficiency of the simple cycle shown in Figure B4.14 if the area shown in B4.14a was 130 J and the area in B4.14b was 89 J.

Answer

$$\eta = \frac{\text{useful work done}}{\text{energy input}} = \frac{(130 - 89)}{130} = 0.32\ (32\%)$$

WORKED EXAMPLE B4.8

Figure B4.15 shows one simplified cycle, ABCA, of a gas in a particular heat engine, during which time 1.3×10^5 J of thermal energy flowed into the gas.

a Calculate the work done during the process AB.

b Name the processes AB and BC.

c Estimate the net useful work done by the gas during the cycle.

d What is the approximate efficiency of the engine?

Answer

a $W = P\Delta V$ = area under AB = $(0.50 \times 10^5) \times (1.0 - 0.20) = 4.0 \times 10^4$ J (done on the gas)

b AB occurs at constant pressure: isobaric compression. BC occurs at constant volume: isovolumetric temperature increase.

c Net work done by gas = area enclosed in cycle = $(1.0 - 0.20) \times (1.3 - 0.5) \times 10^5$ (estimated from a rectangle having about the same area, as judged by eye)

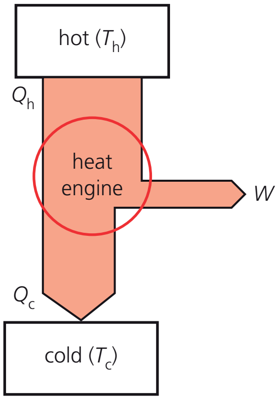
$W = 6.4 \times 10^4$ J

d $$\text{efficiency} = \frac{\text{useful work output}}{\text{total energy input}} = \frac{6.4 \times 10^4}{1.3 \times 10^5} \approx 0.50\ (50\%)$$

The purpose of a heat engine is to do useful work: to transfer thermal energy into mechanical energy (motion). We know that the opposite process, of converting mechanical energy into thermal energy, occurs around us all the time. For example, rubbing our hands together. Such processes are often 100% ‘efficient’, and the inevitable dissipation of useful energy in this way is very familiar in the study of dynamics (Topic A.3).

The laws of thermodynamics show us that, in a cyclical process, *it is not possible to convert thermal energy into work with 100% efficiency*. As will see, in practice, 50% is a good output! This is known as the Kelvin form of the second law of thermodynamics (see later).

Heat engines need a flow of thermal energy, which we know always spontaneously flows from a hotter region to a colder region. Some of that energy, but never all of it, is transferred to do useful work. Figure B4.16 represents this in schematic form.



■ **Figure B4.16** Energy flow in a heat engine

◆ **Carnot cycle** The most efficient thermodynamic cycle. An isothermal expansion followed by an adiabatic expansion; the gas then returns to its original state by isothermal and adiabatic compressions.

A temperature difference, $(T_h - T_c)$, is needed between hot and cold reservoirs, so that there is a resulting flow of thermal energy, which operates the engine. Thermal energy Q_h flows out of the hot reservoir and Q_c flows into the cold reservoir. The difference in thermal energy is transferred to doing useful mechanical work, W.

efficiency of a heat engine:

$$\eta = \frac{W}{Q_h}$$

or:

$$\eta = \frac{(Q_h - Q_c)}{Q_h} = 1 - \frac{Q_c}{Q_h}$$

Carnot cycle

The thermodynamic cycle that produces the maximum theoretical efficiency is called the Carnot cycle.

The **Carnot cycle** is an idealized and reversible four-stage process, as shown in Figure B4.17: an isothermal expansion (AB) is followed by an adiabatic expansion (BC); the gas then returns to its original state by isothermal (CD) and adiabatic compressions (DA). Thermal energy is transferred during the two isothermal stages. By definition, thermal energy is not transferred during the adiabatic changes.

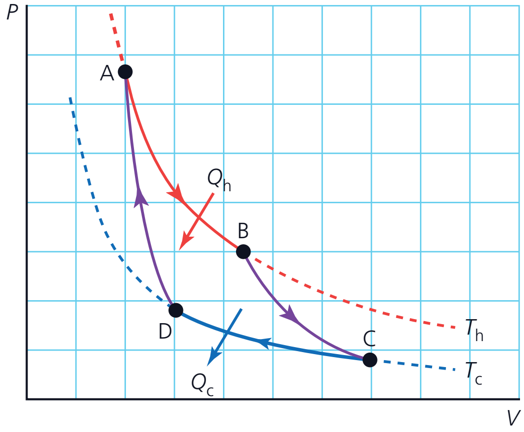
efficiency of a heat engine using the Carnot cycle:

$$\boldsymbol{\eta_{\text{Carnot}} = 1 - \frac{T_c}{T_h}}$$

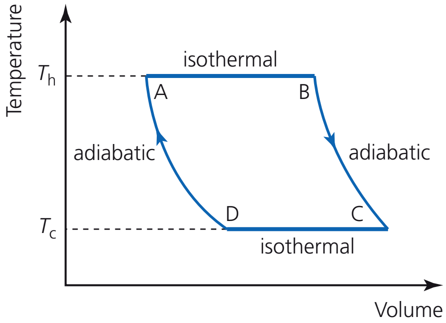
An explanation of the origin of this equation is provided in the *entropy* sub-section.

It is important to realize that this equation represents the most efficient cycle allowed by the laws of physics. This is very different from, for example, processes in which efficiency is limited by energy dissipation due to friction (which also occurs in heat engines).

The Carnot cycle may also be represented on temperature–pressure and temperature–volume graphs (such as shown in Figure B4.18).



■ **Figure B4.17** Carnot cycle



■ **Figure B4.18** Carnot cycle on a volume–temperature graph

WORKED EXAMPLE B4.9

Determine the theoretical thermodynamic efficiencies of Carnot cycles operating between temperatures of:

a 100 °C and 20 °C

b 500 °C and 100 °C

c 150 °C and −150 °C.

Answer

a $\eta_{\text{carnot}} = 1 - \frac{T_c}{T_h} = 1 - \left(\frac{293}{373}\right) = 0.21\ (21\%)$

b $\eta_{\text{carnot}} = 1 - \frac{T_c}{T_h} = 1 - \left(\frac{373}{773}\right) = 0.52\ (52\%)$

c $\eta_{\text{carnot}} = 1 - \frac{T_c}{T_h} = 1 - \left(\frac{123}{423}\right) = 0.71\ (71\%)$

These calculations show us that the efficiency of a Carnot cycle (and others) is limited by the maximum and minimum temperatures that are possible. Extremely high temperatures are technologically difficult to sustain. More importantly, we live in a world which has an average temperature of 288 K.

A waterfall analogy may be helpful. See Figure B4.19. A waterwheel placed in falling water can convert the lost gravitational potential energy into useful kinetic energy of the wheel, but, if for some reason the wheel is placed at a height of *h*/2, a maximum of only 50% of the available energy will be transferred.

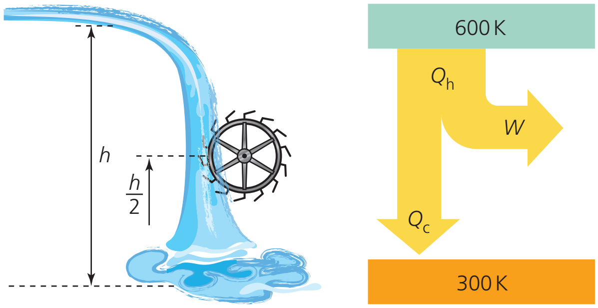


Figure B4.19 Waterfall analogy. Both processes have a maximum efficiency of 50%

If in a Carnot cycle the thermal energy could be transferred between 600 K and 0 K, then the theoretical efficiency would be 100%. But this is not possible, and if we are constrained to an outlet temperature of about 300K, then the maximum possible efficiency is only 50%.

17 a With an outlet temperature of 350 K, calculate the inlet temperature needed to achieve a maximum theoretical efficiency of 45% with a Carnot cycle.

b In practice, why will a higher temperature be needed?

18 Suggest reasons why electrical power stations need cooling towers, such as those seen in Figure B4.20.

Figure B4.20 Cooling towers in Poland

19 Sketch a temperature–pressure diagram for the Carnot cycle.

20 Determine the thermodynamic efficiency of a Carnot cycle in which thermal energy is flowing out of the hot reservoir at a rate of 1.26×10^5 W and into the cold reservoir at a rate of 0.79×10^5 W.

21 Estimate the efficiency of the process shown in Figure B4.16.

22 Figure B4.21 shows the four-stage cycle of a heat engine.

a Which stage is the compression of the gas?

b The temperature at A is 320 K. Calculate the amount of gas in moles.

c Calculate the temperature at point B.

d Estimate the area ABCD. What does it represent?

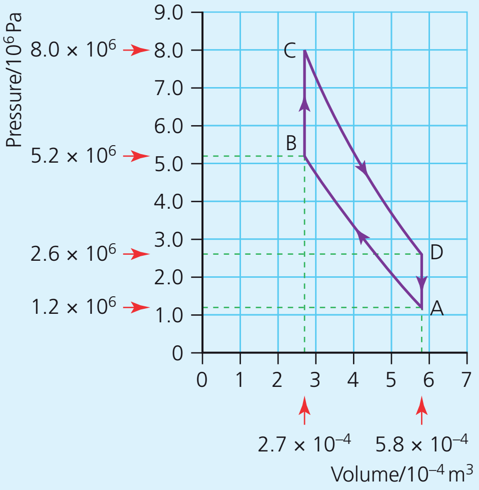


Figure B4.21 The four-stage cycle of a heat engine

23 Using graph paper, make a sketch of the following four consecutive processes in a heat engine. Start your graph at a volume of 20 cm^3 and a pressure of 6.0×10^6 Pa.

a an isobaric expansion increasing the volume by a factor of five

b an adiabatic expansion doubling the volume to a pressure of 1.5×10^6 Pa

c an isovolumetric reduction in pressure to 0.5×10^6 Pa

d an adiabatic return to its original state.

e Mark on your graph where work is done on the gas.

f Estimate the net work done by the gas during the cycle.

Heat pumps

◆ **Heat pump** Device which transfers thermal energy from a colder place by doing work.

A **heat pump** works like a heat engine in reverse, using a work input to enable the transfer of thermal energy from colder to hotter.

Refrigerators and air-conditioners are heat pumps (see Figure B4.22).

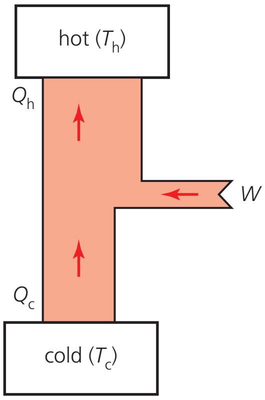


Figure B4.22 Energy transfers in a heat pump

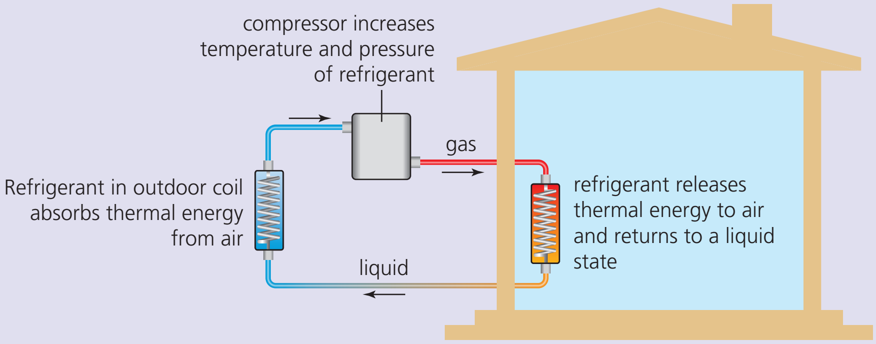
ATL B4A: Research skills, thinking skills

Applying key ideas and facts in new contexts

Using heat pumps for heating homes

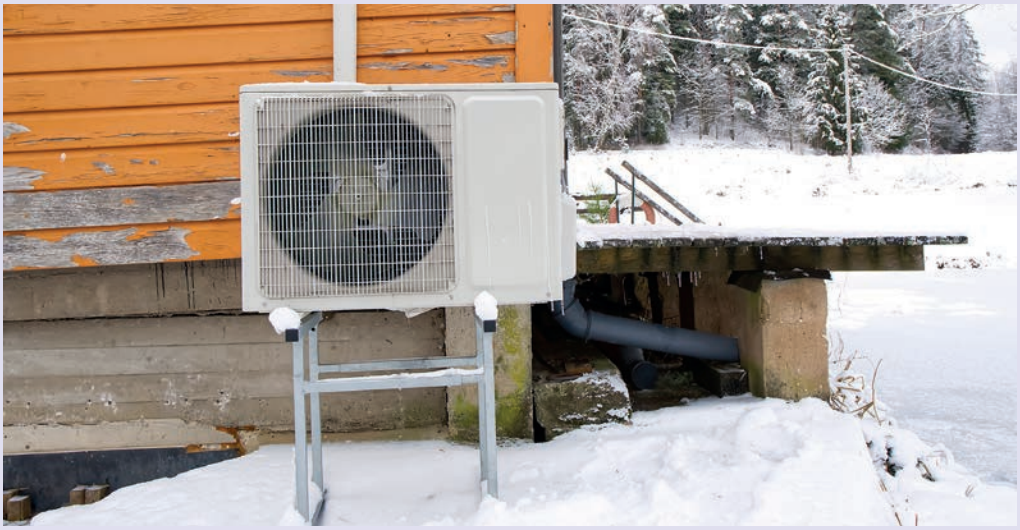
The purpose of refrigerators is to transfer thermal energy from a colder place, inside a refrigerator, to a warmer place (the kitchen). Heat pumps using the same principle can also be used to transfer thermal energy from a colder exterior environment (air or ground) and use it to help to keep the interior of a house warm in winter, although their efficiency may be disappointing in very cold weather, when additional alternative heating may be needed.

Heat pumps use the fact that an evaporating liquid (the 'refrigerant') removes thermal energy (latent heat) from itself and then its surroundings. The thermal energy is released when the gas later is compressed and condenses back to its liquid state. Most heat pumps can be 'reversed' to use as air-conditioners in hot weather. Electrical energy is needed to operate the compressor, as shown in Figure B4.23.



■ **Figure B4.23** Extracting thermal energy from a colder environment

We have seen that the internal energy, *U*, in any substance is proportional to its temperature, (K). It may be surprising to realize that air (for example) at 0 °C has 92% of the internal energy that the same air has at 25 °C.



■ **Figure B4.24** Heat pump outside a home

Why might people choose to install a heat pump to heat their homes, instead of a more conventional fossil-fuel or electric heating system? What could their impact on climate change be?

Using your own research, determine some of the advantages and disadvantages of heat pumps.

Reversible and irreversible changes

SYLLABUS CONTENT

- Processes in real isolated systems are almost always irreversible and consequently the entropy of a real isolated system always increases.

◆ **Irreversible process** A process which cannot be reversed, and in which entropy (see below) always increases. All real macroscopic processes are irreversible.

◆ **Reversible process** A process that can be reversed so that the system and all of its surroundings return to their original states and there is no change in entropy. An impossibility in the macroscopic world.

The required flow of some thermal energy into a cold reservoir in a heat engine is an example of an **irreversible process**, which means that the system *and its surroundings* cannot be returned exactly to their original states.

If the original state of a system *and its surroundings* can be restored exactly, the process is described as **reversible**.

In practice, *all* macroscopic processes can be considered to be irreversible.

Consider watching a swinging pendulum. At first it may seem that the motions keep reversing perfectly. However, if we keep watching, we will notice that the amplitudes decrease, because energy is dissipated out of the system into the surroundings. A video of a pendulum swinging shown in reverse could never be mistaken for the normal behaviour of a pendulum. In fact, almost all events shown in reverse will be obviously just that. Figure B4.25 shows another simple example.

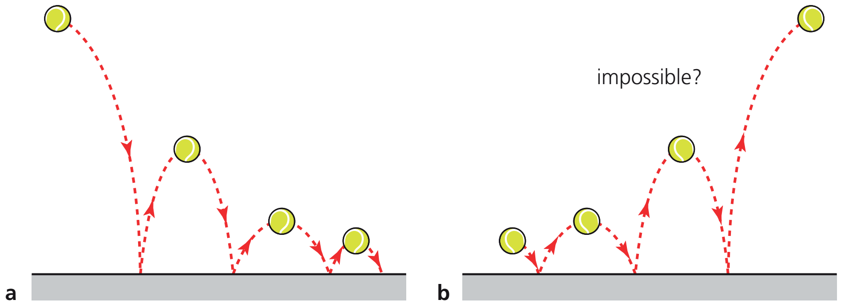


Figure B4.25 Bouncing ball

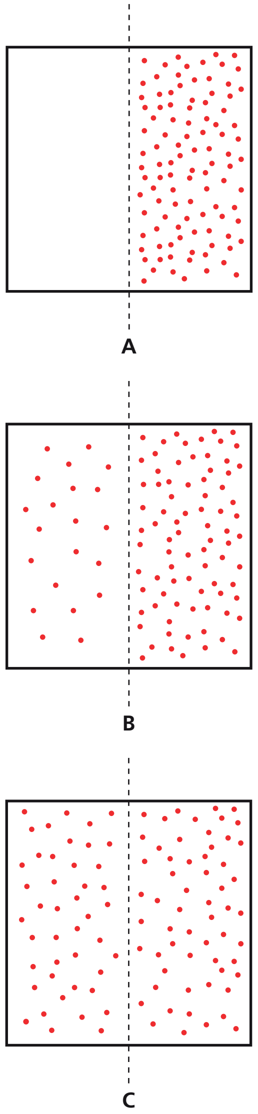
The passage of time seems to be linked to the irreversibility of processes.

So, then we must ask: what is the scientific principle that makes processes irreversible? Why can there be a net energy flow out of a bouncing ball into the ground, but not out of the ground into the ball? The first law of thermodynamics (conservation of energy) does not help us to answer this question.

◆ **Order and disorder (particle)** The way in which particles are arranged, or energy is distributed, can be described in terms of the extent of patterns and similarities.

A bouncing ball (Figure B4.25) has internal energy: its particles have individual *random* potential and kinetic energies. But, in addition, all the particles each have the *same* velocity as the ball as a whole. We need to distinguish between the **ordered** energy of the particles moving in the ball as a whole, and the random **disordered** energies of the same particles. As the ball bounces, more and more ordered energy is transferred to disordered energy. Finally, all the kinetic energies are disordered. The process is irreversible.

Without outside interference, in any process, ordered energy of particles will be *irreversibly* transferred to disordered energy.



■ **Figure B4.26** Gas molecules spreading out in a container

To explain this, we need to consider statistics. Consider an everyday, non-physics example: 20 candy bars are to be distributed among 10 children. The fairest and most ordered way of doing this is to give two bars to each child. However, if the distribution is not controlled, but is entirely random, it is *extremely* unlikely that an even distribution will occur. It is much more likely that, for example, some children will get four and some children will get none. A full statistical analysis can make a reasonably accurate prediction of the overall distribution but cannot predict the number of candy bars given to any particular child.

Continuing the analogy, consider what would happen if we started with an ordered situation in which 10 children each had two candy bars, and then arranged a random re-distribution. The result will be a change to the same overall distribution as described in the previous paragraph.

However, if the distribution was already disordered, further random changes would not affect the overall distribution.

Returning to particle motions, the molecules of a gas move in completely random and uncontrollable ways. What happens to them is simply the most likely outcome. It is theoretically possible for all the randomly moving molecules in a room to go out of an open window at the same time. The only reason that this does not happen is simply that it is statistically *extremely* unlikely.

Consider Figure B4.26, which shows three distributions of the same number of gas molecules in a container. The dotted line represents an imaginary line dividing the container into two equal halves. We can be (almost) sure that A occurred before B, and that B occurred before C. (Probably the gas was released at first into the right-hand side of the container.)

Because it is so unlikely for molecules moving randomly, we simply cannot believe that C occurred before B and A. (In a similar way, we would not believe that if 100 coins were tossed, they could all land ‘heads’ up.) Figure B4.26 only shows about 100 molecules drawn to represent a gas. In even a very small sample of a real gas there will be as many as 10^{19} molecules, turning a highly probable behaviour into a certainty. The simplest way we have of explaining this is that, in the process of going from A to B to C (moving forward in time), the system becomes more disordered.

Similarly, the fact that energy is exchanged randomly between molecules leads to the conclusion that molecular energies will become more and more disordered and spread out as time goes on. Thermal energy will inevitably spread from places where molecules have higher average kinetic energy (hotter) to places with lower average molecular kinetic energies (colder). This is simply random molecular behaviour producing more disorder.

We can be certain that every isolated system of particles cannot spontaneously become more ordered as time progresses. Put simply, this is because everything is made up of particles, and individual atoms and molecules are uncontrollable. Everything that happens occurs because of the random behaviour of individual particles. Of course, we may wish to control and order molecules, for example by turning water into ice, but this would not be an *isolated* system – to impose more order on the water molecules we must remove thermal energy and this will result in even higher molecular disorder in the surroundings.

Two everyday examples may help our understanding: why is it much more likely that a pack of playing cards will be disordered rather than in any particular arrangement? Why is a desk, or a room, much more likely to be untidy rather than tidy? Because, left to the normal course of events, things get disorganized. To produce order from disorder requires intervention and may be difficult, or even impossible. There are a countless number of ways to disorganize a system, but only a relatively few ways to organize it.

Entropy

SYLLABUS CONTENT

- Entropy, S, is a thermodynamic quantity that relates to the degree of disorder of the particles in a system.

◆ **Entropy, S** A measure of the disorder of a thermodynamic system of particles.

◆ **Second law of thermodynamics** The overall entropy of the universe is always increasing. This implies that energy cannot spontaneously transfer from a place at low temperature to a place at high temperature. Or, in the Kelvin version: when extracting energy from a heat reservoir, it is impossible to convert it all into work.

The disorder of a system of particles can be calculated. It is known as the **entropy** of the system.

The concept of entropy, S, numerically expresses the degree of disorder in a system.

Molecular disorder and the concept of entropy are profound and very important ideas. They are relevant everywhere – to every process in every system, to everything that happens anywhere and at any time in the Universe. The principle that molecular disorder is always increasing is neatly summarized by the second law of thermodynamics.

Second law of thermodynamics

SYLLABUS CONTENT

- The second law of thermodynamics refers to the change in entropy of an isolated system and sets constraints on possible physical processes and on the overall evolution of the system.
- The entropy of a non-isolated system can decrease locally, but this is compensated by an equal or greater increase of the entropy of the surroundings.

The **second law of thermodynamics** states that in every process, the total entropy of any isolated system, or the Universe as a whole, always increases.

Figure B4.27 A refrigerator transfers thermal energy from the food and reduces entropy, but where does the energy go?

This is sometimes expressed by the statement ‘entropy can never decrease’. But it should be stressed that it is certainly possible to reduce the ‘local’ entropy of part of a system, but in the process another part of the system will gain *even more* entropy. For example, the growth of a plant, animal or human being reduces the entropy of the molecules that come to be inside the growing body, but there will be an even greater increase in the entropy of all the other molecules in the surroundings that were involved in the chemical and biological processes. Water freezing or the action of a refrigerator, such as seen in Figure B4.27, provide other examples. The internal energy of the contents is reduced as thermal energy is transferred away at the back of the refrigerator. The entropy of the contents is reduced because they are colder, but the entropy of the kitchen is increased by a greater amount because it is hotter.

The statistical analysis of the behaviour of enormous numbers of uncontrollable particles leads to the inescapable conclusion that differences in the macroscopic properties of any system, such as energy, temperature and pressure, must even out over time. This is represented quantitatively by a continuously increasing entropy. This suggests that, eventually, all energy will be spread out, all differences in temperature will be eliminated and entropy will reach a final steady, maximum value. This is often described as the *‘heat death’ of the Universe*.

LINKING QUESTION

- Why is there an upper limit on the efficiency of any energy source or engine?

This question links to understandings in Topic A.3.

Disorder and entropy in macroscopic systems

A few examples:

- Gas in a large volume is more disordered / has greater entropy than the same gas at the same temperature in a smaller volume.
- Gas at a higher temperature is more disordered / has greater entropy than the same gas in the same volume at a lower temperature.
- A liquid is more disordered / has greater entropy than a solid of the same material at the same temperature.

Alternative ways of expressing the second law

Consider any two objects at different temperatures placed in thermal contact in an isolated system with no external influences, as shown in Figure B4.28.

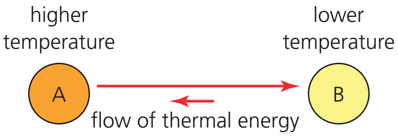


Figure B4.28 Exchanges of energy

Thermal energy can flow from A to B and from B to A, but the net flow of energy is from A to B because the increase in entropy of B will be greater than the decrease in entropy of A (for the same energy transfer). Therefore, the net flow of thermal energy will always be from hotter to colder. This is as we have explained previously in Topic B.1, when discussing particle collisions.. It is an alternative version of the second law of thermodynamics, first expressed by the German physicist *Rudolf Clausius*:

> Thermal energy cannot *spontaneously* transfer from a region of lower temperature to a region of higher temperature.

But we *can* use heat engines to transfer thermal energy from colder to hotter by doing external work (heat pumps – see earlier). Thermal energy always flows *spontaneously* from hotter to colder. Insulation can be used to reduce the rate of energy transfer but can never stop it completely. A third version (the *Kelvin* form) of the second law of thermodynamics is expressed in terms of a thermodynamic cycle as:

> When extracting energy from a heat reservoir, it is impossible to convert it all into work.

Representing entropy mathematically

SYLLABUS CONTENT

- ▶ Entropy can be determined in terms of macroscopic quantities such as thermal energy and temperature, as given by: $\Delta S = \frac{\Delta Q}{\Delta T}$; and also in terms of the properties of individual particles of the system as given by: $S = k_B \ln \Omega$, where k_B is the Boltzmann constant and Ω is the number of possible microstates of the system.

To express the entropy of a system of particles numerically, we need to count the number of ways that the system can be arranged. This is sometimes described as its *multiplicity*. The ‘state’ of the system can be defined by any property or properties which allows it to be distinguished from other states, for example particle positions or distribution of energies.

First, we will consider an everyday, non-physics example: throwing two six-sided dice and adding the numbers shown to obtain a total. Figure B4.29 shows all the possible combinations. How do we explain that a total of seven is the most likely?

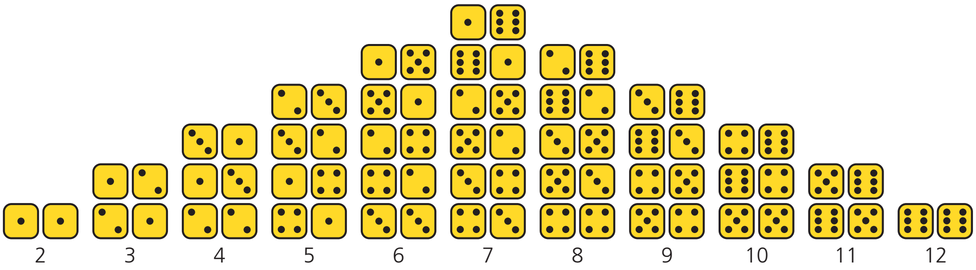


Figure B4.29 Combinations of two dice

There are 36 possible combinations, with totals ranging from 2 to 12. A total of 2 can only be obtained in one way: 1 + 1 = 2. Similarly, a total of 12 can only be obtained in one way: 6 + 6 = 12. However, a total of 7 can be obtained in six ways: 1 + 6, 2 + 5, 3 + 4, 4 + 3, 5 + 2 and 6 + 1.

Put simply: a total of 7 is most likely because it can be produced in the greatest numbers of ways.

◆ **Microstates** The numerous possible combinations of microscopic properties of a thermodynamic system.

Now consider a very simple system of three freely moving gas particles (A, B and C). We will describe the state of the system in terms of the locations of the particles: to the left, or to the right, of an imaginary line dividing their container in half. We can identify 8 (2^3) possible arrangements, as shown in Figure B4.30. These arrangements may be called **microstates**.

All these eight microstates are equally likely if the particles are moving randomly. However, either of the two 'ordered' microstates, shown in the top line, is less likely than any one of the other six 'disordered' microstates. A system which is ordered will inevitably become disordered.

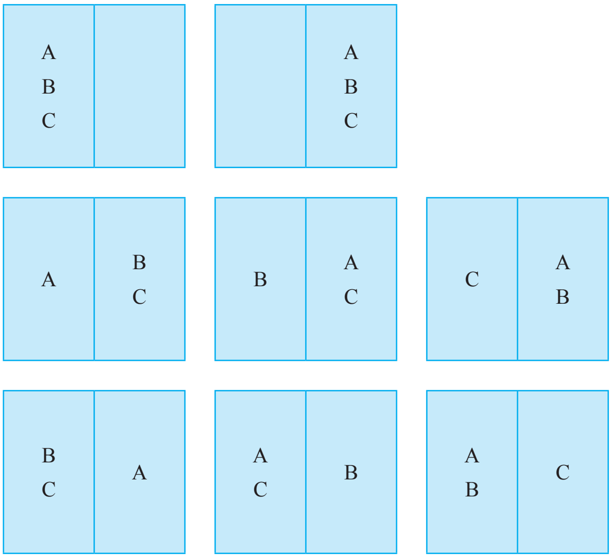


Figure B4.30 Distributions of three particles

In order to establish the principle, this example has used a small number of particles (3) for simplicity. If larger numbers are used, it becomes clear that returning to an original, ordered arrangement is effectively impossible. Extending the example seen in Figure B4.30: if there were 10 particles, the number of possible disordered microsites would be greater than 1000 (2^{10}). And remember that even small gaseous systems will contain 10^{19} or more particles. (So, the probability of all of the particles being in the left-hand side of a small container is one in two to the power of 10 to the power of 19.).

The greater the number of possible microstates of a system, the greater its disorder and the greater its entropy.

The symbol Ω is used to represent the number of possible microstates of a system (its multiplicity). Clearly, Ω will be a large number. $\ln \Omega$ is more manageable.

Entropy of a system of microscopic particles:

$$S = k_B \ln \Omega \quad \text{(SI Unit: J K}^{-1}\text{)}$$

Nature of science: Science as a shared endeavour

Expressing laws as formulas

The considerable importance of the second law of thermodynamics is undoubted and a broad understanding of the associated concept of entropy is becoming more widespread among the general public. However, entropy is a difficult concept to understand well and, similar to many scientific principles, its true meaning requires the precision of mathematics. Entropy can be determined from the equation $S = k_B \ln \Omega$, but there is no easy application of this equation to everyday life.

The equation was devised by the Austrian physicist Ludwig Boltzmann, who considered the equation to be so important that it was famously carved on his memorial in Vienna. See Figure B4.31. (W was used instead of Ω.)

■ **Figure B4.31** Ludwig Boltzmann

WORKED EXAMPLE B4.10

What is the entropy of a system which has 1×10^{22} microstates?

Answer

$S = k_B \ln (1 \times 10^{22}) = (1.38 \times 10^{-23}) \times 50.7 = 7.0 \times 10^{-22}\,\mathrm{J\,K^{-1}}$

Worked example B4.10 is shown simply to illustrate the principle. In practice, such calculations are often unrealistic. Fortunately, *changes* of entropy, ΔS, in macroscopic situations are much easier to calculate. If thermal energy ΔQ is supplied to a system at a *constant temperature* of T:

change of entropy (using macroscopic quantities): $\boldsymbol{\Delta S = \dfrac{\Delta Q}{T}}$ Unit: $\mathrm{J\,K^{-1}}$

◆ **Entropy change** When an amount of thermal energy, ΔQ, is added to, or removed from, a system at temperature T, the change in entropy, ΔS, can be calculated from the equation $\Delta S = \dfrac{\Delta Q}{T}$.

WORKED EXAMPLE B4.11

Calculate a value for the **entropy change** when 5000 J of thermal energy flows out of a hot cup of coffee at 70 °C into the surrounding room at 25 °C. Assume that the temperatures of the coffee and the room are unchanged. (In practice the coffee will cool by about 5 °C.)

Answer

The entropy of the coffee has decreased by:

$$\Delta S = \frac{\Delta Q}{T} = \frac{-5000}{(273 + 70)} = 14.6\,\mathrm{J\,K^{-1}}$$

The entropy of the room has increased by:

$$\Delta S = \frac{\Delta Q}{T} = \frac{+5000}{(273 + 25)} = 16.8\,\mathrm{J\,K^{-1}}$$

Overall change of entropy in the coffee / room system = 16.8 −14.6 = +2.2 $\mathrm{J\,K^{-1}}$

The fact that the entropy has increased is related to the fact that the thermal energy flowed from a higher temperature to a lower temperature. If the coffee and the room were at the same temperature, there would be no flow of thermal energy and no change of entropy. If we imagined the impossible situation in which thermal energy could flow spontaneously out of the cooler room into the hotter coffee, then entropy of the system would decrease – which never happens.

Entropy in the Carnot cycle

The theoretical Carnot cycle is a reversible process, so that there is no overall change of entropy at the end of each cycle. Consider Figure B4.32 and compare it to Figure B4.17.

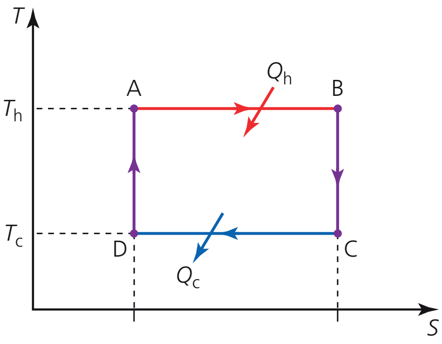
During the isothermal expansion, AB, thermal energy is supplied, the temperature remains constant but the entropy rises as the volume increases:

$$\Delta S = \frac{\Delta Q_h}{T_h}$$

During the adiabatic expansion, BC, the pressure and temperature decrease, but the entropy remains the same.

During the isothermal compression, CD, thermal energy is removed, the temperature remains constant but the entropy falls as the volume decreases:

$$\Delta S = \frac{\Delta Q_c}{T_c}\left(= \frac{\Delta Q_h}{T_h}\right)$$



■ **Figure B4.32** Temperature–entropy diagram for the Carnot cycle

During the adiabatic compression, DA, the pressure and temperature increase, but the entropy remains the same.

Since:

$$\frac{\Delta Q_c}{T_c} = \frac{\Delta Q_h}{T_h}$$

we can now explain the origin of the equation for the efficiency of a Carnot cycle:

$$\eta = 1 - \frac{\Delta Q_c}{Q_h}$$

But:

$$\frac{Q_c}{Q_h} = \frac{T_c}{T_h}$$

So that:

$$\eta_{\text{carnot}} = 1 - \frac{T_C}{T_h}$$

(As highlighted previously.)

ATL B4B : Thinking skills

Applying key ideas and facts in new contexts

The true nature of time has always preoccupied scientists. We have seen that the entropy of any system increases with time, and increasing entropy is sometimes described as representing the ‘arrow of time’. Is it possible that time is only that: an indication of increasing entropy? Does the second law of thermodynamics imply that time travel is just for science fiction stories, and can never be possible in reality?

LINKING QUESTION

- What are the consequences of the second law of thermodynamics to the Universe as a whole?

24 By discussing what happens to the molecules of the gas, explain the entropy change when a balloon bursts.

25 Coffee, sugar and milk are put in hot water to make a drink. Why it is difficult to reverse the process?

26 Imagine a large container of water, separated into two halves by a removable barrier. Half of the water is at 90 °C and the other half is at 20 °C.

- **a** Explain why, in theory, some of the energy in the water is available to do useful work.
- **b** If the barrier is removed and water from the two halves mixes, what will be the final temperature? Assume that no thermal energy is transferred to the surroundings.
- **c** Why can the particles in the system be described as more disordered after the mixing?
- **d** What has happened to:
  - **i** the total energy of the system
  - **ii** the total entropy of the system?
- **e** Explain why it is now impossible for the system to do any useful work.

27 Consider Figure B4.30.

- **a** How many microstates were there if there were four particles instead of three?
- **b** If the particles were moving randomly, what was the probability that all four were in the right-hand half of the container?

28 Use the equation $S = k_B \ln \Omega$ and Figure B4.30 to determine a value for the change in entropy of a system of three particles when the volume in which they can move is doubled.

29 Calculate the entropy of a system which has 10^{30} microstates.

30 Calculate the total entropy change when 900 J of thermal energy is transferred from a hot reservoir at a constant 550 K to a cold reservoir at a constant 275 K.

31 **a** 3.34×10^4 J are needed to melt some ice at 0 °C. Determine the total change in entropy when the ice melts from thermal energy transferred from the air at a temperature of 25 °C.

b What assumptions did you make?

32 Figure B4.33 represents a Carnot cycle.

a How much thermal energy is transferred between states P and Q?

b How much thermal energy is transferred between R and S?

c How much useful work is done in each cycle?

d Use your answers for **a**, **b** and **c** to determine the efficiency of the process.

e Confirm that the same answer can be calculated from the temperatures involved.

f What physical quantity can be calculated from the area enclosed by the cycle?

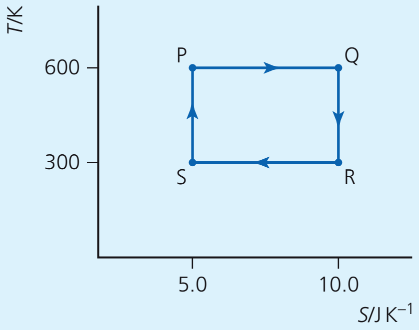


Figure B4.33 A Carnot cycle

33 There are four laws of thermodynamics, but only the first and second are included in this course. They can be summarized in the following humorous form:

- Zeroth: You must play the game.
- *First: You can't win.*
- *Second: You can't break even.*
- Third: You can't quit the game.

What are these comments on the first and second laws suggesting about energy?

Nature of science: Theories

Three versions of the same very important law

The second law of thermodynamics is considered by many physicists to be one of the most important principles in the whole of science. The following quote from Sir Arthur Stanley Eddington (*The Nature of the Physical World*, 1927) may help to convey the importance of this law:

'The law that entropy always increases holds, I think, the supreme position among the laws of nature. If someone points out to you that your pet theory of the Universe is in disagreement with Maxwell's equations – then so much the worse for Maxwell's equations. If it is found to be contradicted by observation – well, these experimentalists do bungle things sometimes. But if your theory is found to be against the second law of thermodynamics, I can give you no hope; there is nothing for it but to collapse in deepest humiliation.'

The second law of thermodynamics can be expressed in different ways depending on the context, and the three versions presented above are slightly different perspectives on the consequences of molecular disorder. Therefore, it is not surprising that, it in the nineteenth century when it was first formulated, the law was the subject of much attention and discussion between prominent scientists in different countries.

B.5

Current and circuits

◆ **Electric charge** Fundamental property of some subatomic particles that makes them experience electric forces when they interact with other charges. Charges can be **positive** or **negative** (SI unit: coulomb, C).

◆ **Opposite charge** Positive and negative charges are described as opposite charges.

◆ **Coulomb, C** The derived SI unit of measurement of electric charge.

◆ **Proton** Subatomic particle with a positive charge ($+1.6 \times 10^{-19}$ C).

◆ **Neutron** Neutral subatomic particle.

◆ **Nucleus** The central part of an atom containing protons and neutrons.

◆ **Electron** Elementary subatomic particle with a negative charge (-1.6×10^{-19} C) present in all atoms and located outside the nucleus.

◆ **Elementary charge, *e*,** 1.6×10^{-19} C

Guiding questions

- How do charged particles flow through materials?
- How are electrical properties of materials quantified?
- What are the consequences of resistance in conductors?

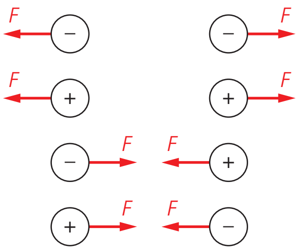
Electric charge

Electric charge is a fundamental property of some subatomic particles, responsible for the forces between them. (Details in Topic D.2.)

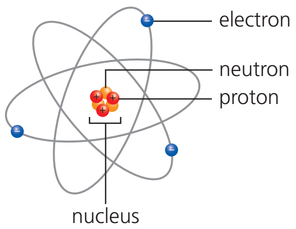
Because there are two kinds of force (attractive and repulsive) as seen in Figure B5.1, we need two kinds of charge to explain the different forces. We call these two kinds of charge, **positive charge** and **negative charge**. The description of charges as ‘positive’ or ‘negative’ has no particular significance, other than to suggest that they are two different types of the same thing. Positive and negative charges are often described as **opposite charges**.

Charges of opposite sign attract each other. Charges of same sign repel each other.

Charge is measured in **coulombs**, C. One coulomb is a relatively large amount of charge and we often use microcoulombs (1 μC = 10^{-6} C) and nanocoulombs (1 nC = 10^{-9} C).



■ **Figure B5.1** Electric forces between similar and opposite charges



■ **Figure B5.2** Simple model of an atom

Figure B5.2 shows a simple visualization of an atom, with three types of subatomic particle. The structure of atoms will be discussed in more detail in Topic E.1.

Protons and **neutrons** are to be found in the small central **nucleus** of the atom. **Electrons** are located in the space around the nucleus.

All protons have a positive charge of $+1.60 \times 10^{-19}$ C and all electrons have a negative charge of -1.60×10^{-19} C.

A charge of magnitude **1.60×10^{-19} C** is called the **elementary charge**. It is given the symbol *e*.

Since 2019, the elementary charge has been defined to be exactly $-1.602\,176\,634 \times 10^{-19}$ C.

◆ **Neutral** Uncharged, or zero net charge.

◆ **Variable** Quantity that can change during the course of an investigation. Variables can be **continuous** or **discrete**. A variable can be measurable (*quantitative*) or just observable (*qualitative*). A quantity being deliberately changed is called the *independent variable* and the measured, or observed, result of those changes occurs in a *dependent variable*. Usually, all other variables will be kept constant (as far as possible); they are called the *controlled variables*.

◆ **Quantized** Can only exist in certain definite (discrete) numerical values.

There are electric forces between the electrons and the protons in the nucleus. Neutrons do not have any charge; they are **neutral**.

Tool 3: Mathematics

Distinguish between continuous and discrete variables

A **continuous variable** can have, in theory, any value (within the available limits), but a **discrete variable** can only have certain values. An everyday example might be buying eggs: you can buy 1, 2, 6, 10 and so on, but not 1.5 or 3.7 eggs. Values of bank notes are another example. A physical quantity which can only have discrete values is described as being **quantized**.

Figure B5.3 a discrete number of eggs

Any quantity of charge consists of a whole number of charged particles, each ±1.60 × 10^{-19} C. Intermediate values are not possible (with the exception of the sub-nuclear particles *quarks*, but they are not included in this course). For example, it is not possible to have charge with a value of 4.00 × 10^{-19} C. We describe this by saying that charge is quantized.

Charge is generally given the symbol q. (Q is also sometimes used, but the same symbol is used for thermal energy.)

One coulomb of negative charge is the total charge of 6.24×10^{18} electrons: $\frac{1}{1.602 \times 10^{-19}}$

Law of conservation of charge

This is one of the few conservation rules in physics (rules which are always true):

The total charge in an isolated system remains constant.

◆ **Ionization** The process by which an atom or molecule becomes an ion. The required energy is called the ionization energy.

For example, if one or more negatively charged electrons are removed from a neutral atom, this law shows us that the remaining atom must have an equal positive charge. The charged atom is then called an ion and the process is called **ionization**.

TOK

The natural sciences

- Should scientific research be subject to ethical constraints or is the pursuit of all scientific knowledge intrinsically worthwhile?

Increasing knowledge could be life-threatening.

Benjamin Franklin (b 1706) was a famous and influential personality in the USA in the eighteenth century. His experiments with static electricity (see Figure B5.4) certainly endangered lives, but his experiments expanded our scientific knowledge. Travelling into orbit, or to the Moon, are among scientific investigations which could be described similarly.

It can be argued that, if individuals are fully informed and prepared to risk their lives for scientific advancement, then that is entirely their own personal choice. Experiments with animals is another matter, and there is a wide range of well-considered opinions on this matter. Opinions may vary with the type of animal involved and the possible benefits to human society. One notorious demonstration of the effects of poisoning and electricity involved a ‘rogue’ elephant named Topsy in 1903.

Figure B5.4 Benjamin Franklin famously flew a kite in a lightning storm as part of his investigations into electricity

Electric currents

SYLLABUS CONTENT

- Direct current (dc), I, as a flow of charged carriers given by: $I = \frac{\Delta q}{\Delta t}$.

◆ **Current (electric), I** A flow of electric charge. Equal to the amount of charge passing a point in unit time: $I = \frac{\Delta q}{\Delta t}$.

◆ **Charge carrier** A charged particle which is free to move (mobile).

◆ **Delocalized electrons** Electrons which are not bound to any particular atom or molecule. Sometimes called **'free' electrons**.

Whenever charges flow from place to place we describe it as an **electric current**. We have identified electrons, protons and ions as charged particles, but in an electric current the term **charge carrier** is often used to describe any moving charge. In order for a charged particle to flow as part of an electric current it has to be relatively 'free' to move. We often refer to *mobile* charge carriers.

Some of the electrons in the atoms of metals have enough energy that they are no longer attracted to a particular metal ion. We say that they are **delocalized electrons**, or **free electrons**. This means that electric currents, carried by *free* electrons, can flow through metals better than through most other materials.

In this topic we will only be considering currents in which the charge carriers are electrons. Figure B5.5 represents the electric current of a flow of free electrons through a metal wire. The negatively charged electrons are attracted to the positive terminal of a battery.

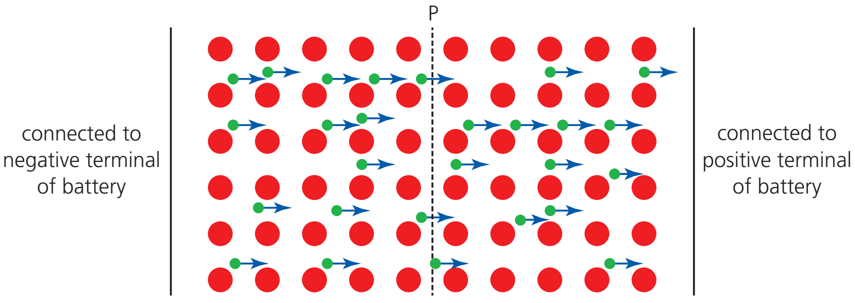


Figure B5.5 Electric current flowing through a metal

When there is no electric current, free electrons normally move around randomly in metals at high speeds, somewhat like molecules in a gas. But, when they form an electric current, a much slower 'drift' speed in the direct of the current is added to the electrons' random movements. This movement cannot be represented in a single diagram such as Figure B5.5.

We define the magnitude of an electric current (given the symbol I) as the amount of charge that passes a point (such as P in Figure B5.5) in unit time:

$$\text{electric current, } \boldsymbol{I} = \frac{\boldsymbol{\Delta q}}{\boldsymbol{\Delta t}}$$

◆ **Ampere, A** SI (fundamental) unit of electric current. $1\,\text{A} = 1\,\text{C}\,\text{s}^{-1}$.

◆ **Direct current (dc)** A flow of electric charge that is always in the same direction.

◆ **Alternating current (ac)** A flow of electric charge that changes direction periodically.

◆ **SI system of units** International system of standard units of measurement (from the French 'Système International') which is widely used around the world. It is based on seven fundamental units and the decimal system.

The SI unit for electric current is the **ampere** (amp), A. Milliamps (mA) and microamps (µA) are also in common use.

When a current only flows in one direction it is called a **direct current** (dc). **Alternating currents** (ac) continuously change direction. They are discussed in Topic D.4.

The ampere (amp) is one of the seven base units of the **SI system**. It is defined to be the current in which 1 C of charge (6.24×10^{18} electrons) passes a point in one second. (Before 2019 the amp was defined differently and more obscurely: as the current in two straight parallel wires of infinite length exactly one metre apart in a vacuum, which results in a magnetic force between them of exactly $2 \times 10^{-7}\,\text{N}\,\text{m}^{-1}$. This is explained in Topic D.3 and need not be understood here.)

Top tip!

Direct currents are generally more useful than alternating currents, but ac is used for transmitting electrical energy around the world as it is more easily transformed to the high voltages needed to reduce energy dissipation in the wires (discussed briefly in Topic D.3).

WORKED EXAMPLE B5.1

The current through an LED desk lamp is 50 mA.

a Calculate the amount of charge which flows through the lamp in 1.0 minute.

b How many electrons flow through the lamp every minute?

Answer

a $I = \frac{\Delta q}{\Delta t} = 50 \times 10^{-3} = \frac{\Delta q}{60}$

$\Delta q = 3.0\,\text{C}$

b $\frac{3.0}{1.60 \times 10^{-19}} = 1.9 \times 10^{19}$

Electrical circuits

Circuit diagrams represent the arrangement of components in a circuit.

In order for there to be a continuous flow of current, **electrical components** and wires need to form a complete (closed) loop, called an **electrical circuit**. However, a current cannot flow unless there is a battery (or other electrical power source) included in the circuit, as shown in Figure B5.6, which is drawn in the conventional style.

A single battery is better described as an **electric cell**. When more than one cell is used, the combination is called a **battery**, although in everyday language, one cell is commonly called a battery.

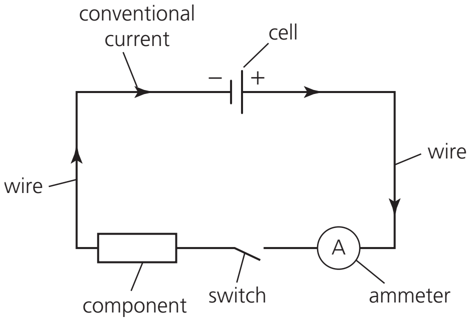


Figure B5.6 Simple electric circuit

The two **terminals** of any battery are labelled positive and negative, and it may be considered that the positive terminal attracts free electrons from the circuit, and the negative terminal repels free electrons. In this way electrons will move around the circuit shown in Figure B5.6 in an anticlockwise direction. However, for historical reasons:

Electric current is *always* shown flowing from positive to negative around any circuit.

This is shown by the arrows in Figure B5.6. This is known as the direction of **conventional current** flow. It was chosen a long time before electrons had been discovered.

Tool 2: Technology

Applying technology to collect data

The magnitudes of electric currents are determined by instruments called **ammeters**, which are connected so that all of the current to be measured flows through them, as shown in Figure B5.6. (This is called being connected *in series*.) Connecting an ammeter in a circuit should not reduce the magnitude of the current it is measuring. Therefore, an **ideal ammeter** will have zero *resistance* to the flow of a current through it. An electronic current sensor responds to the magnetic fields which exist around all currents.

When measuring dc, the current must flow through the ammeter in the correct direction. This is shown by marking the two terminals as positive and negative. Moving through the circuit from the positive terminal on the ammeter, you should arrive at the positive terminal of the battery (or other voltage supply).

◆ **Circuit (electrical)** A complete conducting path that enables an electric current to continuously transfer energy from a voltage source to various **electrical components**.

◆ **Cell (electric)** Device that transfers chemical energy to the energy carried by an electric current.

◆ **Battery** One or more electric cells.

◆ **Terminals (electrical)** Points at which connecting wires are joined to electrical components.

◆ **Conventional current** The direction of flow of a direct current is always shown from the positive terminal of the power source, around the circuit, to the negative terminal. Conventional current is opposite in direction from electron flow.

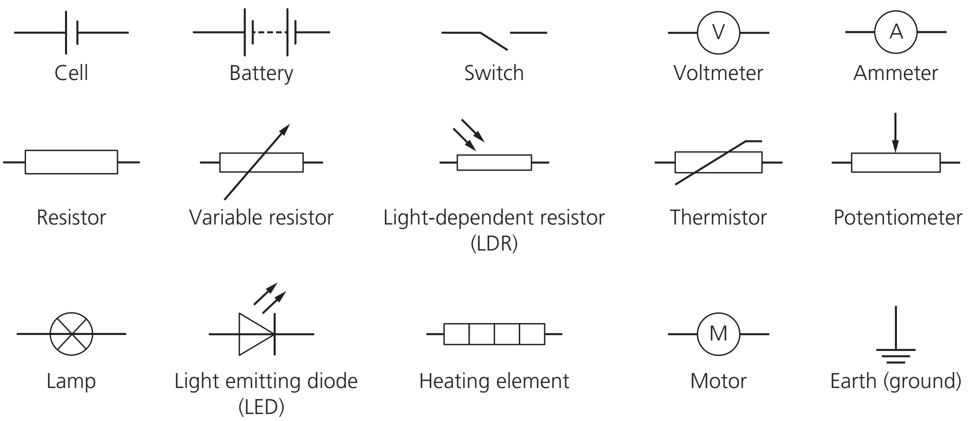
◆ **Ammeter** Instrument that measures electric current.

◆ **Ideal meters** Meters with no effect on the electrical circuits in which they are used. An **ideal ammeter** has zero resistance, and an **ideal voltmeter** has infinite resistance.

Top tip!

In order to understand electrical circuits, it is usually better to consider that the current does *not* begin or end anywhere in particular (at the battery, for example). It is better to consider that the current flows at the same time throughout the circuit, which should be considered as a whole.

(In reality, an electric field moves around the circuit [setting electrons into motion] at a speed close to the speed of light.)



■ **Figure B5.7** Complete list of circuit symbols shown in the IB Physics Data Booklet

Nature of science: Science as a shared endeavour

The use of common symbols and units

The communication of scientific information and ideas between different countries and cultures can be affected by language problems, but this is greatly helped by the use of standard symbols for physical quantities (and units) and for electrical components. Increasingly, English is being used as the international language of science but, naturally, there are many individuals, organizations and countries who prefer to use their own language. Imagine the confusion and risks that could be caused by countries using totally different symbols and languages to represent the circuitry on, for example, a modern international aircraft.

A famous incident occurred in a commercial aircraft flight over Canada in 1983, when the aircraft ran out of fuel because of confusion over the units of volume used for the fuel measurements. Fortunately, there were no serious injuries.

LINKING QUESTION

- In what ways can an electrical circuit be described as a system like the Earth's atmosphere or a heat engine?

This question links to understandings in Topics B.2 and B.4.

1 A carbon atom (carbon-12) contains six protons, six neutrons and six electrons.
  a Sketch this atom in a diagram similar to that seen in Figure B5.2.
  b Calculate the total positive charge in the atom.
  c What is the total negative charge in the atom?
  d We can describe the atom as *neutral*. Explain what this means.
  e If the atom is ionized by the removal of one electron, what is the charge on the ion?

2 2.5×10^{20} electrons flow through a television in one minute.
  a Calculate the total charge which flows in that time.
  b Determine the electric current.

3 Most people find the shape of electricity pylons (that transmit electricity around countries) ugly and would say they spoil the natural beauty of the landscape, for example Figure E5.8.

Figure B5.8 Electricity power lines cross some of the most remote countryside in the world

- **a** If the charge flowing through a point on an overhead power line every hour is three million coulombs, what is the current?
- **b** Some architects have suggested that pylons could be designed with interesting and attractive structures that are more sympathetic to the environment, but such pylon designs are more expensive than the usual designs and many people will not be happy to pay more for their electricity. Sketch a pylon design for your country that is attractive, practicable and probably not too expensive.

4 Explain what is meant by a *free* or *delocalized* electron.

5 Determine the current at point P in Figure B5.9. State its direction.

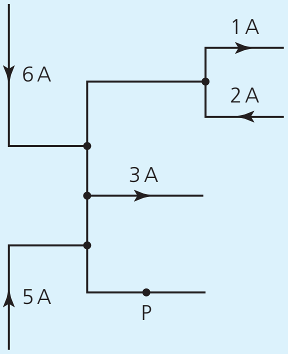


Figure B5.9

Potential difference / voltage

SYLLABUS CONTENT

- ▶ The electric potential difference, V, is the work done per unit charge on moving a positive charge between two points along the path of the current: $V = \frac{W}{q}$.

◆ **Voltage** See *potential difference*.

◆ **Volt** Derived unit of measurement of potential difference. 1 V = 1 J C^{-1}.

◆ **Potential difference, V** The energy transfered by unit positive charge (1 C) moving between two points. Commonly referred to as voltage.

The term **voltage** is familiar to everyone. Electricity is usually provided to our homes at 110 V or 230 V, and batteries of various lower voltages are used to provide energy to electronic devices.

The voltage of a battery, or other source of electrical energy, is a measure of how much *energy* it can supply to the charge carriers flowing through it. One **volt** means that one joule of energy is transferred by each coulomb of charge moving *between* two specified points.

1 V = 1 joule/coulomb (J C^{-1})

Voltage has become the widely used term for the physical quantity that is measured by volts. However, the correct term is **potential difference**, commonly shortened to p.d. The symbol V is used for potential difference, the same letter as used for its unit, V.

The electric potential difference, V, is the work done per unit charge on moving a positive charge between two points along the path of the current: $\boldsymbol{V = \frac{W}{q}}$.

In this topic we are only concerned with free electrons moving in electric circuits, but in Topic D.2, for HL students, we will discuss the movements of both positive and negative charges more generally. When we consider potential differences then, we will need to consider the nature of the charge and the direction of movement more carefully.

We may refer to the potential difference *across* a battery (or other electrical energy source) that is *supplying* energy to a circuit, or to the potential difference across any component(s) that is *using* energy in the circuit.

Tool 2: Technology

Applying technology to collect data

Potential differences are measured using voltage sensors (various designs) or **voltmeters**, which are connected as shown by the many voltmeters in Figure B5.10. A voltmeter is always connected *across* (in parallel with) the component(s) it is checking. An **ideal voltmeter** has infinite resistance, so that no current flows through it and it does not affect the p.d. it is measuring. As with ammeters, voltmeters must be connected correctly when using dc circuits: moving through the circuit from the positive terminal on the voltmeter, you should arrive at the positive terminal of the battery (or other voltage supply).

Considering Figure B5.10, if the p.d. supplied to the circuit, V_s is 12 joules to every coulomb (12 V), then when the switch is closed, $V_1+V_2+V_3$ must also equal 12 V (J/C), because the energy transferred into the circuit must be equal to the energy 'used' by the components in the circuit. It is assumed that no energy is transferred in the connecting wires or battery.

Suppose $V_1 = 3$ V and $V_2 = 4$ V, then V_3 must equal 12 − 3 − 4 = 5 V. V_{12} is measuring the same as $V_1 + V_2$, so it should read 7 V.

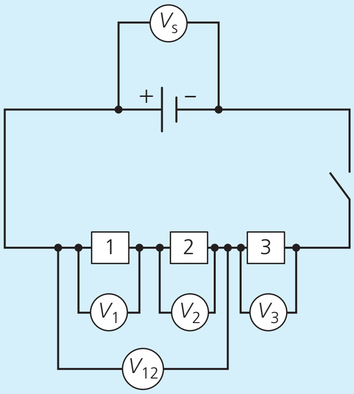


Figure B5.10 Connecting voltmeters

◆ **Voltmeter** An instrument used to measure potential difference (voltage).

◆ **Observer effect** When the act of observation, or measurement, changes the phenomenon being observed.

6 Outline what would happen in the circuit shown in Figure B5.6 if the ammeter was replaced by a voltmeter.

7 Explain why replacing the voltmeter, V_s (Figure B5.10) with an ammeter would be a bad idea.

8 Consider the circuit shown in Figure B5.10. If the battery supplied 12 V, the reading on V_1 was 2 V and V_{12} showed a voltage of 5 V, state the readings on the other three voltmeters.

9 400 C of electric charge flow through a lamp in one hour in a country where the electricity mains are supplied at 230 V.
   - a Calculate the current in the lamp.
   - b How much energy is supplied from the mains to the lamp in this time?

Nature of science: Measurements

The observer effect

We have discussed the use of ammeters and voltmeters to make electrical measurements and referred to the use of 'ideal' meters which will not affect the values of the currents and voltages that they are measuring. However, when taking *any* scientific measurement, we need to consider the possibility that the act of taking the measurement will change what is being measured.

When measuring the pressure in a car tyre, as in Figure B5.11, some of the air in the tyre must flow into the pressure gauge. This will result in a reduction of pressure in the tyre, although it will probably be a very small change.

Many types of thermometer need to absorb or emit thermal energy until they reach thermal equilibrium with their surroundings. This may affect the temperature of the locations that they are measuring. (An infrared thermometer does not have this problem.)

As a non-physics example, doctors are well aware that a patient's blood pressure may well rise when it is being measured because of psychological effects.

Figure B5.11 Measuring the air pressure in a car tyre

Using a gravitational analogy may help to explain potential difference

If you are taking the Standard Level examination, you do not need to worry about the deeper meaning of the term *potential difference*; just think of it as *voltage*, and you do not need to remember the following explanation.

A mass will fall towards Earth if it is free to do so. We have explained this by stating that there is a gravitational force acting downwards on it. It then moves from a position of higher gravitational potential energy to a position of lower gravitational potential energy. Alternatively, we can say that any mass may move because of a *difference in gravitational potential* (*energy*).

(The concept of *potential,* defined as potential energy/mass, is introduced for HL students in Topics D.1 and D.2.)

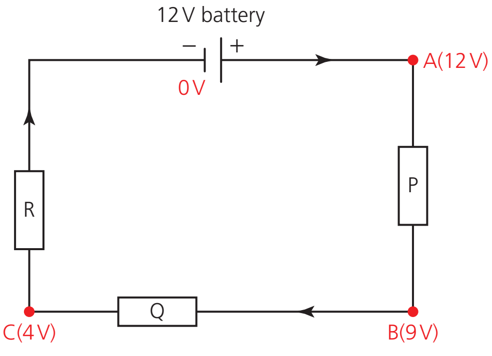


Figure B5.12 Potentials at points around a series circuit

The analogy between gravitational fields and electric fields (introduced in Topics D.1 and D.2) is useful. We will use it now:

A charge will move because of a *difference in electrical potential (energy)* – potential difference, shortened to p.d. – if it is free to do so. It will move from higher electrical potential energy to lower electrical potential energy. A battery provides this p.d. In a similar way, a pump can raise water to a greater height, increasing its gravitational potential energy, but if the water is free to move, it will then fall back down.

Consider the circuit shown in Figure B5.12. We can label the negative side of the battery with a voltage (potential) of 0 V. Since there is a p.d. of 12 V across the battery, the voltage at point A is 12 V. As we move around the circuit, the voltage decreases. For example, it could be 9 V at B and 4 V at C.

A voltmeter connected across P will record a p.d. of (12 – 9) = 3 V.

A voltmeter connected across Q will record a p.d. of (9 – 4) = 5 V.

A voltmeter connected across R will record a p.d. of (4 – 0) = 4 V.

3 + 5 + 4 = 12 V.

Tool 1: Experimental techniques

Recognize and address relevant safety, ethical and environmental issues

Using electrical circuits is an important part of any physics course. For obvious safety reasons, you should work with low voltages. This is done with batteries providing, typically, 9 V or less, but most commonly 1.5 V. However, there is an environmental impact here, as the large number of batteries eventually need to be disposed of. Low voltage (LT) adjustable supplies typically provide 12–15 V. They are connected to the mains supply and need to be checked regularly, including ‘earthing’, for safety.

Electrical resistance

SYLLABUS CONTENT

- Electric resistance and its origin.
- Electrical resistance given by: $R = \frac{V}{I}$.
- Ohm’s law.
- Ohmic and non-ohmic behaviour of electrical conductors, including the heating effect of resistors.
- Properties of electrical conductors and insulators in terms of mobility of charge carriers.

◆ **Conductor (electrical)** A material through which an electric current can flow because it contains significant numbers of mobile charges (usually free electrons).

◆ **Resistance (electrical)** Ratio of potential difference across a conductor to the current flowing through it. $R = \frac{V}{I}$ (SI unit: ohm, Ω).

◆ **Insulator (electrical)** A non-conductor. A material through which a (significant) electric current cannot flow, because it does not contain many charge carriers.

When the same potential difference (voltage) is connected across different electrical components, the currents produced will vary. In general, if the currents are relatively large, the components are described as good **electrical conductors** with low **electrical resistance**. If the currents are small, or negligible, the material is described as a good **electrical insulator**, with a high electrical resistance.

A few substances, most notably silicon and germanium, are described as **semiconductors** because their ability to conduct electricity falls between the obvious conductors and insulators. The electrical behaviour of these materials provides the basis of the electronics industry.

Good electrical conductors, metals, are usually also good thermal conductors. This is because free electrons are important in both processes.

To discuss the origin of electrical resistance we can refer back to Figure B5.5. The greater the number of mobile change carriers (free electrons) in a given volume of the material, the lower we would expect the resistance to be. When the free electrons move through the conductor they will collide / interact with the vibrating metal ions and this is the cause of electrical resistance. We know that particle vibrations in a solid decrease at lower temperatures, so resistance can be reduced by cooling a metal. Conversely, the resistance of a metal will increase if it gets hotter.

Metals are good conductors because they have a large number of mobile charge carriers (free electrons) in unit volume. The vibration of metal ions creates resistance to the flow of electrons.

Electrical resistance, *R*, is defined quantitatively as follows:

$$\text{electrical resistance} = \frac{\text{p.d.}}{\text{current}} \qquad \boldsymbol{R = \frac{V}{I}}$$

SI unit: **ohm, Ω** ($1\,\Omega = 1\,\text{V}\,\text{A}^{-1}$)

◆ **Semiconductor** Material (such as silicon) with a resistivity (explained later in this section) between that of conductors and insulators. Such materials are essential to modern electronics.

◆ **Ohm, Ω** The derived SI unit of electric resistance. $1\,\Omega = 1\,\text{V}/1\,\text{A}$.

◆ **Fundamental units** Units of measurement that are not defined as combinations of other units.

◆ **Derived units** Units of measurement that are defined in terms of other units.

Tool 3: Mathematics

Apply and use SI units

There are seven **fundamental (basic) units** in the SI system: kilogram, metre, second, ampere, mole, kelvin (and candela, which is not part of this course). The quantities, names and symbols for these fundamental SI units are given in Table B5.1.

They are called 'fundamental' because their definitions are not combinations of other units (unlike metres per second, or Newtons, for example). You are not expected to learn the definitions of these units.

■ **Table B5.1** Fundamental SI units used in this course

Quantity	**Name**	**Symbol**
length	metre	m
mass	kilogram	kg
time	second	s
electric current	ampere	A
temperature	kelvin	K
amount of substance	mole	mol

Derived units of measurement

All other units in science are combinations of the fundamental units. For example, the unit for volume is m^3 and the unit for speed is $\text{m}\,\text{s}^{-1}$. Combinations of fundamental units are known as **derived units**.

Sometimes derived units are also given their own name (Table B5.2). For example, the unit of force is kg m s^{-2}, but it is usually called the newton, N. All derived units will be introduced and defined when they are needed during the course.

Table B5.2 Some named derived units

Derived unit	Quantity	Combined fundamental units
newton (N)	force	kg m s^{-2}
pascal (Pa)	pressure	$\text{kg m}^{-1}\,\text{s}^{-2}$
hertz (Hz)	frequency	s^{-1}
joule (J)	energy	$\text{kg m}^2\,\text{s}^{-2}$
watt (W)	power	$\text{kg m}^2\,\text{s}^{-3}$
coulomb (C)	charge	A s
volt (V)	potential difference	$\text{kg m}^2\,\text{s}^{-3}\,\text{A}^{-1}$
ohm (Ω)	resistance	$\text{kg m}^2\,\text{s}^{-3}\,\text{A}^{-2}$

Note that you are expected to write and recognize units using superscript format, such as m s^{-1} rather than m/s. The unit for acceleration, for example, should be written m s^{-2}, not m/s^2.

Express a derived unit in terms of fundamental units

The ohm is a derived unit, but all derived units can be reduced to their fundamental components:

$$\text{V} = \text{J C}^{-1} = \text{Nm} \times \text{C}^{-1} = \text{kg m s}^{-2} \times \text{m} \times \text{A}^{-1} \times \text{s}^{-1} = \text{kg m}^2\,\text{s}^{-3}\text{A}^{-1}$$

$$\Omega = \frac{\text{V}}{\text{A}} = \frac{\text{kg m}^2\,\text{s}^{-3}\,\text{A}^{-1}}{\text{A}} = \text{kg m}^2\,\text{s}^{-3}\text{A}^{-2}$$

LINKING QUESTIONS

- How does a particle model allow electrical resistance to be explained? (NOS)
- What are the parallels in the models for thermal and electrical conductivity? (NOS)

These questions link to understandings in Topic B.1.

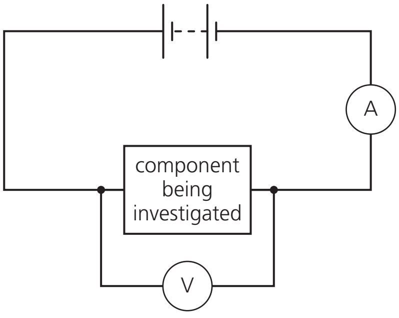


Figure B5.13 Determining the resistance of a component

Determining resistance values experimentally

The resistance of a component can be determined as shown in Figure B5.13, recording a pair of values for p.d., V, and current, I.

WORKED EXAMPLE B5.2

The current through an electrical component was 0.78 A when a p.d. of 4.4 V was applied across it. Calculate its resistance.

Answer

$$R = \frac{V}{I} = \frac{4.4}{0.78} = 5.6\,\Omega$$

The value of the resistance obtained should not be assumed to be constant. It may be, but resistances can also change depending on other factors, as discussed later.

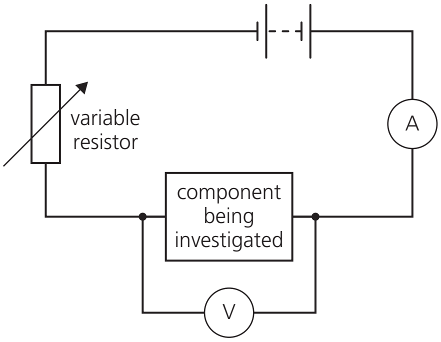


Figure B5.14 Varying current and potential difference using a variable resistor

I-V characteristics

The resistive properties of a component can be fully investigated by measuring the values of range of different currents produced by varying the p.d. across it. This can be done by using the circuit shown in Figure B5.13 but replacing the battery with a source of *variable* voltage. Alternatively, if only a fixed voltage supply is available, the circuit used in Figure B5.14 can be used. See the worked example later in this topic – in the *Using variable resistors* section (which also explains how using a variable resistor as *potentiometer* is the best method).

◆ ***I–V* characteristic:** Graph of current–p.d., representing the basic behaviour of an electrical component.

◆ **Ohm's law** The current in a conductor is proportional to the potential difference across it, provided that the temperature is constant.

◆ **Ohmic (and non-ohmic) behaviour** The electrical behaviour of an ohmic component is described by Ohm's law. A non-ohmic device does not follow Ohm's law.

◆ **Filament lamp** Lamp that emits light from a very hot metal wire. Also called an incandescent lamp.

Common mistake

Resistance cannot be determined from the gradient of a p.d.–current graph, unless the component is ohmic.

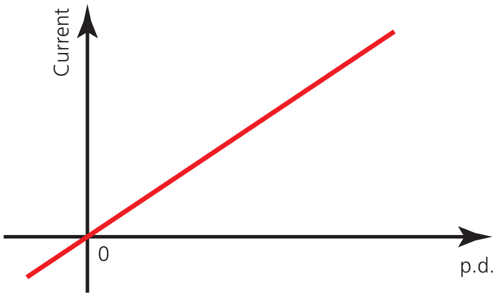
The results of these experiments can be shown on current–p.d. graphs. They are called ***I–V* characteristics**.

I–V characteristics are the best way to represent the electrical behaviour of components.

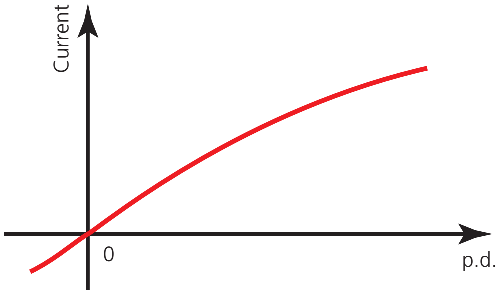
The simplest possible relationship is that the current and p.d. are proportional to each other, as shown in Figure B5.15. This relationship shows that the resistance ($R = V/I$) is constant. The relationship is called **Ohm's law**, and any component that behaves like this is described as being **ohmic**. Metal wires at constant temperatures are ohmic.

Ohm's law: at constant temperature, the current through a metallic conductor is proportional to the p.d. across it: $I \propto V$

Figure B5.15 should be compared to Figure B5.16 which shows the *I–V* characteristic of a metal wire that gets hot. The most common example of this type of **non-ohmic** behaviour is shown by a **filament lamp**. If we took pairs of values for V / I from Figure B5.16, it would show that the resistance ($R = V / I$) increases when the current is greater. This is because, when the current is greater, there are more collisions / interactions between the free electrons and the vibrating metal ions. So that more energy is transferred to the ions, their vibrations increase and the temperature rises.



■ **Figure B5.15** Ohm's law for an ohmic resistor



■ **Figure B5.16** A current–p.d. graph for metal wire that gets hot, such as a filament lamp

ATL B5A: Research skills, thinking skills

Use search engines and libraries effectively; provide a reasoned argument to support a conclusion

Types of lighting

■ **Figure B5.17** Incandescent filament lamp

Incandescent electrical lamps like that shown in Figure B5.17 were the most popular means of lighting throughout the world for more than one hundred years. Because they need to get very hot to emit light, incandescent lamps are very *inefficient*. They have been replaced by more efficient **fluorescent** lighting and, especially, LED lighting. LED stands for **light emitting diode**.

A **diode** is an electrical component that allows an electric current to pass through it in only one direction. Modern diodes are made from semiconductors, and some of these have the very useful property of emitting light when a current is passing through them. Figure B5.18 shows the *I–V* characteristic of a typical diode.

◆ **Incandescent** Emitting light when very hot.

◆ **Fluorescent lamp** Lamp that produces light by passing electricity through mercury vapour at low pressure.

◆ **Light-emitting diodes (LEDs)** Small semiconducting diodes that emit light of various colours at low voltage and power.

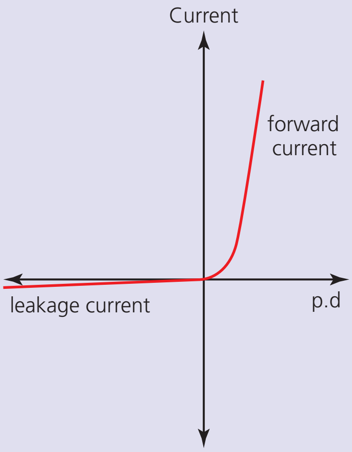
◆ **Diode** An electrical component that only allows current to flow in one direction.

Top tip!

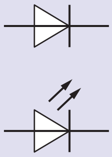
Negative values of current and p.d. represent values in the reverse directions, obtained by turning the battery (or other energy source) the other way around. Simple resistors will behave the same way for currents in either direction, but many other components, diodes for example, need to be connected the 'right' way around.

When connected in one way, a 'forward' current is produced, and the diode has very little resistance, as long as the p.d. is greater than a certain minimum value. When connected the other way around, the diode has a large resistance, although a small 'leakage' current is possible. The arrowhead on the circuit symbol shows the 'forward' direction for the current (Figure B5.19).

Individual LEDS are low voltage components and a number of them must be connected *in series* to produce lighting bright enough for a whole room. See Figure B5.20.



■ **Figure B5.18** I–V characteristic for a diode



■ **Figure B5.19** The circuit symbols for a diode and a LED

■ **Figure B5.20** Rings of small LEDs in a ceiling light

Search online to find the relative efficiencies of incandescent, fluorescent and LED lighting. What kinds of lighting do you use at home?

◆ **Peer review** Evaluation of scientific results and reports by other scientists with expertise in the same field of study.

Nature of science: Science as a shared endeavour

Peer review or competition between scientists?

Georg Ohm's famous law was published in Germany in 1827 in the form that the current in a wire, I, is proportional to $(A/L)V$ where A is the cross-sectional area of a uniform metal wire of length L. Two years earlier, in England, Peter Barlow had incorrectly proposed 'Barlow's law' in the form I was proportional to $\sqrt{(A/L)}$, but with no reference to the key concept of voltage, V. It is not unusual for two or more different scientists, or groups of scientists, to be investigating similar areas of science at the same time, often in different countries.

In the worldwide, modern scientific community, with its quick and easy mass communication, new experimental results and theories are quickly subjected to close scrutiny. New ideas are reviewed carefully by other scientists and experts working in the same field, in a process called **peer review**. But 200 years ago, when Ohm was carrying out his research, things were very different. At that time, social factors and the reputation, power and influence of the scientist were sometimes as important in judging new ideas as the value of the work itself. The story of Barlow and Ohm is particularly interesting because in the early stages, the incorrect theory proposed by Barlow was more widely believed.

■ **Figure B5.21** Georg Ohm

Resistors

All electrical components have resistance, although the resistance of some things, for example connecting wires and ammeters, usually have **negligible** resistance. A component manufactured for its specific resistive properties is called a **resistor**. Resistors are important components in all electrical circuits. Figure B5.22 shows the appearance of a few typical fixed resistors.

◆ **Negligible** Too small to be significant.

◆ **Resistor** A resistance made to have a specific value or range of values.

◆ **Variable resistor** A resistor (usually with three terminals) that can be used to control currents and/or potential differences in a circuit.

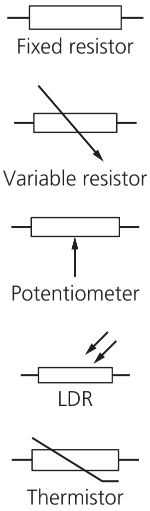
◆ **Potentiometer** Variable resistor (with three terminals) used as a potential divider. (see later)

◆ **Light-dependent resistor (LDR)** A resistor, the resistance of which depends on the light intensity incident upon it.

◆ **Thermistor** (negative temperature coefficient) A resistor that has less resistance when its temperature increases. Also called a temperature-dependent resistor

■ **Figure B5.22** Fixed value resistors

Apart from resistors of fixed value, later in this topic we will discuss the use of **variable resistors**, **potentiometers**, **light-dependent resistors (LDRs)** and **thermistors**. Their circuit symbols are shown in Figure B5.23.



■ **Figure B5.23** Circuit symbols for resistors: fixed, variable, potentiometer, LDR, thermistor

10 What voltage is needed to make a current of 56 mA pass through a 675 Ω ohmic resistor?

11 **a** Calculate the operating resistance of a 230 V domestic water heater if the current through it is 8.4 A.

b Explain why you would expect that the resistance would be less when it is first turned on.

12 What current flows through a 3.7 kΩ resistor when there is a p.d. of 4.5 V across it?

13 Calculate the p.d. across a 68.0 Ω resistor if 120 C of charge flows through it in 60 s.

14 Explain what it means if a component is described as non-ohmic.

15 State one reason why the resistance of a component may

a increase as it gets hotter

b decrease as it gets hotter.

16 Sketch an *I–V* characteristic for a component that has a resistance which decreases as the current through it gets larger.

Electrical resistivity

SYLLABUS CONTENT

▶ resistivity as given by $\rho = \frac{RA}{L}$

Investigations into how the resistances of metal wires depends on their dimensions can be undertaken using a circuit similar to that shown in Figure B5.13 or Figure B5.14. The currents should be kept low (or turned on for only short times) to avoid any significant temperature changes in the wires. Three important conclusions can be reached:

1 The resistance of a uniform wire is proportional to its length, *l*

(Assuming the wire has constant thickness and does not change temperature.) As shown by the example in Figure B5.24.

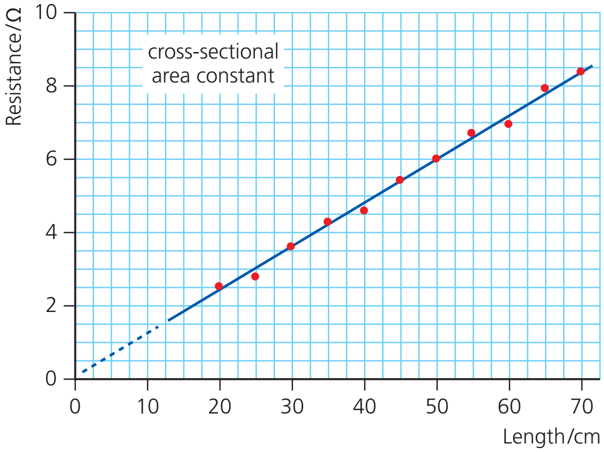


Figure B5.24 Variation of resistance with length of a metal wire

2 The resistance of a uniform wire is inversely proportional to its cross-sectional area, *A*

(Assuming the wire has constant length and does not change temperature.) As shown by the example in Figure B5.25. If the same data is used to plot a resistance–1/area graph, it will produce a straight line through the origin.

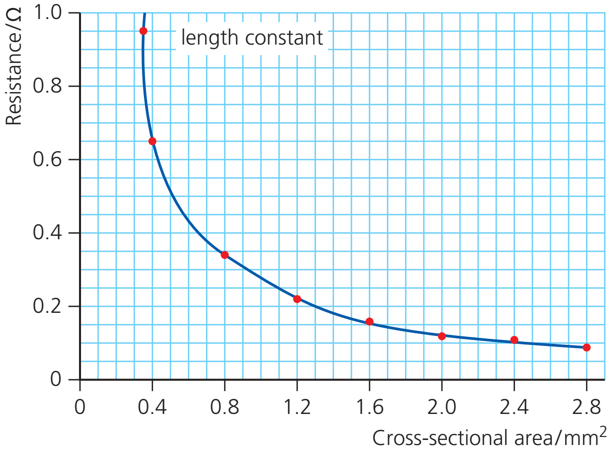


Figure B5.25 Variation of resistance with cross-sectional area of a metal wire

3 The resistance of a uniform wire depends on the metal from which it is made

Combining the last two results, we get:

$$R \propto \frac{l}{A}$$

or:

$$R = \text{constant} \times \frac{l}{A}$$

where the value of the constant depends on the resistive properties of the particular metal.

The constant is called the **resistivity** of the metal, and it is given the symbol *ρ*.

◆ **Resistivity, *ρ*** Resistance of a specimen of a material that has a length of 1 m and cross-sectional area of 1 m².

$$\textbf{resistivity, } \boldsymbol{\rho} = \frac{\boldsymbol{RA}}{\boldsymbol{L}} \qquad \text{SI unit: } \Omega\,\text{m}$$

Common mistake

Note that the SI unit for resistivity is Ω m. Many students think (wrongly) that the unit is ohms per metre, $\Omega\,m^{-1}$.

Clearly, the resistance of a material, like a wire, depends on its shape. For this reason, we cannot refer to the resistance of, for example, aluminium, because we have not specified its shape. The resistivity of a material can be considered as the resistance of a length of one meter, with a cross-sectional area of $1\,m^2$. In other words, the resistance of a cube of the material with sides of 1 m. This is a very large piece of a material, so it is not surprising that the resistivities of good conductors have very low values in SI units. See Table B5.3, which also includes the very high resistivities of some good insulators.

■ **Table B5.3** Resistivities of various substances at 20 °C

Material	**Resistivity/Ω m**
silver	1.6×10^{-8}
copper	1.7×10^{-8}
aluminium	2.8×10^{-8}
iron	1.0×10^{-7}
nichrome (used for electric heaters)	1.1×10^{-6}
carbon (graphite)	3.5×10^{-5}
germanium	4.6×10^{-1}
sea water	$\approx 2 \times 10^{-1}$
silicon	6.4×10^{2}
glass	$\approx 10^{12}$
quartz	$\approx 10^{17}$
Teflon (PTFE)	$\approx 10^{23}$

Variation of resistivity with temperature

As already explained, the resistivity of metals will increase with temperature because of the increased vibrations of the metal ions. The number of free electrons (charge carriers) in metals will not increase significantly unless temperatures are extreme. However, it can be very different with semiconductors and insulators.

The number of charge carriers (per cubic metre) in non-metals can increase significantly with rising temperatures, so that their resistance can decrease considerably as they get hotter. For example, glass is usually described as an insulator, but at 500 °C many types of glass can become good conductors.

WORKED EXAMPLE B5.3

a Determine the resistance of a nichrome wire at 20 °C, if it has a length of 1.96 m and a radius of 0.21 mm.

b Explain why the answer would be different at 100 °C.

Answer

a $$\rho = \frac{RA}{L}$$

Using data from Table B5.3:

$$1.1 \times 10^{-6} = \frac{R \times \left(\pi \times (0.21 \times 10^{-3})^2\right)}{1.96}$$

$R = 16\,\Omega$

b Increased vibrations of metal ions would cause an increase in resistance of the wire.

17 If the wire used to produce the results shown in Figure B5.24 had a resistivity of $4.9 \times 10^{-7}\,\Omega$m, calculate the cross-sectional area of the wire.

18 a Use values taken from the graph in Figure B5.25 to show that the resistance was inversely proportional to the area.

b If the wire had a length of 0.56 m, determine its resistivity.

19 Calculate the length of aluminium wire which will have the same resistance as a 1.0 m length of copper wire of the same thickness.

20 The central cable of a high voltage power cable, as seen in Figure B5.26, is made from aluminium and has an effective cross-sectional area of $3.4\,\text{cm}^2$.

a Predict what length of this cable will have a resistance of $1.0\,\Omega$.

b Suggest a reason why the cable is made of thinner strands of aluminium, rather than a single, thicker wire.

Figure B5.26 Power cable

21 A glass rod of length 10 cm and diameter 0.50 cm was heated until its resistance became $10\,\Omega$. Estimate the resistivity of the glass at this high temperature.

22 a Calculate the ratio of highest / lowest resistivity as seen in Table B5.3.

b Explain the difference.

Connecting two or more components in the same circuit

SYLLABUS CONTENT

- Combinations of resistors in series and parallel circuits:
  Series circuits
  $I = I_1 = I_2 = \ldots$
  $V = V_1 + V_2 + \ldots$
  $R_s = R_1 + R_2 + \ldots$
  Parallel circuits
  $I = I_1 + I_2 + \ldots$
  $V = V_1 = V_2 = \ldots$
  $\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots$

Components can be connected in series, in parallel, or as a combination of the two. We will use resistors to illustrate the possibilities.

Figure B5.27 shows three different resistors in a **series connection**. All the current follows the same path. Because of the law of conservation of charge, the charge per second (current) flowing into each resistor must be the same as the current flowing out of it and into the next resistor.

◆ **Series connection**
Two or more electrical components connected such that there is only one path for the electrical current, which is the same through all the components.

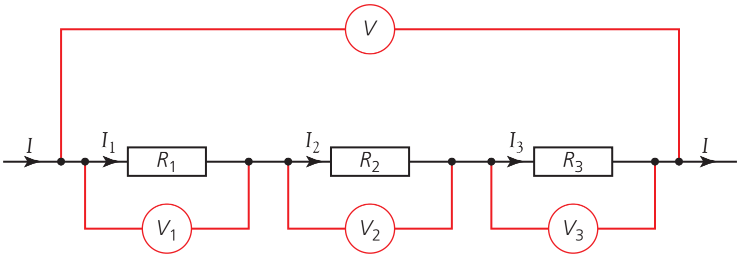


Figure B5.27 Three resistors in series

Currents in series: $\boldsymbol{I = I_1 = I_2 = \ldots}$

The sum of the separate potential differences must equal the potential difference across them all, V, so that:

Potential differences in series: $\boldsymbol{V = V_1 + V_2 + \ldots}$

Using $V = IR$ for the individual resistors, we get $IR_s = IR_1 + IR_2 + IR_3$, so that we can derive an equation for the single resistor, R_s, which has the same resistance as the combination.

Total resistance of **resistors in series**: $\boldsymbol{R_s = R_1 + R_2 + \ldots}$

Figure B5.28 shows three resistors connected in a **parallel connection**. The current splits into three and they follow different paths between the same two points. Because the resistors are all connected between the same two points, they must all have the same potential difference, V, across them.

◆ **Resistors in series** Resistors connected one after another so that the same current passes through them all. $R_s = R_1 + R_2 + \ldots$

◆ **Parallel connection** Two or more electrical components connected between the same two points, so that they have the same potential difference across them.

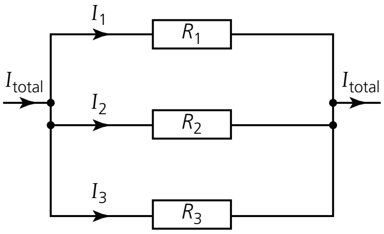


Figure B5.28 Three resistors in parallel

Potential differences in parallel: $\boldsymbol{V = V_1 = V_2 \ldots}$

The law of conservation of charge means that:

Currents in parallel: $\boldsymbol{I = I_1 + I_2 + \ldots}$

Applying $I = \dfrac{V}{R}$ throughout gives:

$$\frac{V}{R_p} = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3}$$

Cancelling the V gives us an equation for the single resistor, R_p, which has the same resistance as the combination:

Total resistance of resistors in parallel can be determined from: $\boldsymbol{\dfrac{1}{R_p} = \dfrac{1}{R_1} + \dfrac{1}{R_2} + \ldots}$

◆ **Resistors in parallel** Resistors connected between the same two points so that they all have the same potential difference across them. $\dfrac{1}{R_p} = \dfrac{1}{R_1} + \dfrac{1}{R_2} + \ldots$

All the electrical equipment in our homes is wired in parallel because, in that way, each device is connected to the full supply voltage and can be controlled with a separate switch.

Tool 3: Mathematics

Understand the significance of uncertainties in raw and processed data

How many significant figures are there in 5000 and 8000 (as discussed in Worked example B5.4)?

This example highlights a common problem. Without knowing the context in which this data is presented, we cannot be sure how many significant figures they have. If these are mathematical quantities, then their value is precisely defined and all figures are significant. Equally, if these are measurements of some kind, we would need to know the uncertainty in the measurement to know which were the significant figures. It would be better if the question presented the data in scientific notation!

For numerical data provided in this book, as a rule and for simplicity, we will generally assume that all digits are significant.

WORKED EXAMPLE B5.4

A 5000 Ω resistor and 8000 Ω resistor are connected in series.

a Calculate their combined resistance.

b What is the current through each of them if they are connected to a 4.5 V battery?

c What is the potential difference across the 5000 Ω resistor?

d Repeat these three calculations for the same resistors in parallel with each other.

Answer

a $5000 + 8000 = 13\,000\,\Omega$

b $I = \frac{V}{R} = \frac{4.5}{13\,000} = 3.5 \times 10^{-4}$ A through both resistors (seen on calculator as $3.46... \times 10^{-4}$)

c $V = IR = (3.46 \times 10^{-4}) \times 5000 = 1.7\,\text{V}$

d $\frac{1}{R} = \frac{1}{5000} + \frac{1}{8000} = \frac{13}{40\,000}$

$R = \frac{40\,000}{13} = 3100\,\Omega$

Both resistors have a p.d. of 4.5 V across them.

Current through 5000 Ω:

$I = \frac{V}{R} = \frac{4.5}{5000} = 9.0 \times 10^{-4}\,\text{A}$

Current through 8000 Ω:

$I = \frac{V}{R} = \frac{4.5}{8000} = 5.6 \times 10^{-4}\,\text{A}$

WORKED EXAMPLE B5.5

The lamps shown in Figure B5.29 are all the same.

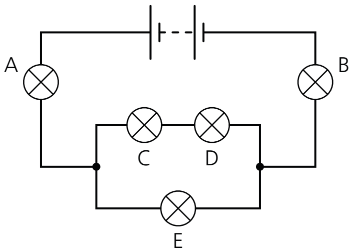


Figure B5.29 Five lamps in circuit

a Compare the brightness of all the lamps (assuming that they are all alight).

b If all the lamps have the same constant resistance of 2.0 Ω, what is the total resistance of the circuit?

Answer

a Lamps A and B will have the same brightness because the same current flows through them both. That same current will be split between lamp E and lamps C and D, so that these three must all be dimmer than lamps A and B.

Lamps C and D will have the same brightness because they are in series with each other.

Lamp E will be brighter than lamps C or D because a higher current will flow through it.

b C and D together will have a resistance of 2.0 + 2.0 = 4.0 Ω.

E in parallel with C / D will have a combined resistance of 1.3 Ω ($1/R = 1/2 + 1/4$).

Total resistance = 2.0 + 2.0 + 1.3 = 5.3 Ω

23 Draw a circuit diagram to represent the following arrangement: two lamps, A and B, are connected to a 12 V battery with a switch such that it can control lamp A only (lamp B is always on). An ammeter is connected so that it can measure the total current in both lamps and a voltmeter measures the p.d. across the battery.

24 Calculate the four possible total resistances that can be made by combining three 10 Ω resistors.

25 Figure B5.30 shows a simple circuit in which the ammeter and voltmeter have been connected in the wrong positions.

- **a** Predict the readings that you would expect to see on the meters. Explain your answer.
- **b** When the positions of the meters were swapped to their correct positions, what readings would you expect to see on the meters?

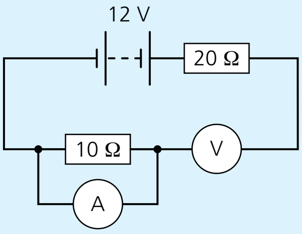


Figure B5.30 Simple circuit with ammeter and voltmeter in wrong positions

26 a Calculate the current that flows through a 12.0 Ω resistor connected to a p.d. of 9.10 V.

- **b** An ideal ammeter would display this current accurately, but what value will an ammeter of resistance 0.31 Ω record?
- **c** Determine the percentage error when using the ammeter in this way.

27 a Determine the currents I_1, I_2 and I_3 in Figure B5.31.

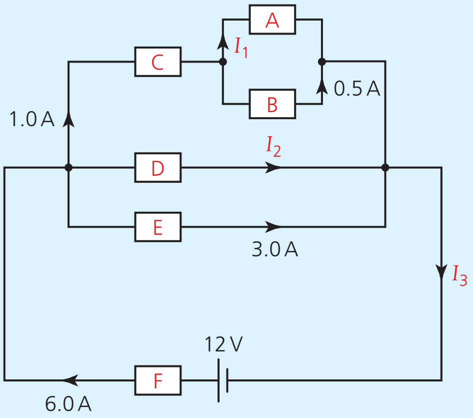


Figure B5.31 Circuit diagram

- **b** If resistor C has a value of 5.0 Ω and resistor E has a value of 2.0 Ω, determine values for the four unknown resistances.

28 Consider the circuit shown in Figure B5.32.

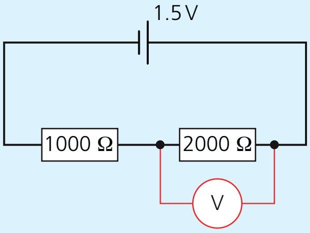


Figure B5.32

- **a** Calculate the current in the circuit before the voltmeter is connected.
- **b** What is the voltage across the 2000 Ω resistor (before the voltmeter is connected)?
- **c** Determine the voltages that will be measured if voltmeters with the following resistances are connected in turn across the 2000 Ω resistor:
  - **i** 5000 Ω
  - **ii** 50 000 Ω.

Emf and internal resistance

SYLLABUS CONTENT

- Cells provide a source of emf.
- Chemical cells and solar cells as the energy sources in circuits.
- Electric cells are characterized by their emf, ε, and internal resistance, r, as given by: $\varepsilon = I(R + r)$.

Sources of electrical energy

As individuals, if we wish to use electrical and electronic devices, we need to transfer energy to them using electrical currents. There are several possibilities, including:

- Most homes (but not all) are provided with a p.d. from the '**mains electricity**' generated at electric power stations by various means (most commonly from fossil fuels).
- A home-based **electrical generator** can be used to generate a p.d. from burning a fuel.
- Batteries (also called chemical cells or electric cells) use chemical reactions to provide a p.d. They can be single-use or rechargeable.
- **Solar cells** (also called **photovoltaic cells**) use radiant energy from the Sun to produce a p.d. (see Figure B5.33).
- **Wind generators** use the kinetic energy of moving air to produce a p.d.
- A **dynamo** on a bicycle (for example) can transfer kinetic energy to electrical energy for the lamp.

All of these energy sources have their advantages and disadvantages. These may be assessed by considering:

- convenience of use
- power available
- potential difference available
- whether they contribute to pollution and/or global warming
- whether the energy source is renewable
- whether the power supply is continuous
- whether they supply ac or dc (and the implications of that)
- whether the source is mobile, or fixed to a particular location
- internal resistance of supply (see below)
- cost.

◆ **Mains electricity** Electrical energy supplied to homes and businesses by cables from power stations. Also called *utility power.*

◆ **Generator (electrical)** Device that converts kinetic energy into electrical energy.

◆ **Solar cell** Device which converts light and infrared directly into electrical energy. Also called **photovoltaic cell**. A collection of solar cells connected together electrically is commonly called a solar panel.

◆ **Wind generator:** Device that transfers the kinetic energy of wind into electrical energy.

◆ **Dynamo:** A type of electricity generator that produces direct current.

LINKING QUESTION

- What are the advantages of cells as a source of electrical energy?

This question links to understandings in Topic A.3.

ATL B5B: Thinking skills

Asking questions based upon sensible scientific rationale

Imagine that your family have bought a remote house in the countryside for a holiday home, but it has no mains electricity supply. The electricity company can provide a new cable to the house, but the cost would be high. What information would you need to consider in order to decide how to provide energy to the home? You may prefer a renewable energy source, but are they always the best choice?

Emf

The **electromotive force (emf)** of a battery, or any other source of electrical energy, is defined as the total energy transferred in the source per unit charge passing through it.

Electromotive force is given the symbol ε and its unit is the volt, V (J C^{-1}). The name electromotive force can cause confusion because it is not a force. For this reason, it is commonly just called 'emf'.

For example, a battery with an emf of 12 V can transfer a total of 12 J to every coulomb of charge that flows through it. However, some of that energy will be transferred within the source itself, as explained in the next section.

■ **Figure B5.33** Solar cells collecting energy in the day to power lamps at night

◆ **Electromotive force (emf), ε** The total energy transferred in a source of electrical energy per unit charge passing through it.

◆ **Internal resistance, r** Sources of electrical energy, for example batteries, are not perfect conductors. The materials inside them have resistance in themselves, which we call internal resistance.

Internal resistance

Cells, batteries and other sources of electrical energy are not perfect conductors of electricity. The materials from which they are made all have resistance, called the **internal resistance** of the source, and given the symbol r.

Batteries have resistance, called *internal resistance*, typically less than 1 Ω.

If the internal resistance of a battery is much less than the resistance of the rest of the circuit, its effect can usually be ignored and, as a result, many examination questions refer to batteries or cells of 'negligible internal resistance'. But in other examples, in circuits with high currents and/or low resistance, the internal resistance of an energy source can have a significant effect on the circuit. Internal resistances can vary with temperature and the age of the battery, but in this course, we will assume that they are constant.

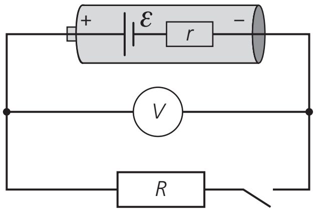


Figure B5.34 A cell in a simple circuit

Figure B5.34 shows a battery connected to an external resistance, R. In this example, the circuit symbols for the battery and the internal resistance are shown separately to aid understanding, but in practice they are combined and there is no accessible point between them. The voltmeter is assumed to be 'ideal'. To analyse the circuit, we will need to add the internal and external resistances together.

When the switch is open and there is no current flowing in the circuit ($I = 0$), there will be no voltage across the internal resistance (since $V_r = Ir$) and an ideal voltmeter will display the true value of the emf, ε, of the battery.

When the switch is closed and a current, I, flows, there will be a p.d. of $V_r = Ir$ across the internal resistance, so that the reading on the voltmeter will fall.

Consider a numerical example of Figure B5.34: The battery has an emf of 9.0 V and an internal resistance 0.4 Ω. When the switch is open, the voltmeter will read 9.0 V, but if a current of 2.5 A flows, $V_r = 2.5 \times 0.4 = 1.0$ V. This is commonly called '**lost volts**'. The voltmeter reading will fall from 9.0 V to (9.0 – 1.0) = 8.0 V.

◆ **Lost volts** Term sometimes used to describe the voltage drop (becoming less than the emf) that occurs when a source of electrical energy delivers a current to a circuit. Lost volts (Ir) increase with larger currents.

◆ **Terminal potential difference** The potential difference across the terminals of a battery (or other voltage supply) when it is supplying a current to a circuit (less than the emf).

It should be clear that an ideal voltmeter connected across a battery will only show the emf of the battery if there is no current flowing. At all other times the p.d. will be less, by an amount which depends on the magnitude of the current at that moment.

The p.d. across the terminals of a battery is equal to the p.d. applied to the circuit and is known as the **terminal p.d.**, V_t.

total energy transferred by the cell (per coulomb) = energy transferred to the circuit (per coulomb) + energy transferred inside the cell (per coulomb)

emf of cell = terminal p.d. across circuit + 'lost volts' due to internal resistance of battery

$$\varepsilon = V_t + V_r$$

The same current flows through both resistances, so that using $V = IR$ gives:

$$\varepsilon = IR + Ir$$

or:

$$\boldsymbol{\varepsilon = I(R + r)}$$

WORKED EXAMPLE B5.6

A battery with an emf of 1.5 V and internal resistance 0.82 Ω is connected in a circuit with a 5.6 Ω fixed resistor.

a Calculate the current in the circuit.

b If a (ideal) voltmeter is connected across the terminals of the battery when the current is flowing, what reading will it show?

Answer

a $$I = \frac{\varepsilon}{R + r} = \frac{1.5}{(5.6 + 0.80)} = 0.23\ \text{A}$$

b $V = IR = 0.23 \times 5.6 = 1.3\ \text{V}$

Choosing batteries

The emf and the internal resistance of any source of electrical energy are its defining features, although physical size and mass are also important, especially since they will affect the amount of energy that can be stored. 1.5 V is the most common emf produced by an electric cell. They are often connected in series to make a battery which has a greater overall emf, but this will also increase the overall internal resistance. 1.5 V cells connected in parallel will still produce an overall emf of 1.5 V, but the total internal resistance will be reduced. Table B5.4 shows the most common types of batteries. The use of mAh to represent energy storage is explained towards the end of this topic.

■ **Table B5.4** Typical properties of some common batteries

				Energy capacity	
Type of battery		**emf / V**	**Internal resistance/Ω**	**mAh**	**kJ (approx.)**
mobile phone		3.7	0.1	1400	20
AA or AAA		1.5	0.2	1000	5
button / coin battery		3.0	10	100	1
car battery		12	0.01	50 000	2000

Tool 3: Mathematics

Use of units whenever appropriate

The energy stored in a battery *of known voltage* is usually given in terms of the current it could supply for one hour at that voltage. Table B5.4 includes the example of a mobile phone battery, quoted at 1400 mAh. With a voltage of 3.7 V, it can supply a current of 1400 mA (1.4 A) for one hour.

Energy = *VIt* (as explained below in *electrical power section*) = 3.7 × 1.4 × 3600 = 1.9×10^4 J (≈ 20 kJ)

Figure B5.35 E-bike for hire

E-bikes have become very popular. A typical battery might have a mass of about 3 kg and supplies 50 Ah (9 MJ) with an emf of 50 V and internal resistance of 0.1 Ω. This should be enough to travel approximately 30 km on mostly level ground.

Inquiry 2: Collecting and processing data

Designing

A student set up the following circuit to investigate the emf and internal resistance of an AA battery.

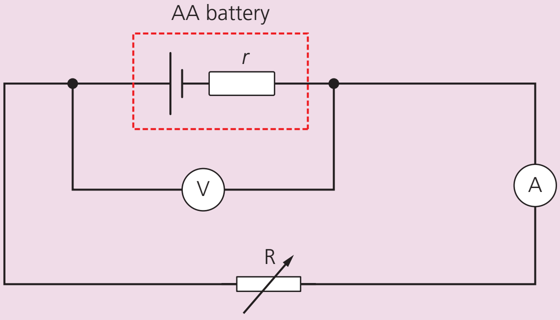


Figure B5.36 Investigating internal resistance

The following results (Table B5.5) were obtained when the value of the variable resistance was changed:

Table B5.5 Current and voltage measurements

I/A	*V*/V
1.0	1.0
0.6	1.2
0.4	1.3
0.2	1.4

The student concluded, by making calculations from these results, that the battery had an emf of 1.5 V and an internal resistance of 0.5 Ω. (Explain how she came to these conclusions.)

The student's teacher said that the conclusions were acceptable, but the experiment was not accurate enough. She suggested that there were ways in which the investigation could be improved, including the drawing of an appropriate graph.

Design an improved investigation using the same arrangement and explain how the data obtained can be used to accurately determine the emf and internal resistance of the battery.

29 Three cells, each of emf 1.5 V and internal resistance 0.20 Ω were connected together. Determine their combined emf and internal resistance if they were connected:

a in series

b in parallel

30 A very high-resistance voltmeter shows a voltage of 12.5 V when it is connected across the terminals of a battery that is not supplying a current to a circuit. When the battery is connected to a lamp, a current of 2.5 A flows and the reading on the voltmeter falls to 11.8 V.

a State the emf of the battery.

b Calculate the internal resistance of the battery.

c What is the resistance of the lamp?

31 When a battery of emf 4.5 V and internal resistance 1.1 Ω was connected to a resistor, the current was 0.68 A.

a What was the value of the resistor?

b If the resistor was replaced with another of twice the value, predict the new current.

c State an assumption that you made when answering part **b**.

32 If a connecting wire is connected by mistake across a battery or power supply, it is an example of a **short circuit**.

a Calculate the current that flows through a battery of emf 12.0 V and internal resistance 0.25 Ω if it is accidentally 'shorted'.

b Suggest what will happen to the battery.

33 A 1.50 V cell with an internal resistance of 0.100 Ω is connected to a resistance of 500 Ω. Outline why it would be reasonable to assume that the cell has 'negligible internal resistance'. Include a calculation.

◆ **Short circuit** An unwanted (usually) electrical connection that provides a low resistance path for an electric current. It can result in damage to the circuit, unless the circuit is protected by a fuse or circuit breaker.

ATL B5C: Research skills

Use search engines and libraries effectively

Use the internet to research into the latest developments into the design of batteries for electric vehicles, including their weight, charging possibilities and the range of the cars on a fully charged battery. Consider how you will verify the reliability of your sources.

Nature of science: Global impact of science

Scientific responsibility

Battery storage is seen as useful to society despite the well-known environmental issues surrounding their manufacture and disposal. Should scientists be held morally responsible for the long-term consequences of their inventions and discoveries?

Most, if not all, scientific and technological developments have some unwanted, and/or unexpected, side-effects. Most commonly these may involve pollution, the threat of the misuse of new technologies and the implications for an overcrowded world. Or maybe a new technology will result in dramatic changes to how societies function; changes that will have both benefits and disadvantages, many of which will be a matter of opinion.

Should more effort be made to anticipate the possible negative aspects of scientific research and development? Perhaps that is unrealistic, because predicting the future of anything, especially the consequences of as-yet unfinished research, is rarely successful. Of course, there are some extreme areas of research that most people would agree should never be allowed; nuclear or biological weapons, for example. It is important to appreciate that a key feature of much scientific research is that it involves the investigation of the unknown.

If 'society' decides that it wishes to control some area of scientific and technological research because the possible negative consequences are considered to be greater than the possible benefits, who makes those decisions and who monitors and controls the research (especially in this modern international world)? Is it reasonable to expect scientists to be responsible for their own discoveries and inventions? Or will human imagination, the motivation to explore the unknown and the desire (of some people) for fame, power or wealth, inevitably result in every possible new scientific and technological discovery being developed?

◆ **Rheostat** Variable resistance used to control current.

Using variable resistors

SYLLABUS CONTENT

- Resistors can have variable resistance.

Variable resistors come in a wide variety of shapes and sizes. Two examples can be seen in Figure B5.37. **a** is small in size but has a high resistance. Part **b** is larger in size because it is designed to carry much bigger currents, although it has a lower resistance. It is often called a **rheostat**.

Figure B5.37 Two different variable resistors

Many variable resistances have three terminals, one at each end of the resistance and a third movable / sliding contact between them. One possible use has already been seen in Figure B5.14. Only two of the three terminals are being used in that experiment. If the magnitude of the variable resistance is increased, the overall resistance of the circuit increases and the current decreases. At the same time, the p.d. across the variable resistance rises, while the p.d. across the component decreases.

WORKED EXAMPLE B5.7

Consider Figure B5.14. Suppose that the battery supplies a constant 12 V (and has negligible internal resistance), the component has a fixed resistance of 24 Ω and the variable resistor can vary from 0 to 48 Ω.

a Calculate the current in the circuit when the variable resistance is set to:

i 0 **ii** 24 Ω **iii** 48 Ω.

b Determine the p.d. across both components with the same three settings.

Answer

a **i** $I = \frac{V}{R} = \frac{12}{(24 + 0)} = 0.50\,\text{A}$

ii $\frac{12}{(24 + 24)} = 0.25\,\text{A}$

ii $\frac{12}{(24 + 48)} = 0.17\,\text{A}$

b p.d. across variable resistor = $IR = 0.50 \times 0 = 0\,\text{V}$. p.d. across component = 12 V

p.d. across variable resistor = $IR = 0.25 \times 24 = 6.0\,\text{V}$. p.d. across component = 6.0 V

p.d. across variable resistor = $IR = 0.167 \times 48 = 8.0\,\text{V}$. p.d. across component = 4.0 V

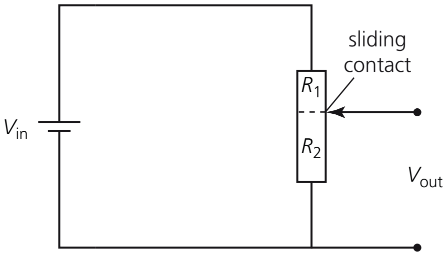
Note that it is not possible to reduce the p.d. across the component any lower than 4.0 V in this arrangement. However, using the same apparatus connected in a different way, it is possible to vary the p.d. across the component from 0 to 12 V. See next section.

Potentiometers

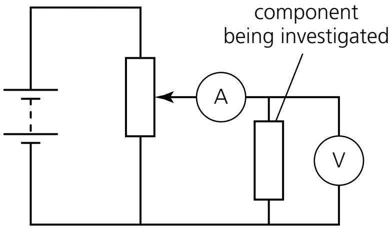
A potentiometer is the name we give to a three terminal variable resistor when its sliding contact is being used to produce a varying p.d.

Used as shown in Figure B5.38, the variable resistor (potentiometer) can provide a p.d., V_{out}, to another part of the circuit, which varies continuously from zero to the full p.d. of the battery, V_{in}. The maximum voltage will be obtained with the sliding contact at the top of the variable resistor (as shown) and the voltage will be zero with the contact at the bottom (when both connections to the other circuit come from the same point).

A potentiometer provides the best way of varying the p.d. across a component in order to investigate its I–V characteristics (see Figure B5.39).



■ **Figure B5.38** A variable resistor used as a potentiometer



■ **Figure B5.39** A circuit for investigating I–V characteristics of electrical components

When a potentiometer is connected as the input into a circuit, the value of V_{out} cannot be confirmed without considering the effect of the resistance of the rest of that circuit. Generally, the resistance of the circuit should be much higher than the resistance of the potentiometer, unlike in Worked example B5.8.

WORKED EXAMPLE B5.8

Consider Figure B5.39, using components of the same value as in the last worked example: supply voltage is a constant 12 V, the component being investigated has a resistance of 24 Ω and the variable resistor (potentiometer) can vary from 0 to 48 Ω.

Determine the p.d. across the component when the sliding contact on the potentiometer is:

- **a** at the top
- **b** at the bottom
- **c** at the mid-point.

Answer

- **a** Connected across the full supply p.d.: 12 V
- **b** Not connected across the supply p.d.: 0 V
- **c** Half of the potentiometer (24 Ω) is in parallel with the 24 Ω component, so together they have a resistance of 12 Ω. This means that the other half of the potentiometer (24 Ω) is in series with 12 Ω.

  There will be (24/36) × 12 = 8 V across the top half of the potentiometer and only 4 V across the bottom half, and the component.

Potential-dividing circuits

Figure B5.40 shows an LDR and a thermistor, which are both semi-conducting components. These variable resistors are used as electrical sensors in circuits which control lighting and heating.

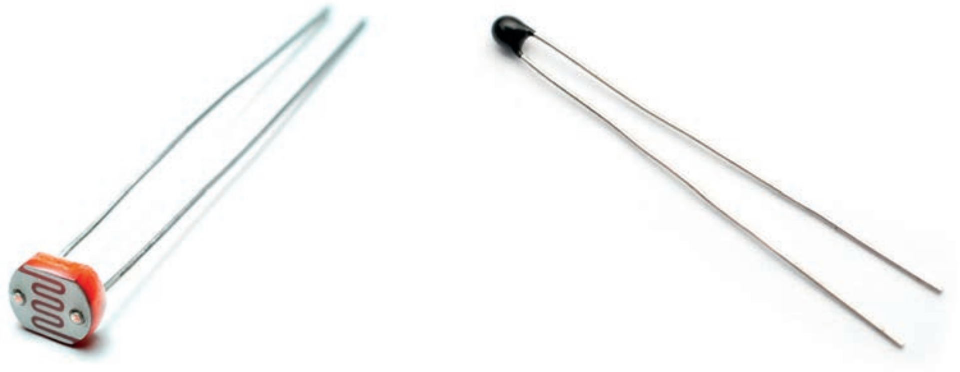


Figure B5.40 LDR (left) and thermistor (right) for sensing changes in light intensity and temperature

◆ **Potential-dividing circuit** Two resistors used in series with a constant potential difference across them. When one resistance is changed, the potential difference across each resistor will change, and this can be used for controlling another part of the circuit.

Figure B5.41 shows how the resistance of an LDR changes with light intensity. (Note that the resistance scale is not linear, it is logarithmic.) In bright sunlight the LDR has a resistance of about $10^2\,\Omega$, but in the dark its resistance rises to about $10^6\,\Omega$ (10 000 × greater). More light energy releases more charge carriers in the LDR.

Sensors are connected in **potential-dividing circuits** as shown (using an LDR) in Figure B5.42.

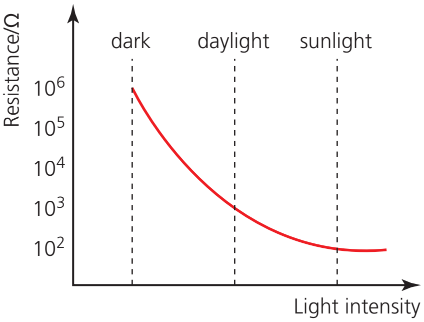


Figure B5.41 Variation of resistance of an LDR with light intensity

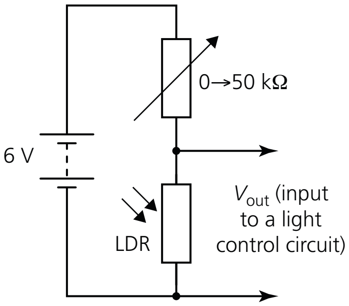


Figure B5.42 An LDR in a potential-dividing circuit

A potential-dividing circuit produces an output voltage which is a fraction of the supply voltage, dependent on the ratio of the values of the resistors it contains.

In this circuit, the p.d. of the supply, V_s, (6.0 V) is always shared / divided between the variable resistance and the LDR in a ratio depending on their resistances, as shown in the following worked example. The p.d. across one of the resistors (in this example, the LDR) is used as the input to another circuit.

Tool 2: Technology

Use sensors

An LDR in a potential-dividing circuit can produce outputs that can be used to roughly *compare* light intensities at different times or locations. However, if more accuracy is required, the output would need to be calibrated to measure the intensity of the light falling on it. This would require comparison with another, reliable light meter, such as shown in Figure B5.43, which usually displays light intensity in the SI unit *lux* (not needed in this course).

■ **Figure B5.43** Light meter

WORKED EXAMPLE B5.9

Consider the circuit shown in Figure B5.42. The light comes on if V_{out} is greater than 5.0 V If the variable resistance, R_v, is set to 50 kΩ, calculate a value for V_{out}, – which can be the input to the light control circuit – when the room is

a in bright sunlight

b in a dark room. Take values from Figure B5.41.

Answer

a Assuming that the current in both resistors is the same:

$$I = \frac{V_s}{R_{total}} = \frac{V_{LDR}}{R_{LDR}} \left(= \frac{V_v}{R_v} \right)$$

So that:

$$V_{LDR} = V_s \times \left(\frac{R_{LDR}}{R_{total}} \right) = 6.0 \times \frac{100}{100 + (50 \times 10^3)} = 0.12\,\text{V}$$

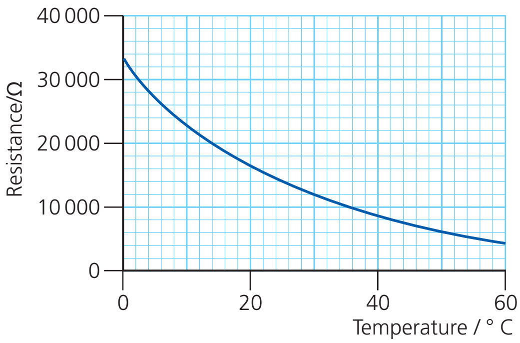
This p.d. will not be enough to turn the lighting on, as required in a bright environment.

b $$V_{LDR} = V_s \times \left(\frac{R_{LDR}}{R_{total}} \right) = 6.0 \times \frac{1.0 \times 10^6}{(1.0 \times 10^6) + (50 \times 10^3)} = 5.7\,\text{V}$$

This p.d. will turn the lighting on automatically, as required in a dark environment.

The light intensity at which the lighting is turned on, or off, can be adjusted by the choice of value for the variable resistor. In practice the input resistance of the lighting circuit will also have to be considered.

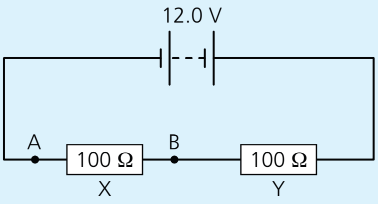
◆ **Thermostat** Component that is used with a heater or cooler to maintain a constant temperature.



■ **Figure B5.44** Variation of resistance with temperature for a thermistor

Figure B5.44 shows how the resistance of a semi-conducting thermistor changes with temperature. Thermal energy releases more charge carriers in the thermistor, so that its resistance decreases as it gets hotter. This can be used as a sensor (called a **thermostat**) to control temperature. (There is another kind of thermistor which has *greater* resistance if its temperature rises.)

34 a Calculate the potential difference between points A and B in Figure B5.45.



■ **Figure B5.45**

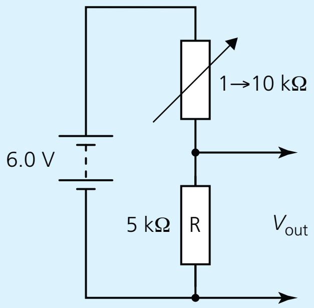
b A student wants to connect a lamp which is rated at 2.5 A, 6.0 V. Calculate the working resistance of the lamp.

c The student thinks that the lamp will work normally if he connects it between A and B, in parallel with X. Discuss if the lamp will work as he hopes.

d Another student thinks that the lamp will work if resistor X is removed (while the lamp is still connected between A and B). Calculate the new voltage across the lamp. Will the lamp work normally now?

e Suggest how the lamp could work normally using a 12 V battery.

35 The value of the variable resistor in Figure B5.46 can be changed continuously from 1 kΩ to 10 kΩ.



■ **Figure B5.46**

a Calculate the maximum and minimum potential differences, V_{out}, that can be obtained across R.

b State (without calculations) how your answer will change if V_{out} is connected across another 5 kΩ.

c Estimate the percentage change in the resistance of the thermistor represented in Figure B5.44, as the temperature changes from 0 °C to 60 °C.

36 a Draw a potential-dividing circuit that could be used to control the temperature of, for example, a refrigerator.

b Make a list of household electrical devices that have thermostats inside them.

37 Describe a laboratory experiment that could be used to obtain results similar to those seen in Figure B5.44. Include a fully annotated diagram.

◆ **Power (electrical)** The rate of dissipation of energy in a resistance.

Electrical power

SYLLABUS CONTENT

▶ Electrical power, P, dissipated by a resistor given by: $P = IV \equiv I^2R = \frac{V^2}{R}$.

If the current through a resistor is, for example, 3 A, then 3 C of charge is passing through it every second. If there is a potential difference across the resistor of 6 V, then 6 J of energy is being transferred by every coulomb of charge (to internal energy). The rate of transfer of energy is 3 × 6 = 18 joules every second (watts).

More generally, we can derive an expression for the **electrical power** dissipated to internal energy in a resistor by considering the definitions of p.d. and current, as follows:

$$\frac{\text{energy transferred}}{\text{time}} = \frac{\text{charge flowing through resistor}}{\text{time}} \times \frac{\text{energy transferred in resistor}}{\text{charge flowing through resistor}}$$

$$\frac{W}{\Delta t} = \frac{\Delta q}{\Delta t} \times \frac{W}{\Delta q}$$

or:

power = current × potential difference $\boldsymbol{P = IV}$

Because $V = IR$, this can be rewritten in two other useful ways:

$$\boldsymbol{P = I^2R = \frac{V^2}{R}}$$

To calculate the total energy transferred in a given time, we know that energy = power × time, so that:

electrical energy = VIt

The heating effect of a current passing through a resistor

Whenever any current passes through any resistance, energy will be transferred to internal energy and then transferred as thermal energy.

This has always been one of the most widespread applications of electricity, including heating water and heating air.

WORKED EXAMPLE B5.10

An electric iron is labelled as 230 V, 1100 W.

- **a** Explain exactly what the label means.
- **b** Calculate the resistance of the heating coil of the iron.
- **c** Explain with a calculation what would happen if the iron was used in a country where the mains voltage was 110 V.

Answer

a The label means that the iron is designed to be used with 230 V and, when correctly connected, it will transfer energy at a rate of 1100 joules every second.

b $P = \frac{V^2}{R}$

$1100 = \frac{230^2}{R}$

$R = 48.1\,\Omega$

c $P = \frac{V^2}{R} = \frac{110^2}{48.1} = 251\text{ W}$

The iron would transfer energy at 0.25 times the intended rate and would not get hot enough to work properly. If an iron designed to work with 110 V was connected to 230 V it would begin to transfer energy at about four times the rate it was designed for; it would overheat and be permanently damaged.

Recall from Topic A.3 that efficiency, η, in terms of energy transfer or power is given by:

$$\eta = \frac{\text{useful work}}{\text{input energy}}$$

So:

$$\eta = \frac{E_{\text{output}}}{E_{\text{input}}} = \frac{P_{\text{output}}}{P_{\text{input}}}$$

Tool 3: Mathematics

Use of units whenever appropriate

When we buy a battery or pay for the mains electricity connected to our homes, we are really buying the energy that is transferred by the electric current. In most countries mains electrical energy is sold by the **kilowatt hour**:

1 kWh is the amount of energy transferred by a 1 kW device in one hour, which is the equivalent of 1000 J per second for 3600 s, or **3.6 × 10^6 J (3.6 MJ)**.

◆ **Kilowatt hour, kWh**
The amount of electrical energy transferred by a 1 kW device in 1 hour.

LINKING QUESTION

- How can the heating of an electrical resistor be explained using other areas of physics?

This question links to understandings in Topic B.1.

WORKED EXAMPLE B5.11

In a city where the cost of electrical energy is $ 0.14 for each kWh, predict how much it will cost to operate an air conditioner with an average power of 1500 W for four hours a day for a week.

Answer

Total cost = energy supplied in kWh × 0.14

= 1.5 × 4 × 7 × 0.14 = $ 5.90

38 A 12 V potential difference is applied across a 240 Ω resistor.

a Calculate:

i the current

ii the power

iii the total energy transferred in 2 minutes.

b What value resistor would have twice the power with the same voltage?

c What p.d. will double the power with the original resistor?

39 A 2.00 kW household water heater has a resistance of 24.3 Ω.

a Calculate the current that flows through it.

b What is the mains voltage?

c Show that the overall efficiency of the heater is approximately 85% if 1.0 × 10^5 J are transferred to the water every minute.

40 a Determine the rate of production of thermal energy if a current of 100 A flowed through an overhead cable of length 20 km and resistance of 0.001 ohm per metre.

b Comment on your answer.

41 a Calculate the power of a heater that will raise the temperature of a metal block of mass 2.3 kg from 23 °C to 47 °C in 4 minutes (specific heat capacity = 670 J kg^{-1} $°C^{-1}$).

b Draw a circuit diagram to show how the heater should be connected to a 12 V supply and suitable electrical meters so that the power can be checked.

42 a An electric motor is used to raise a 50 kg mass to a height of 2.5 m in 24 s. The voltage supplied to the motor was 230 V but it was only 8.0% efficient. Determine the current in the motor.

b State two reasons why this process has a low efficiency.

43 a Calculate the value of resistance that would be needed to make a 1.25 kW water heater in a country where the mains voltage is 110 V.

b What current flows through the heater during normal use?

44 If a kWh of electrical energy costs 6.2 rupees, predict how long a 150 W television can be on for a total cost of 100 rupees?

45 In 2022, the best mobile phone batteries were rated at 3.7 V, 5000 mAh.

a Calculate how much energy (J) they store.

b If such a battery is used with an average current of 100 mA, predict how many hours before the battery would be completely discharged.

c Estimate how long completely recharging the battery will take at an average rate of 5 W.

d Suggest why phone manufacturers do not install batteries which store more energy.

ATL B5D: Thinking skills

Applying key ideas and facts in new contexts

Solar panels connected to an outdoor night light are widely available for sale on the internet. Some advertisements make claims which are vague or unrealistic, or only apply to ideal conditions. Suppose you wanted to buy a solar panel with a 25 W LED spotlight which was on for 12 hours every night. Research and compare the products available. Which would you choose, and why?

ATL B5E: Self-management skills

Setting learning goals and adjusting them in response to experience

Reaching the end of Theme B means that you will soon be halfway through the content of the IB Physics syllabus (if you followed the order of this book). Ask yourself some questions and give honest replies.

- Are you doing as well as you hoped at the beginning of the course?
- Could you realistically be doing much better?
- Do you spend enough time studying physics to achieve your goals?
- Are your study methods effective, or are you too easily distracted?
- Are you finding the material interesting? If not, why not?
- Is any of the course content difficult to understand? If so, why?
- Are you using the school's resources effectively?
- Do you use the help of your fellow students and teachers?
- Do you care enough about your physics grade to want to work harder?

Honest answers to these, or similar questions, should lead to setting achievable goals for the rest of the course.

C.1 Simple harmonic motion (SHM)

Guiding questions

- What makes the harmonic oscillator model applicable to a wide range of physical phenomena?
- Why must the defining equation of simple harmonic motion take the form it does?
- How can the energy and motion of an oscillation be analysed both graphically and algebraically?

Oscillations

◆ **Oscillation** Repetitive motion about a fixed point.

An **oscillation** is a regularly repeated backwards-and-forwards movement about the same central point, and along the same path.

Figure C1.1 Oscillations of a humming bird's wings

The importance of the study of oscillations should be apparent from the very wide range of examples, both in physics and more generally. A few scientific examples:

- oscillations of the planets around the sun
- oscillations of the human heart
- oscillations of clocks (mechanical and electronic)
- oscillations of engines and motors
- oscillations of atoms within molecules
- oscillations within musical instruments
- oscillations producing human speech and within the eardrum
- oscillations of waves on water
- oscillations of light and other electromagnetic waves
- oscillations of tides on the ocean
- oscillations of electric currents.

◆ **Simple harmonic motion (SHM)** An idealized oscillation that maintains a constant amplitude and frequency.

◆ **Isochronous** Describing events that take equal times.

Simple harmonic motion (SHM) is a simplified theoretical model representing oscillations. It is the starting point for the study of all oscillations.

In perfect SHM the oscillations always take the same time and there are no resistive forces, so that they continue oscillating indefinitely with no loss of energy and constant amplitude.

This time-keeping property is described as being **isochronous**.

Nature of science: Models

A 'model' in science means a simplified representation of a more complex situation. A model may take many forms, for example: a description in words, a drawing, a theory, an equation, a 3-D construction, a computer program or simulation, and so on.

SHM is a simple model (of oscillations) in a complex world. As the list above suggests, we are surrounded by oscillations, but few, if any, are perfect simple harmonic oscillators. Real oscillators are complex and various. To understand them, we need to first make sure we understand simplified versions.

Such simplifying models are found throughout physics and they are powerful and very useful. They should not be dismissed because of their basic assumptions.

Terms used to describe oscillations and SHM

SYLLABUS CONTENT

- A particle undergoing simple harmonic motion can be described using time period, T, frequency, f, angular frequency, ω, amplitude, equilibrium position, and displacement.
- The time period in terms of frequency of oscillation and angular frequency as given by:

$$T = \frac{1}{f} = \frac{2\pi}{\omega}$$

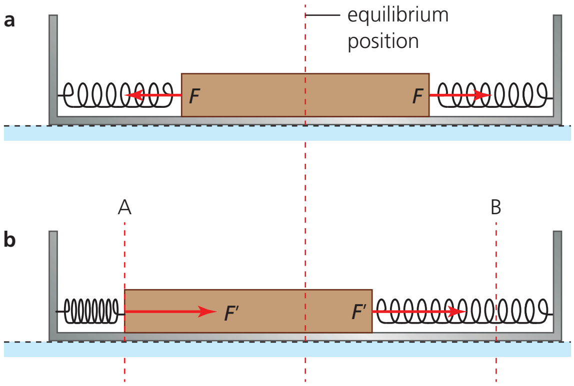
Although we commonly talk about oscillating masses or oscillating objects, a discussion of perfect SHM often refers to a point mass, or a 'particle'.

Until a mass is displaced by a resultant force, it will remain in its **equilibrium position**, where the resultant force is zero.

◆ Equilibrium position
Position in which there is no resultant force acting on an object.

If a mass is then displaced, oscillations may occur if there is always a resultant *restoring force* pulling, and/or pushing, it back towards its equilibrium position. The mass will gain kinetic energy as it moves back where it came from. It will then pass through the equilibrium position (because of its inertia), so that the displacement is then in the opposite direction. Kinetic energy is transferred to some form of potential energy in the system, and when the kinetic energy has reduced to zero, the mass will stop and then reverse its motion. And so on. The motion of a mass between two identical springs on a friction-free surface is a good visualization of this. See Figure C1.2.

At all times the springs are stretched, but in part **a** of the diagram, the forces from the springs on the mass are equal and opposite. In part **b**, there is a resultant force to the right on the mass from the springs. When released, the mass will accelerate, reach its greatest speed in the centre, and then decelerate until it stops at B. And so on.



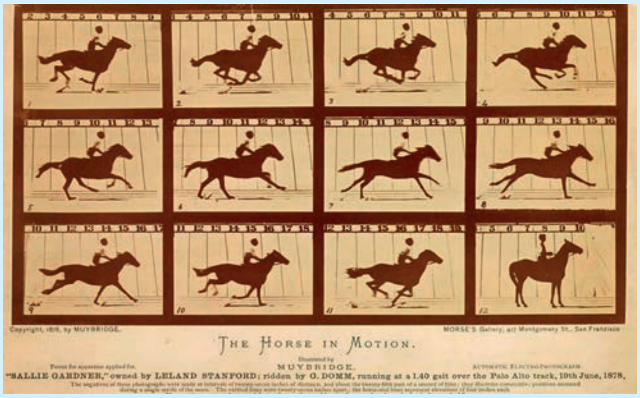
■ **Figure C1.2** Oscillations of a mass between two springs

Tool 2: Technology

Carry out image analysis and video analysis of motion

Making observations and measurements on fast-moving objects provides many technical difficulties. In the 1870s and 1880s, Eadweard Muybridge was the first to try to analyse motion (most notably of horses) by using quickly taken photographs. See Figure C1.3.

The times between each photograph were known and the position of the horse could be judged from the lines in the background. From this information, the average speed of the horse between each picture could be calculated (and, if relevant, the horse's acceleration). The pictures also revealed previously unconfirmed information about how a horse's legs moved.



■ **Figure C1.3** Famous photographs of a horse in motion (Eadweard Muybridge)

The same principles apply to *video analysis* using a modern video camera, or smart phone app. The motion to be analysed should be just in front of a suitable measurement grid. The video can be replayed in slow motion, or frame-by-frame.

Alternatively, a software program can be used which enables the position of a videoed object to be tracked and measured.

Oscillations similar to that seen in Figure C1.2 can be analysed using a suitable video camera or smart phone. A scale calibrated in millimetres can be placed behind the oscillating mass and a few seconds of action recorded while the mass is oscillating. Replaying the motion in slow motion, or frame-by-frame, can provide information on times and displacements.

Common mistake

The amplitude of an oscillation is *not* the distance between its extreme positions (which equals twice the amplitude). A single movement between the extremes is *not* a complete oscillation (it is half an oscillation).

The displacement, x, of an oscillator is its distance from the equilibrium position in a specified direction. Displacement is a vector quantity. (This term should be familiar from Topic A.1.) The displacement varies continuously, both in size and direction, during an oscillation.

The **amplitude**, x_0, of an oscillation is its maximum displacement while oscillating.

In the idealized example (SHM), where there is no energy dissipation, the amplitude will remain constant. More realistically, each oscillation will have an amplitude which is less than the one before (assuming there is no driving force).

One *oscillation* is completed every time that an oscillating mass returns to a certain position, moving in the same direction.

A compete oscillation is sometimes called a **cycle**. An object which oscillates is called an **oscillator**.

The **time period, T,** of an oscillation is the time taken for one complete oscillation. Unit: s.

The **frequency, f,** of the motion is the number of oscillations in unit time (per second). Unit: hertz, Hz. A frequency of 1 Hz means that there is one oscillation per second.

$$\text{frequency, } T = \frac{1}{f}$$

- ◆ **Amplitude** Maximum displacement of an oscillation (or wave).
- ◆ **Cycle (oscillation)** One complete oscillation.
- ◆ **Oscillator** Something which oscillates.
- ◆ **Time period, T** Time taken for one complete oscillation.
- ◆ **Frequency, f** The number of oscillations per unit time, (usually per second). $f = \frac{1}{T}$ (SI unit: hertz, Hz).

The time period and frequency of an oscillation provide exactly the same information. Typically, we prefer to use the one which is greater than one.

The time period and frequency of most practical oscillations remain constant (they are *isochronous*), even when the amplitude reduces because of energy dissipation.

In physics we often deal with high frequencies, so the following units are in common use: kHz (10^3 Hz), MHz (10^6 Hz), GHz (10^9 Hz).

Connection between SHM and circular motion

Sometimes we may describe circular motion as an oscillation, and the last two terms described above (frequency and time period) should already be familiar because they were introduced in the circular motion sub-topic in A.2.

There is a close connection between oscillations and circular motion. Indeed, viewed from the side, motion in a circle has exactly the same pattern of movement as a simple oscillation.

Figure C1.4 shows a particle moving in a circle at constant speed. Point P is the projection of the particle's position onto the diameter of the circle.

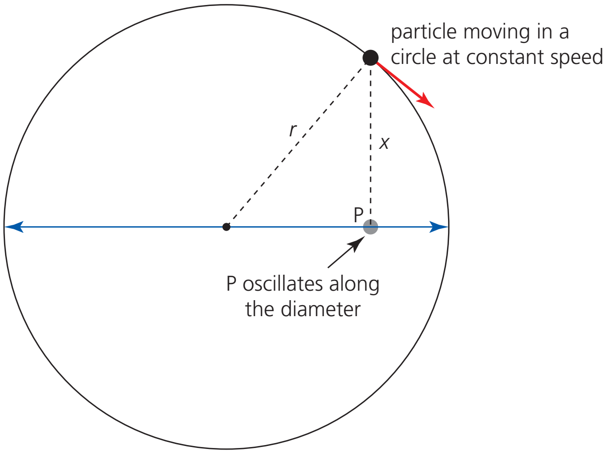


Figure C1.4 Comparing SHM to motion in a circle

As the particle moves in a circle, point P oscillates backwards and forwards along the diameter with the same frequency as the particle's circular motion, and with an amplitude equal to the radius, r, of the circle.

One complete oscillation of point P can be considered as equivalent to the particle moving through an angle of 2π radians.

The concept of angular velocity, ω, was introduced in Topic A.2 as a key quantity to describe the physical reality of motion in a circle. Because of the close **analogy** between circular motion and SHM (described above), ω is also important in the mathematical description of SHM. However, in the study of SHM, ω is known as **angular frequency**. Unit: rad s^{-1}

◆ **Analogy** Applying knowledge of one subject to another because of some similarities.

◆ **Angular frequency, ω** Similar to angular velocity, but used to represent the frequency of an oscillation in rad s^{-1} (because of the mathematical similarities between uniform circular motion and simple harmonic oscillations.) $\omega = 2\pi f$.

$$\text{angular frequency, } \boldsymbol{\omega} = \frac{\mathbf{2\pi}}{\boldsymbol{T}} = \mathbf{2\pi f}$$

Remember that, although the terms period, frequency and angular frequency are all used to describe oscillations, they are just different ways of representing exactly the same information.

TOK

Knowledge and the knower

- How do we acquire knowledge?
- How do our expectations and assumptions have an impact on how we perceive things?

Analogies, correlations and causal relationships

An *analogy* is made when we compare an understanding, process or phenomenon in one area of knowledge to another seemingly unrelated area, and we see similarities. This might enable us to understand some deeper, underlying process that causes both the phenomena – or maybe an analogy just makes it easier to understand what is going on.

If an analogy proves to be useful, but has little, or no other validity, does that justify its use, and does it increase our knowledge of the system to which it applied? What is the difference between a correlation, a causal link and an analogy?

Consider, for example, applying the mathematics of oscillations to variations in animal populations, or to economic cycles.

WORKED EXAMPLE C1.1

Consider the oscillator shown in Figure C1.2. The oscillating mass has a length of 12 cm and the distance between A and B is 18 cm.

a Calculate the amplitude of the oscillation.

b Determine the displacement of the oscillator when its end is in position B.

c State when the mass has its greatest kinetic energy.

d If the period of the oscillator was 1.5 s, calculate its

  i frequency

  ii angular frequency.

Answer

a $\frac{(18 - 12)}{2} = 3\text{ cm}$

b 3 cm to the right

c When it passes through its equilibrium position.

d **i** $f = \frac{1}{T} = \frac{1.0}{1.5} = 0.67\text{ Hz}$

  ii $\omega = 2\pi f = 2 \times \pi \times 0.67 = 4.2\text{ rad s}^{-1}$

1 The central processing unit of a lap-top computer operates at 3.2×10^9 cycles per second.
  a Express this frequency in
    i megahertz
    ii gigahertz.
  b Calculate the time period of each cycle.

2 When a guitar string was plucked (once) it oscillated with a frequency of 196 Hz.
  a Determine the *angular* frequency of this oscillation.
  b Suggest how you would expect the frequency and amplitude to change in the next few seconds. Explain your answer

3 a Show that the angular frequency of the Earth spinning on its axis is approximately $7 \times 10^{-5}\,\text{rad}\,\text{s}^{-1}$.
  b Determine the total angle (rad) through which it will rotate in 100 hours.

4 A car engine was measured to have 3755 rpm (revolutions per minute). Calculate its angular frequency.

5 Using slow-motion video replay, the angular frequency of the oscillations of a humming bird's wings was found to be $272\,\text{rad}\,\text{s}^{-1}$.
Determine how many times it beat its wings in one minute.

Two commonly investigated oscillators

SYLLABUS CONTENT

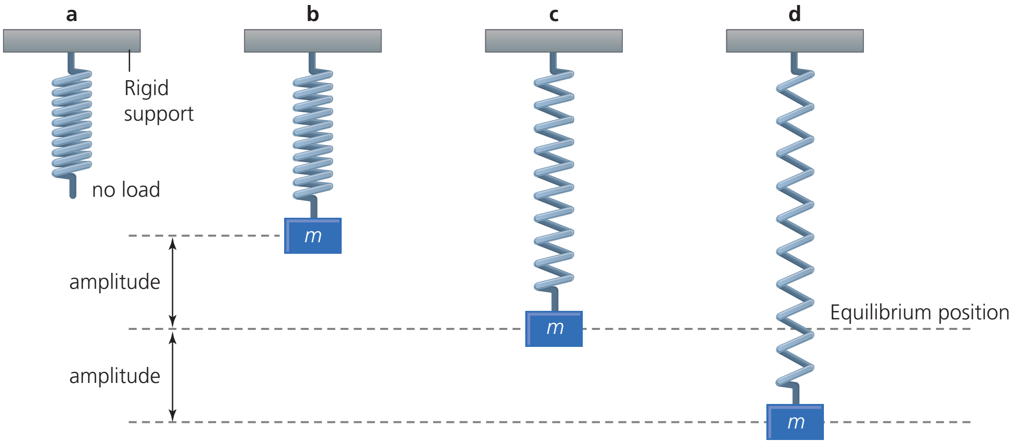
- The time period of a mass–spring system as given by:
$$T = 2\pi\sqrt{\frac{m}{k}}$$
- The time period of a simple pendulum as given by:
$$T = 2\pi\sqrt{\frac{l}{g}}$$

A mass–spring system and a simple pendulum are important for a number of good reasons. These include:

- The proportional relationships between force and displacement are easily understood.
- Their periods of oscillation have a convenient time for measurement in a school laboratory.
- They are good approximations to simple harmonic motion.
- Energy is dissipated slowly, so that the oscillations continue for a long enough time that allows for accurate measurements to be completed.
- They act as starting points for understanding many similar, but more complex, oscillators.

Mass-spring system

We have already briefly discussed a mass oscillating horizontally between two springs (Figure C1.2). A more common arrangement is shown in Figure C1.5. Of course, in this arrangement, the force of gravity (the weight of the mass) also acts vertically, but it does not affect the results because it is constant. The system would have the same time period if it was moved to a location where the gravitational field strength was different.

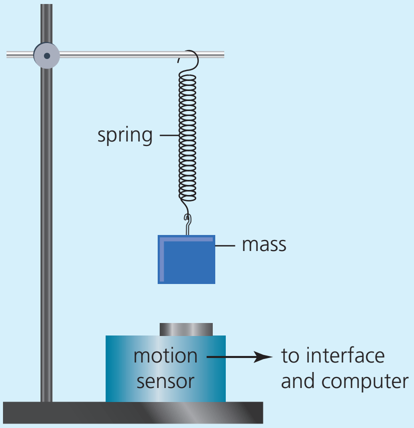


■ **Figure C1.5** Different positions of a mass, *m*, oscillating on a spring

Tool 2: Technology

Use sensors

Position sensors are useful in many aspects of the study of motion, including mechanical oscillations. Figure C1.6 shows how digital data can be collected, which will be processed later by the computer.



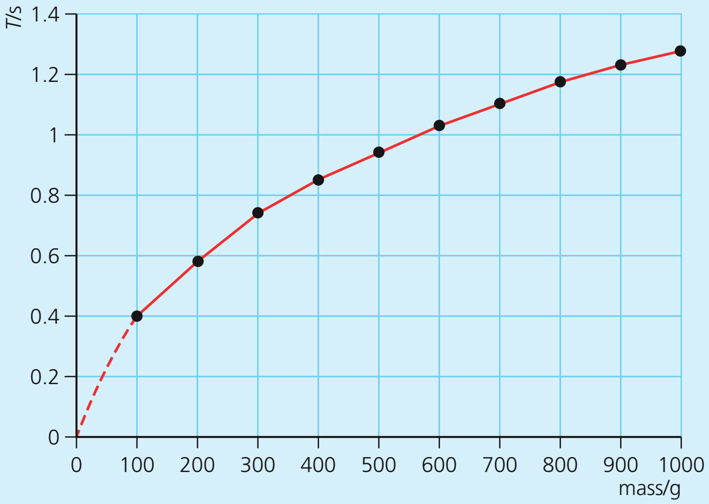
■ **Figure C1.6** Using a sensor to investigate oscillations of a mass on a spring.

We will assume that the spring is never overstretched and that it obeys Hooke's law (see Topic A.2):

$$F_H = -kx$$

Or, in words: the size of the restoring force acting on the mass = spring constant × displacement from its equilibrium position. The spring constant is a measure of the spring's stiffness (= force / deformation).

Laboratory investigations of the time periods of a mass on a spring are straightforward, especially those involving using different masses on the same spring. For typical results see Figure C1.6. Different springs, or combinations of springs, used with the same mass, can be used to investigate the effect of the spring constant, *k*, on the period.



■ **Figure C1.7** Variation of time period, *T*, with mass, *m*, on a spring

Tool 3: Mathematics

Linearize graphs

You may have learnt how to linearize graphs in Topic B.1. The graph seen in Figure C1.6 appears that it might represent the relationship $T^2 \propto m$. Check this by using information from the graph to draw a T^2–m graph (uncertainties in T are ± 0.05s). It should produce a straight line through the origin. Use the graph and the relationship shown below to determine a value for the spring constant, k.

The exact relationship for the SHM of a mass on a spring is as follows:

Time period of a mass-spring system:

$$T = 2\pi\sqrt{\frac{m}{k}}$$

Since $\omega = \frac{2\pi}{T}$, this equation can also be written as: $\omega^2 = \frac{k}{m}$.

This equation applies perfectly only to a point mass acted upon by a separate simple spring system that has a well-defined stiffness (k) and no (significant) mass. There is an extremely large number of other oscillating mechanical systems that have similarities to this simple model but are much more complex. For example, oscillations in buildings and bridges. Figure C1.8 shows a simpler example, in which the mass is spread uniformly along the oscillating system, a ruler.

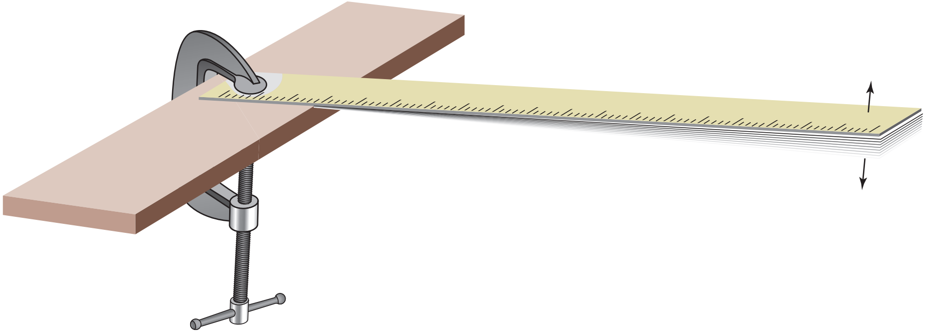


Figure C1.8 Oscillating ruler

Inquiry 3: Concluding and evaluating

Concluding

A student wants to investigate an oscillating ruler, similar to that seen in Figure C1.8, but she varied the mass of the ruler by taping various masses near to its free end. She predicted that the time period of the oscillator could be determined from the same equation as for a mass oscillating vertically on the end of a spring. Table C1.1 summarizes her raw data. Assume that the uncertainties are low.

Table C1.1 Results of vibrating ruler experiment

Mass on end of blade / g	Time period / s
0	too quick to measure
40	0.55
80	0.76
120	0.91
160	1.04
200	1.15

Process the results and reach a conclusion. State whether the student predicted correctly. If not, suggest a possible reason.

WORKED EXAMPLE C1.2

A 200 g mass was placed on the end of a long spring and increased its length by 5.4 cm.

a Determine the spring constant, k, of the spring.

b If the mass is displaced a small distance from its equilibrium position and undergoes SHM, calculate the frequency of oscillations.

c Suggest a possible reason why oscillations with greater amplitude may not be simple harmonic.

Answer

a $F_H = -kx$

$0.200 \times 9.8 = -k \times (5.4 \times 10^{-2})$

$k = -36\,\text{N m}^{-1}$ (the negative sign shows that force and displacement are in opposite directions)

b $T = 2\pi\sqrt{\frac{m}{k}} = 2 \times \pi \times \sqrt{\frac{0.200}{36}} = 0.47\,\text{s}$

$f = \frac{1}{T} = \frac{1}{0.47} = 2.1\,\text{Hz}$

c The spring may become overstretched, so that Hooke's law is no longer applicable.

Simple pendulum

◆ **Pendulum** A weight, which is suspended below a pivot, which is able to swing from side to side. The weight is sometimes called the **pendulum bob**. The concept of a **simple pendulum** is a point mass on the end of an inextensible string.

There are many different designs of **pendulum**, all of which involve a mass swinging from side to side under the effects of gravity. We use the term '**simple pendulum**' to describe the simplest possible model of a pendulum: a spherical mass swinging on a rod, or string (both of which have negligible mass), from a rigid, frictionless support. See Figure C1.9. The mass is often called a **pendulum 'bob'**.

Because the mass is spherical and the rod / string has negligible mass, all of the mass of the pendulum can be assumed to be acting as a point mass at the centre of the sphere. The displacement of the motion can be considered to be the angle between the rod / string and the vertical. The restoring force is provided by gravity: consider Figure C1.10.

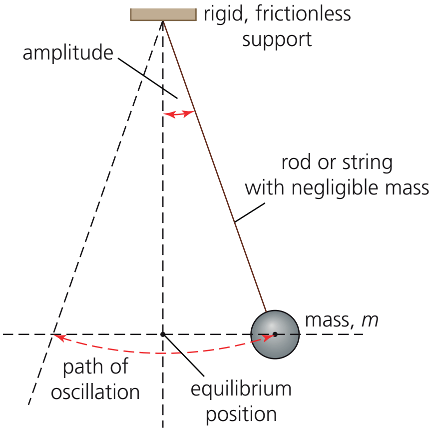


Figure C1.9 Simple pendulum

The weight, mg, of the pendulum is conveniently resolved into two perpendicular components: a force $mg\sin\theta$ provides the restoring force bringing the pendulum back towards its equilibrium position; a force $mg\cos\theta$ keeps the rod / string in tension.

An alternative way of looking at this situation: the pendulum has only two forces acting on it: tension in the rod / string and weight, mg. The resultant of these is the restoring force, $mg\sin\theta$. Practical laboratory investigations of a simple pendulum will confirm that, *for small amplitudes*, the time period, T, depends only its length, l (measured from its point of support to its centre of mass). See Figure C1.11. Period is *not* affected by its mass. This is because doubling the mass, for example, will also result in doubling the restoring force.

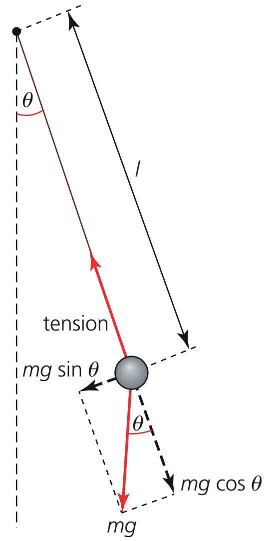


Figure C1.10 Components of the weight of the pendulum

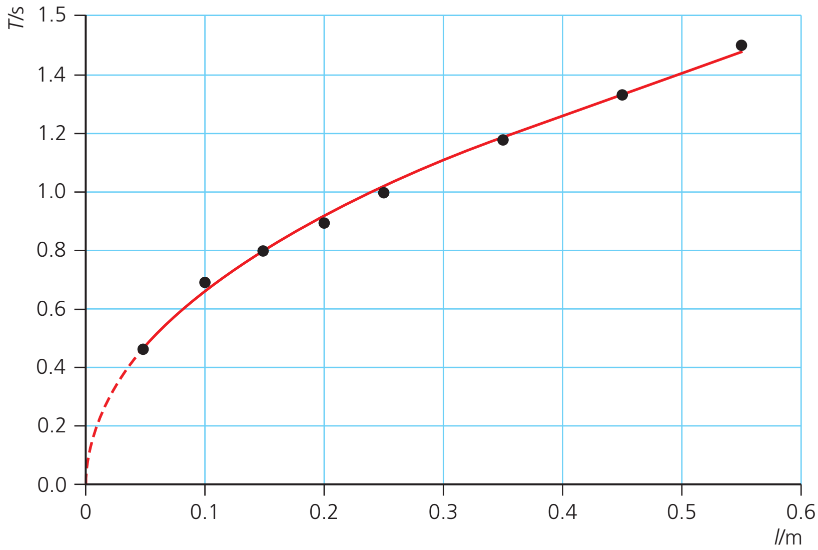


Figure C1.11 Results of simple pendulum investigation

The exact relationship for the SHM of a simple pendulum (for small amplitudes) is as follows.

Time period of a simple pendulum:

$$T = 2\pi\sqrt{\frac{l}{g}}$$

Top tip!

The data of Figure C1.11 can be linearized. A graph of T^2–l should be a straight line through the origin. A simple pendulum experiment can be used to determine a value for the gravitational field strength, g. The graph of T^2–l will have a gradient of: $4\pi^2/g$

Tool 3: Mathematics

Use approximation and estimation

For the motion of a simple pendulum to be a good example of SHM, we require that the restoring force ($mg\sin\theta$) is proportional to the displacement (θ). For very small angles, values of $\sin\theta$, $\tan\theta$ and θ (in rad) are almost identical, so that this condition is satisfied. But, as the angle increases, the difference gets greater. Determine the minimum angle for which there is at least a 1% difference between $\sin\theta$, $\tan\theta$ and θ (in rad).

Figure C1.12 Mount Nevado Huascarán in Peru is reported to have the lowest gravitational field strength on Earth; its summit (6768 m) is one of the farthest points on the surface of the Earth from Earth's centre

ATL C1A: Research skills, thinking skills

Use search engines effectively; providing a reasoned argument to support conclusions

The equation highlighted above shows that the period of a simple pendulum would also change if its value was checked in a location where the gravitational field strength had a different value. Using your own research, investigate the variations of the strength of the gravitational field around the Earth. Then state and explain whether experiments in school laboratories would be able to measure those differences.

WORKED EXAMPLE C1.3

Determine a value for the time period of a simple pendulum of length 85.6 cm at a location where the gravitational field strength is

a 9.81 N kg^{-1}

b 1.63 N kg^{-1} (on the Moon's surface).

c This equation predicts that the ratio of the time periods at two different locations is:

$$\sqrt{\frac{g_1}{g_2}}.$$

Do your answers to **a** and **b** confirm that?

Answer

a $T = 2\pi\sqrt{\frac{l}{g}} = 2 \times \pi \times \sqrt{\frac{0.856}{9.81}} = 1.86\,\text{s}$

b $2 \times \pi \times \sqrt{\frac{0.856}{1.63}} = 4.55\,\text{s}$

c $\frac{T_M}{T_E} = \frac{4.55}{1.86} = 2.45$

$$\sqrt{\frac{g_E}{g_M}} = \sqrt{\frac{9.81}{1.63}} = 2.45$$

Yes, the same ratio is confirmed.

Nature of science: Measurements

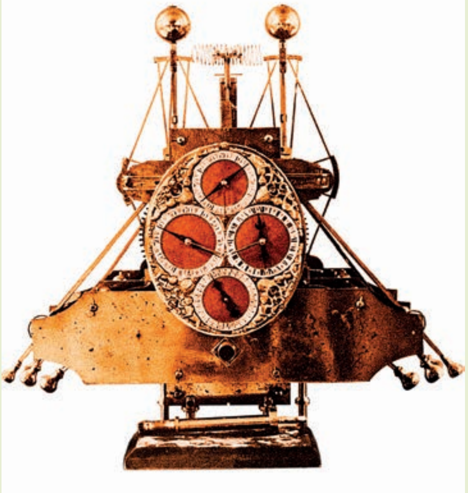
Technological developments in the measurement of time

Famously, Galileo was the first to discover the constant time periods of pendulums, but it was not until about sixty years later, in 1656, that the first pendulum clock was invented by the young Dutch scientist and inventor, Christiaan Huygens. Huygens is widely considered to be one of the greatest scientists and astronomers of all time. By the end of the eighteenth century, the best pendulum clocks could be made with an inaccuracy of less than one second a day. They were used throughout the world as the most accurate timekeepers for more than 250 years.

Accurate timekeepers were also essential for navigation on long journeys by ship. East–west distances (longitude) could be determined from observation of the stars or planets, combined with knowledge of the exact time of observation.

■ **Figure C1.13** Christiaan Huygens

■ **Figure C1.14** Clocks like this were found in many homes



■ **Figure C1.15** James Harrison's first marine timekeeper (1735)

However, pendulum clocks were not designed to cope with the motions of ships on the ocean! The British government offered a valuable reward to anyone who could design a clock that would remain accurate at sea, and hence 'solve the problem of longitude'. It was won by a clockmaker named James Harrison. His design included oscillating spheres on springs, rather than a pendulum.

Today we expect to know the exact time, at any time and place that we want, and we consider such precision to be normal, unworthy of comment. But, before pendulum clocks, most people were unaware of the time to the nearest hour, never mind the nearest minute, or second. (And they probably would not have understood any need for such accuracy!)

It is impossible to over-estimate the importance of the accurate measurement of time in modern life. The physics principle behind the pendulum clock is easily understood, but it took great technological skill (for that era) to manufacture the accurate clocks which had such considerable benefits on everyday life.

6 Determine what mass will have a period of exactly one second when oscillating on the end of a spring which has a spring constant of 84 N m^{-1}.

7 Calculate the angular frequency of a mass–spring system which involves a mass of 1000 g and a spring with $k = 100\,\text{N m}^{-1}$.

8 A student investigated the effect of using springs of different stiffness (k) on the periods of oscillations, T, of the same mass.
Sketch the T–k graph that should be obtained.

9 Perhaps the world's most famous pendulum was made by Léon Foucault in Paris in 1851. The bob had a mass of 28 kg and its length was 67 m.
   a Calculate the period of this pendulum (the gravitational field strength in Paris is 9.81 N kg^{-1}).
   b Suggest reasons why it could continue to swing for a long time without any continuous energy input.
   c Use the internet to find out why this pendulum was so important.

10 Figure C1.11 shows the results of an investigation into a simple pendulum.
   a Calculate the angular frequency which describes oscillations of a pendulum of length 0.50 m.
   b Determine a value for the strength of the gravitational field from these results.

11 Figure C1.16 shows a girl starting to bounce on a trampoline. To begin with, her feet remain in contact with the rubber sheet and her movement can be considered to be SHM. The girl has a mass of 38 kg and the sheet stretched down by 33 cm when she was standing still in the middle.
   a Calculate a value for the spring constant of the trampoline.
   b Determine the time period of her bounces while she remains in contact with the trampoline's surface.

Figure C1.16 Girl on trampoline

12 Figure C1.17 shows a spring which is part of a car's suspension system.

Figure C1.17 The suspension system of a car

   a Estimate its spring constant by considering how much the wing of a car will depress if you push down hard on it (or sit on it).
   b By considering that the wheel effectively oscillates on the spring, estimate
      i its time period of oscillation
      ii how far a car travelling at 12 m s^{-1} moves during one oscillation.
   c The suspension also incorporates a shock absorber (damper). Discuss why this is necessary.

LINKING QUESTIONS

- How can greenhouse gases be modelled as simple harmonic oscillators?
- What physical explanation leads to the enhanced greenhouse effect? (NOS)

These questions link to understandings in Topic B.2.

Conditions that produce SHM

SYLLABUS CONTENT

- Conditions that lead to simple harmonic motion.
- The defining equation of simple harmonic motion as given by: $a = -\omega^2 x$.

SHM will occur if the restoring force, F, is proportional to the displacement, x, but the force always acts back towards the equilibrium position:

$$F \propto -x$$

The negative sign is important here. It shows us that the force acts in the opposite direction to the displacement: an increasing force *opposes* increasing displacement.

For an oscillating mass on a spring which obeys Hooke's law ($F_H = -kx$) this condition is obviously satisfied. For a simple pendulum, the restoring force, $mg\sin\theta$, is (almost) proportional to the displacement angle, θ, if the angle is small.

To define SHM we need to refer to the motion, rather than the force. Since $F = ma$, for a constant mass, we can write $a \propto -x$.

◆ **Simple harmonic motion (SHM)** Oscillations in which the acceleration, a, is proportional to the displacement, x, and in the opposite direction, directed back to the equilibrium position. $a \propto -x$.

Simple harmonic motion is defined as an oscillation in which the acceleration is proportional to the displacement, but in the opposite direction (always directed back towards the equilibrium position):

$$a \propto -x$$

This is represented by the graph shown in Figure C1.18. A graph of the restoring force against displacement would look similar.

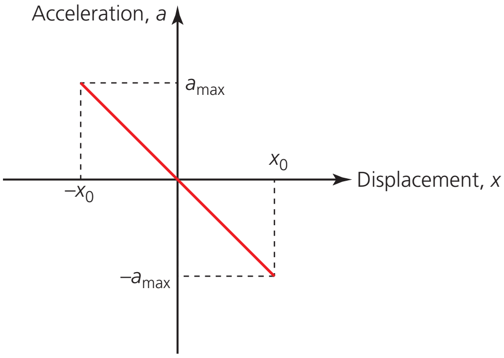


Figure C1.18 Acceleration–displacement graph for SHM

If, for example, the maximum displacement (amplitude) of an SHM is doubled, the restoring force and acceleration will also double. This will result in the mass taking the same time for each oscillation, because it is moving twice the distance at twice the average speed. This explains a defining feature of SHM: amplitude does not affect time period.

We can rewrite the equation above as: $a = -\text{constant} \times x$.

The constant must involve frequency because the magnitude of the acceleration is greater if the frequency is higher. The constant can be shown to be equal to ω^2. So that the defining equation of SHM can be written as:

$$\boldsymbol{a = -\omega^2 x}$$

WORKED EXAMPLE C1.4

A mass oscillates horizontally between two springs with a frequency of 1.4 Hz.

a Calculate its angular frequency.

b Determine its acceleration when

  i its displacement is 1.0 cm to the right

  ii its displacement is 4.0 cm to the left

  iii it passes through its equilibrium position.

Answer

a $\omega = 2\pi f = 2\pi \times 1.4 = 8.8\,\text{rad s}^{-1}$

b **i** $a = -\omega^2 x = -(8.8)^2 \times (+0.010) = -0.77\,\text{m s}^{-2}$ (to the left)

  ii $a = -\omega^2 x = -(8.8)^2 \times (-0.040) = +3.1\,\text{m s}^{-2}$ (to the right)

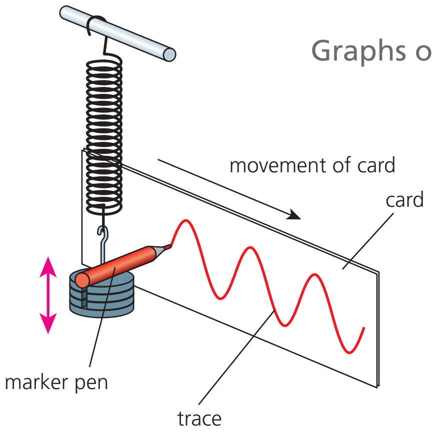
  iii $a = -\omega^2 x = -(8.8)^2 \times 0.0 = 0.0\,\text{m s}^{-2}$

Graphs of SHM

A mass oscillating on a spring could be used with a marker pen to produce a record of the oscillation, as shown in Figure C1.19. If the card moves at a constant speed, the record (trace) produced is effectively a graph of displacement against time. **Data logging** with an appropriate sensor will improve this experiment.

The waveform seen in Figure C1.19 is commonly described as being **sinusoidal** in shape.

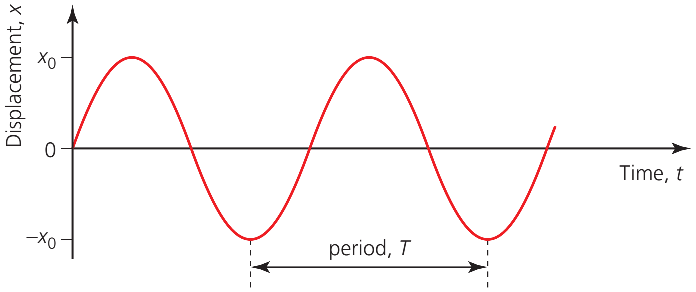
The graph in Figure C1.20 shows the variation in displacement, x, with time, t, for the idealized model of a mass moving with perfect SHM. Here we have *chosen* that the particle has zero displacement at the start of the timing. The motion has an amplitude of x_0. Alternatively, we might have chosen the graph to start with the particle at maximum displacement (or any other displacement).



■ **Figure C1.19** Recording the motion of an oscillation

◆ **Data logging** Connecting sensors to a computer with suitable software to enable physical quantities to be measured and processed digitally.

◆ **Sinusoidal** In the shape of a sine wave (usually equivalent to a cosine wave).



■ **Figure C1.20** Displacement–time graph for simple harmonic motion. Timing was started when the particle had zero displacement

Using knowledge from Topic A.2, we can use this graph to calculate the velocity of the oscillating mass at any particular moment by determining the gradient of the displacement–time graph at that moment:

$$\text{velocity, } v = \frac{\text{change in displacement}}{\text{change in time}} = \frac{\Delta x}{\Delta t}$$

Similarly, the acceleration at any given time can be found from the gradient of the velocity–time graph:

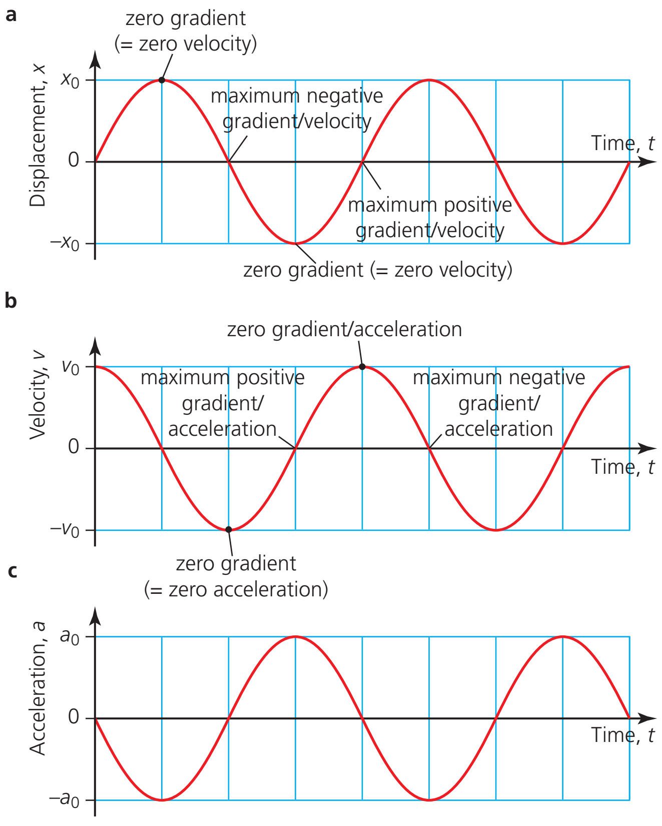
$$\text{acceleration, } a = \frac{\text{change in velocity}}{\text{change in time}} = \frac{\Delta v}{\Delta t}$$

Using this information, three separate but interconnected graphs of motion can be drawn and compared, as shown in Figure C1.21.

The velocity graph has its maximum value, v_0, when the displacement, x, is zero, and the velocity is zero when the displacement is at its maximum, $\pm x_0$.

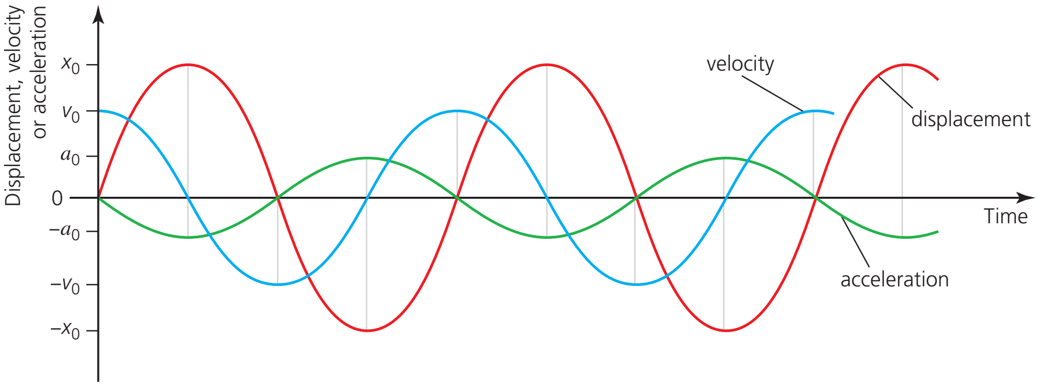
The acceleration has its maximum value when the velocity is zero and the displacement is greatest. This is to be expected, because when the displacement is greatest, the restoring force, acting in the opposite direction, is greatest.

SHM graphs of displacement, velocity and acceleration are all sinusoidal in shape, but their maximum values occur at different times.



■ **Figure C1.21** Graphs for simple harmonic motion starting at displacement $x = 0$: **a** displacement–time **b** velocity–time **c** acceleration–time

Figure C1.22 shows all three SHM graphs drawn on the same axes, so that they can be more easily compared. Note that the amplitudes of the three graphs are arbitrary; they are not interconnected and should not be compared.



■ **Figure C1.22** Comparing displacement, velocity and acceleration for SHM, with timing starting at displacement $x = 0$

LINKING QUESTION

- How can the understanding of simple harmonic motion apply to the wave model? (NOS)

This question links to understandings in Topic C.2.

■ **Figure C1.23** Four oscillations out of phase

Phase difference

This is a convenient point to introduce the important concept of **phase difference**. Figure C1.23 shows an everyday example. The motions of the four children all have the same frequency because all the swings are the same length, but they are each at different points in their oscillations: we say that they are *out of phase* with each other.

A phase difference occurs between two similar oscillations if they have the same frequency, but their maximum values do not occur at the same time.

◆ **Phase difference** When oscillators that have the same frequency are out of phase with each other, the difference between them is defined by the angle (usually in terms of π radians) between the oscillations. Phase differences can be between 0 and 2π radians.

The three graphs shown in Figure C1.21 all have the same frequency and sinusoidal shape, but their peaks occur at different times: there is a phase difference between the waves. This could be quantified by referring to the fraction of a time period, T, that occurs between their peaks. The first peak of the displacement graph occurs $T/4$ after the first peak of the velocity graph. The first peak of the acceleration graph occurs $3T/4$ after the first peak of the velocity graph.

But, remembering that one complete oscillation can be considered as equivalent to moving through an angle of 2π radians, phase differences are more usually quoted in terms of π.

Examples of phase differences

- One quarter of an oscillation, $\frac{T}{4} = \frac{2\pi}{4} = \frac{\pi}{2}$ radians (90°)
- One half of an oscillation, $\frac{T}{2} = \frac{2\pi}{2} = \pi$ radians (180°)
- Three quarters of an oscillation, $\frac{3T}{4} = 2\pi \times \frac{3}{4} = \frac{3\pi}{2}$ radians (270°)

(A phase difference of $\frac{3\pi}{2}$ radians is equal in magnitude to a phase difference of $\frac{\pi}{2}$ radians, but in some circumstances, we might be concerned about which peak came 'first'.)

Referring back to Figure C1.21, can see that there is a phase difference of $\frac{\pi}{2}$ between displacement and velocity, and a phase difference of π between displacement and acceleration.

13 An oscillator moves with SHM and has a time period of 2.34 s.

How far must it be displaced from its equilibrium position in order that its acceleration is $1.00\,\text{m}\,\text{s}^{-2}$?

14 During SHM a mass moves with an acceleration of $3.4\,\text{m}\,\text{s}^{-2}$ when its displacement is 4.0 cm. Calculate its:

a angular frequency **b** time period.

15 A mass oscillating on a spring performs exactly 20 oscillations in 15.8 s.

- **a** Determine its acceleration when it is displaced 62.3 mm from its equilibrium position.
- **b** State any assumption that you made when answering **a**.

16 Look at the graph in Figure C1.24 which shows the motion of a mass oscillating on a spring. Determine:

- **a** the amplitude
- **b** the time period
- **c** the displacement after 0.15 s
- **d** the displacement after 1.4 s.

17 a Sketch a displacement–time graph showing two complete oscillations for a simple harmonic oscillator which has a time period of 2.0 s and an amplitude of 5.0 cm.

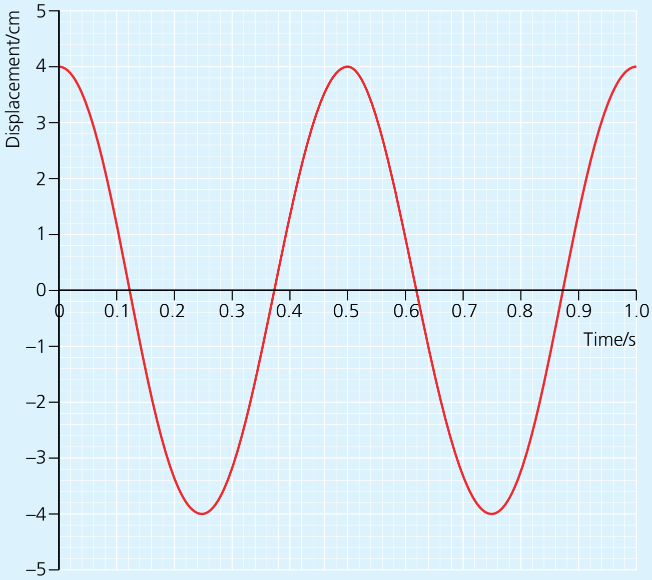
- **b** Add to the same axes the graph of an oscillator which has twice the frequency and the same amplitude.
- **c** Add to the same axes the graph of an oscillator that has an amplitude of 2.5 cm and the same frequency but which is ¼ of an oscillation out of phase with the first oscillator.

18 a Sketch a velocity–time graph showing two complete oscillations for a simple harmonic oscillator which has a frequency of 4 Hz and a maximum speed of $4.0\,\text{c}\,\text{m}\text{s}^{-1}$.

b On the same axes sketch graphs to show the variation of:

i displacement **ii** acceleration

for the same oscillation.



■ **Figure C1.24** Motion of a mass oscillating on a spring

There is more about phase difference, graphs of motion, and the equations which represent them, towards the end of this topic (for HL students).

Energy changes during SHM

SYLLABUS CONTENT

▶ A qualitative approach to energy changes during one cycle of an oscillation.

When an object which can oscillate is pushed, or pulled, away from its equilibrium position, against the action of a restoring force, work will be done and potential energy will be stored in the oscillator. For example, a spring will store elastic potential energy and a simple pendulum will store gravitational potential energy. When the object is released, it will gain kinetic energy and lose potential energy as the restoring force accelerates it back towards the equilibrium position. Its kinetic energy has a maximum value as it passes through the equilibrium position and, at the same time, its potential energy is minimized. As it moves away from the equilibrium position, kinetic energy decreases as the restoring force opposes its motion and potential energy increases again.

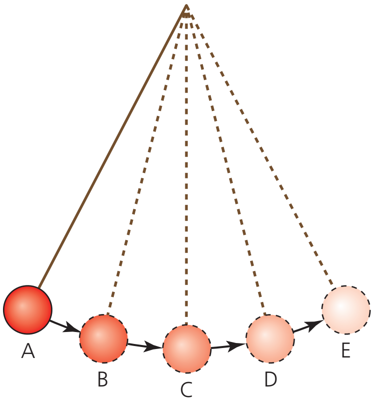


Figure C1.25 A swinging pendulum

As an example, consider Figure C1.25, in which a simple pendulum has been pulled away from its equilibrium position, C, to position A. While it is held in position A, it has zero kinetic energy and its change of gravitational potential energy (compared to position C) is greatest. When it is released, gravity provides the restoring force and the pendulum exchanges potential energy for kinetic energy as it moves through position B to position C. At C it has maximum kinetic energy and the change in potential energy has reduced to zero. The pendulum then transfers its kinetic energy back to potential energy as it moves through position D to position E. At E, like A, it has zero kinetic energy and a maximum change of potential energy. The process then repeats every half time period.

If the pendulum was a perfect simple harmonic oscillator, there would be no energy losses, so that the sum of the potential energy and the kinetic energy would be constant and the pendulum would continue to reach the same maximum vertical height and maximum speed every oscillation. In practice, frictional forces will result in energy dissipation and all the energies of the pendulum will progressively decrease.

All mechanical oscillations involve a continuous exchange of energy between kinetic energy and some form of potential energy.

For a simple pendulum, the potential energy *E*p is in the form of gravitational potential energy (see Topic A.3 where ΔE_p was used instead of E_p):

$$E_p = mg\Delta h$$

For a mass on a spring, the potential energy is in the form of elastic potential energy (see Topic A.3, where the symbol E_H was used instead of E_p and Δx was used instead of x):

$$E_p = \frac{1}{2}kx^2$$

Its maximum value is:

$$E_{p_{max}} = \frac{1}{2}kx_0^2$$

which shows us that generally:

The total energy of an SHM is proportional to its amplitude squared.

Kinetic energy of the mass can be determined from (Topic A.3):

$$E_k = \frac{1}{2}mv^2$$

Figure C1.26 represents these exchanges during several oscillations of perfect SHM.

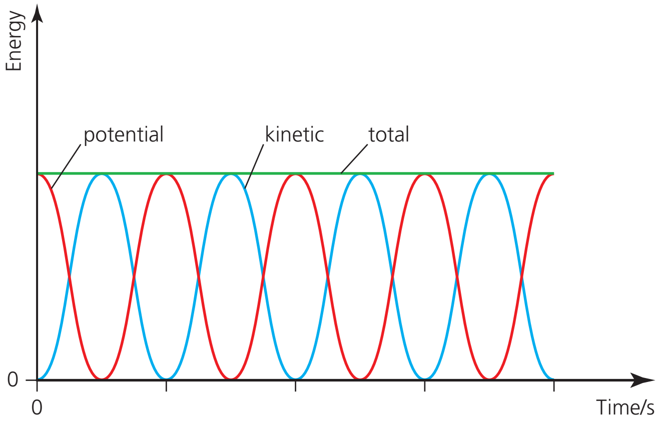


Figure C1.26 Energy changes during oscillation of perfect SHM

◆ **Damping** Occurs when resistive forces act on an oscillating system, dissipating energy and reducing amplitude.

In more realistic situations, the energies of the system will decrease over time. This is shown in Figure C1.27. Energy dissipation from an oscillating system is called **damping**.

Figure C1.28 shows the variations in energy with displacement during each oscillation of SHM.

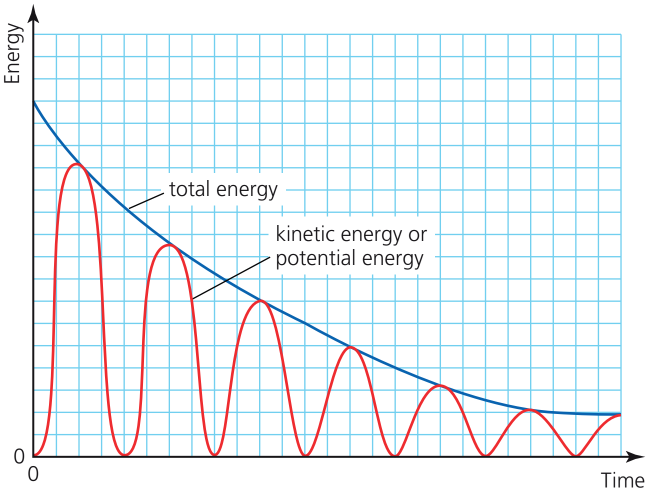


Figure C1.27 Energy dissipation in a practical oscillator

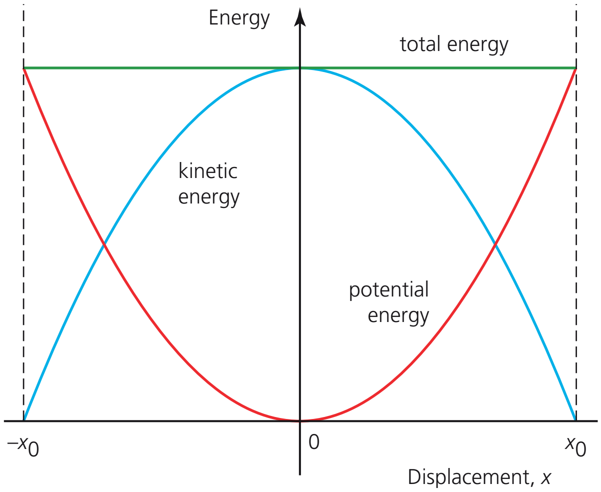


Figure C1.28 Variation of energies of a simple harmonic oscillator with displacement

WORKED EXAMPLE C1.5

A mass oscillating in a mass–spring system has a maximum kinetic energy of 0.047 J.

a If its maximum speed was 85 cm s^{-1}, determine its mass.

b State the maximum value of its potential energy.

c Determine the spring constant if the amplitude of the oscillation was 1.9 cm.

Answer

a $E_k = \frac{1}{2}mv^2$

$0.047 = \frac{1}{2} \times m \times 0.85^2$

$m = 0.13\,\text{kg}$

b 0.047 J

c $E_{p_{max}} = \frac{1}{2}kx_0{}^2$

$0.047 = \frac{1}{2} \times k \times (1.9 \times 10^{-2})^2$

$k = 2.6 \times 10^2\,\text{N m}^{-1}$

19 A simple pendulum, of mass 100 g, is released from rest at its maximum displacement, which is 5.0 cm higher than its central position. It then swings with SHM and a frequency of 1.25 Hz.

a Calculate its maximum potential energy.

b Sketch a graph with fully labelled axes to show the variations in potential energy of the pendulum in the first 1.6 s after it was released.

c Determine the maximum speed of the pendulum.

20 The total energy seen in Figure C1.27 shows a nearly 90% decrease in six oscillations.

a Describe where this energy has been transferred to.

b If the oscillating mass was passing through its equilibrium position at time $t = 0$, state what kind of energy is represented by the red line.

c It is suggested that the total energy decreases to the same fraction each oscillation. Analyse data from the graph to check if this is true.

21 a Use the equation:

$$E_p = \frac{1}{2}kx^2 \text{ where } k = 12\,\text{N m}^{-1}$$

to calculate values of elastic potential energy stored in a mass–spring system for SHM displacements, x, of 0, ±0.05 m, ±0.10 m, ±0.15 m, ±0.20 m. Its amplitude was 0.20 m.

b Sketch a graph to display these results.

c On the same axes, add a graph to represent the variations in kinetic energy of the mass.

d If the mass was 50 g, determine its maximum speed.

ATL C1B: Social skills

Resolving conflicts during collaborative work

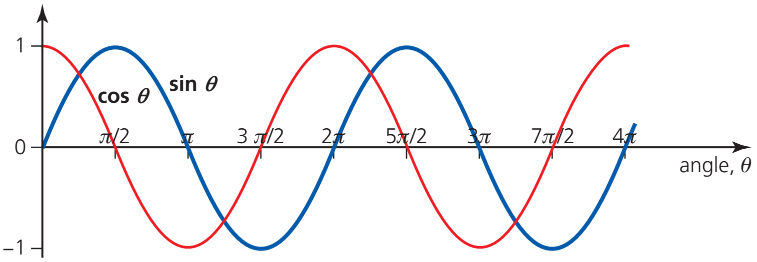
A group of four students was asked by their teacher to investigate four different oscillators and report back to the rest of their group one week later. One of the four students was chosen to be the team-leader. Three of the students worked well, but the fourth showed no interest and did little. What should the team-leader do about this situation?

Calculating displacements and velocities during SHM

SYLLABUS CONTENT

- ▶ A particle undergoing simple harmonic motion can be described using phase angle.
- ▶ Problems can be solved using the equations for simple harmonic motion as given by:
  $x = x_0 \sin(\omega t + \phi)$
  $v = \omega x_0 \cos(\omega t + \phi)$

Figure C1.29 shows the variations of a perfect sine wave and compares it to a cosine wave. The two waves have identical shapes, but there is a phase difference of $\pi/2$ between them.



■ **Figure C1.29** Comparing sine and cosine waves

The shape of the graph of SHM shown in Figure C1.20 is identical to the sine wave shown in Figure C1.29 and it could be represented by the equation of the form $x = x_0 \sin\theta$.

We know from Topic A.4:

$$\text{angular frequency, } \omega = \frac{\Delta\theta}{\Delta t}\left(\text{or } \frac{\theta}{t}\right)$$

so that: θ (radians) = ωt. And then we then can rewrite: $x = x_0 \sin\theta$ as $x = x_0 \sin \omega t$.

This equation assumes that the initial value of x = 0 when t = 0. Under those circumstances, it enables us to calculate the displacement of a SHM at any time we choose (if the amplitude and frequency are known).

◆ **Phase angle** The difference in angular displacement of an oscillation compared to an agreed reference point. Expressed in terms of π radians.

In order to write an equation which allows for all other possibilities, we need to return to the concept of phase difference. We have already discussed the phase difference between two oscillations, but now we will refer to the **phase angle** of a *single* oscillator (sometimes just called *phase*). The oscillations are compared to a theoretical sine wave, which has zero displacement at time $t = 0$.

The phase angle, ϕ, of an oscillation is the fraction of an oscillation (expressed in terms of π) that occurs between when it has zero displacement and a sine wave which has zero displacement at time $t = 0$.

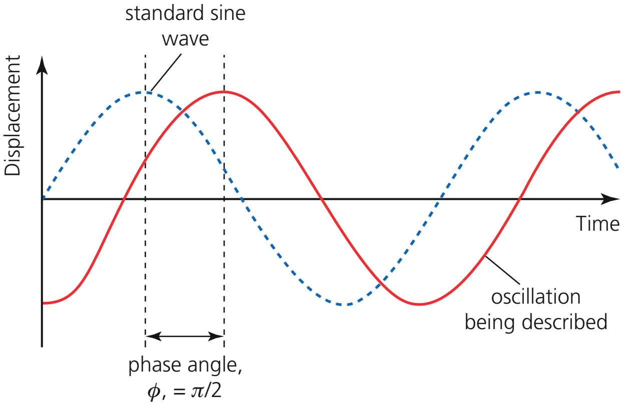


Figure C1.30 Phase angle of an oscillator

Figure C1.30 shows an example.

We can now write a full equation to describe the variation of displacement with time for any SHM:

displacement, $\boldsymbol{x = x_0 \sin(\omega t + \phi)}$

For example, if the lines in Figure C1.30 show the variations in displacement of two oscillators with the same frequency and amplitude, the dotted blue line can be represented by:

$$x = x_0 \sin \omega t$$

while the red line has the equation:

$$x = x_0 \sin\left(\omega t + \frac{3\pi}{2}\right)$$

We have seen (Figure C1.22) that the variations of displacement, x, and velocity, v, for the same SHM oscillator are $\pi/2$ out of phase with each other. So that, the equation representing the velocity of a SHM involves a cosine:

velocity, $\boldsymbol{v = \omega x_0 \cos(\omega t + \phi)}$

Common mistake

The presence of ω at the start of the term on the right-hand side need *not* be explained here, but it is a common mistake to leave it out of calculations.

LINKING QUESTION

- How can circular motion be used to visualize simple harmonic motion?

This question links to understandings in Topic A.2.

The maximum velocity, v_0, will occur when the cosine in the equation has a value of one:

maximum velocity, $v_0 = \omega x_0$

To help to understand these equations, consider first the simplest possible numerical example, with $T = f = 1$ and $x_0 = 1$, for an oscillation beginning at $x = 0$ when $t = 0$. The phase angle is 0, so that the equation for displacement reduces to $x = \sin(2\pi t)$.

Using this equation at, for example, $t = 0.20\,\text{s}$, leads to $x = +0.95\,\text{m}$, and at $t = 0.60\,\text{s}$, $x = -0.59\,\text{m}$

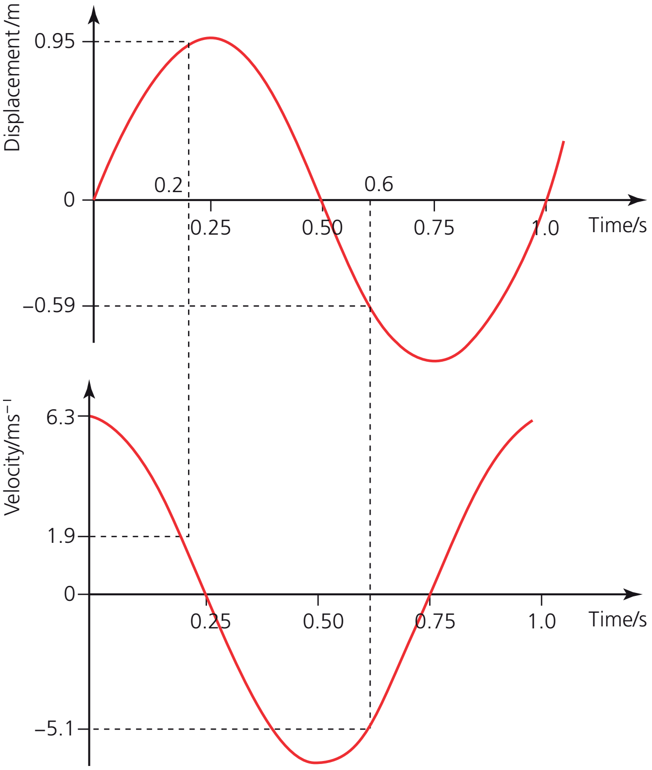
The negative sign shows that the displacement at $t = 0.60\,\text{s}$ was in the opposite direction to the initial displacement (just after $t = 0$).

Using $v = 2\pi \times \cos(2\pi t)$ with the same data, we get: at $t = 0.20$ s, $v = +1.9\,\mathrm{m\,s^{-1}}$, and at $t = 0.60$ s, $v = -5.1\,\mathrm{m\,s^{-1}}$.

The maximum velocity, $v_0 = \omega x_0 = 2\pi = 6.3\,\mathrm{m\,s^{-1}}$.

The negative sign shows that the velocity at $t = 0.60$ s was in the opposite direction to the velocity when $t = 0$.

This data is shown in Figure C1.31.



■ **Figure C1.31** Numerical data for displacement and velocity of SHM shown on a graph

WORKED EXAMPLE C1.6

A mass oscillates with SHM of frequency 2.7 Hz and amplitude 1.7 cm. If its phase angle is $\pi/2$, calculate:

a its displacement after 2.0 s

b its velocity after 3.0 s

c its maximum velocity.

Answer

Remembering that $\omega = 2\pi f = 2 \times \pi \times 2.7 = 5.4\pi\,\mathrm{rad\,s^{-1}}$:

a $x = x_0 \sin(\omega t + \phi) = (1.7 \times 10^{-2}) \times \sin\left((5.4\pi \times 2.0) + \dfrac{\pi}{2}\right) = -1.4 \times 10^{-2}\,\mathrm{m}$

The negative sign shows that the displacement was in the opposite direction to the displacement just after $t = 0$.

b $v = \omega x_0 \cos(\omega t + \phi) = (5.4\pi \times 1.7 \times 10^{-2}) \times \cos\left((5.4\pi \times 3.0) + \dfrac{\pi}{2}\right) = -0.17\,\mathrm{m\,s^{-1}}$

The negative sign shows that the velocity at $t = 3.0$ s was in the opposite direction to the velocity when $t = 0$.

c $v_0 = \omega x_0 = 5.4\pi \times (1.7 \times 10^{-2}) = 0.29\,\mathrm{m\,s^{-1}}$

22 A mass is oscillating between two springs with a frequency of 1.5 Hz and amplitude of 3.7 cm. It has a speed of 34 cm s^{-1} as it passes through its equilibrium position and a stopwatch is started.
Calculate its displacement and velocity 1.8 s later.

23 An object of mass 45 g undergoes SHM with a frequency of 12 Hz and an amplitude of 3.1 mm.
- **a** Determine its maximum speed and kinetic energy.
- **b** What is the object's displacement 120 ms after it is released from its maximum displacement?

24 A mass is oscillating with SHM with an amplitude of 3.8 cm. Its displacement is 2.8 cm at 0.022 s after it is released from its maximum displacement.
Calculate a possible value for its frequency.

25 A simple harmonic oscillator has a time period of 0.84 s and its speed is 0.53 m s^{-1} as it passes through its mean (equilibrium) position.
- **a** Calculate its speed 2.0 s later.
- **b** If the amplitude of the oscillation is 8.9 cm, what was the displacement after 2.0 s?

26 The water level in a harbour rises and falls with the tides, with a time of 12 h 32 min for a complete cycle. The high tide level is 8.20 m above the low tide level, which occurred at 4.10 am.
If the tides rise and fall with SHM, determine the level of the water at 6.00 am.

27 Discuss what the area under a velocity–time graph of an oscillation represents.

Calculating energy changes during SHM

SYLLABUS CONTENT

- Problems can be solved using the equations for simple harmonic motion as given by:

$$v = \pm\omega\sqrt{x_0^{\,2} - x^2}$$

$$E_T = \frac{1}{2}m\omega^2 x_0^{\,2}$$

$$E_p = \frac{1}{2}m\omega^2 x^2$$

We have already discussed the energy exchanges that occur during SHM qualitatively (see Figures C1.26, C.127 and C.128). Now we will interpret those changes in more mathematical detail.

We saw in Topic A.3 that elastic potential energy can be determined from:

$$E_p = \frac{1}{2}kx^2$$

where k is the spring constant.

We also know from earlier in this sub-topic that for a mass–spring oscillator:

$$\omega^2 = \frac{k}{m}$$

which leads to:

potential energy, $\boldsymbol{E_p = \frac{1}{2}m\omega^2x^2}$

When the mass is at its maximum displacement, x_0, its velocity has reduced to zero. It has zero kinetic energy and it has its maximum potential energy - which is then equal to the total energy, E_T, of the SHM:

total energy, $\boldsymbol{E_T = \frac{1}{2}m\omega^2x_0^{\,2}}$

Since at any point, total energy = potential energy + kinetic energy.

kinetic energy, $E_k = \frac{1}{2}m\omega^2x_0^{\,2} - \frac{1}{2}m\omega^2\,x^2 = \frac{1}{2}m\omega^2\,(x_0^{\,2} - x^2)$

The instantaneous velocity, v, at any point in the oscillation can be found by equating:

$$E_k = \frac{1}{2}mv^2$$

with the previous equation. Which leads to:

$\boldsymbol{v = \pm\,\omega\sqrt{(x_0^{\,2} - x^2)}}$

WORKED EXAMPLE C1.7

An oscillator of mass 750 g oscillates with SHM with an amplitude of 3.47 cm and a period of 1.44 s.

- **a** Calculate its total energy.
- **b** When it has a displacement of 2.15 cm determine:
  - **i** its potential energy
  - **ii** its kinetic energy.
- **c** When its kinetic energy is 2.00×10^{-3} J, what is its displacement?
- **d** Calculate the velocity of the mass when its displacement is 2.00 cm.
- **e** What is the maximum velocity of the mass?

Answer

a $E_T = \frac{1}{2}m\omega^2 x_0^2$

$$= 0.5 \times 0.750 \times \left(\frac{2\pi}{1.44}\right)^2 \times 0.0347^2 = 8.60 \times 10^{-3}\ \text{J}$$

b i $E_p = \frac{1}{2}m\omega^2 x^2$

$$= 0.5 \times 0.750 \times \left(\frac{2\pi}{1.44}\right)^2 \times 0.0215^2 = 3.30 \times 10^{-3}\ \text{J}$$

ii $E_k = E_T - E_p = (8.60 \times 10^{-3}) - (3.30 \times 10^{-3}) = 5.30 \times 10^{-3}\ \text{J}$

c $E_k = \frac{1}{2}m\omega^2(x_0^2 - x^2)$

$$2.00 \times 10^{-3} = 0.5 \times 0.750 \times \left(\frac{2\pi}{1.44}\right)^2 \times (x_0^2 - x^2)$$

$$(x_0^2 - x^2) = 2.80 \times 10^{-4}$$

$$x^2 = 0.0347^2 - (2.80 \times 10^{-4})$$

$$x = 9.24 \times 10^{-4}\ \text{m}$$

d $v = \pm\omega\sqrt{(x_0^2 - x^2)} = \left(\frac{2\pi}{1.44}\right) \times \sqrt{(3.47^2 - 2.00^2)} = \pm 12.4\ \text{m s}^{-1}$ (in either direction)

e v_{max} occurs when displacement is zero.

$v_{max} = \pm\omega\sqrt{(x_0^2 - 0^2)} = \pm\omega x_0 = \pm\left(\frac{2\pi}{1.44}\right) \times 3.47 = \pm 15.1\ \text{m s}^{-1}$ (in either direction)

28 A mass of 480 g is suspended on a spring of stiffness 132 N m^{-1}.

- **a** If it undergoes SHM, calculate its time period.
- **b** Calculate its angular frequency.
- **c** If the oscillations have an amplitude of 3.2 cm, determine:
  - **i** its maximum kinetic energy
  - **ii** its maximum speed.
- **d** Calculate how much potential energy is stored in the system when the displacement is 3.2 cm.

29 An SHM oscillator has a mass of 0.42 kg and a total energy of 1.7 J. If its frequency is 5.7 Hz, determine the amplitude of its motion.

30 A student stretched a vertical spring by placing a mass of 100 g on its end. A second 100 g mass was added and the length of the spring increased by a further 4.7 cm.

- **a** Assuming that it obeyed Hooke's law, determine the spring constant.
- **b** The combined mass of 200 g was then displaced by 5.0 cm so that it oscillated with SHM. What was its period?
- **c** Calculate how much energy was stored in these oscillations.
- **d** Show that the value of the ratio E_k/E_p when the displacement was 2.0 cm was about 5/1.
- **e** Determine by what factor the total energy would increase if the amplitude was increased to 8 cm.

31 A mass was oscillating with SHM at a frequency of 7.6 Hz.

- **a** If its maximum speed was 1.4 m s^{-1}, determine the amplitude of its motion.
- **b** If the mass was 54 g, determine the kinetic energy of it when its displacement was 1.8 cm.

C.2 Wave model

Guiding questions

- What are the similarities and differences between different types of waves?
- How can the wave model describe the transmission of energy as a result of local disturbances in a medium?
- What effect does a change in the frequency of oscillation or medium through which the wave is travelling have on the wavelength of a travelling wave?

What is a wave?

SYLLABUS CONTENT

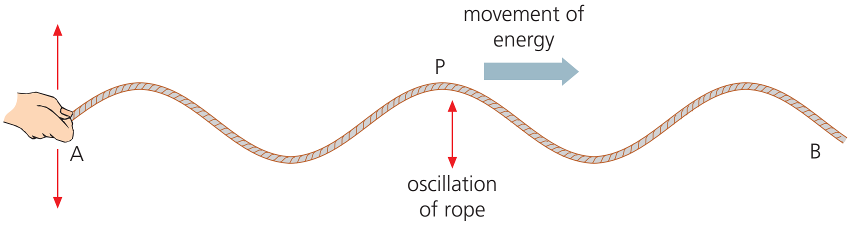
- ▶ The differences between mechanical waves and electromagnetic waves.

■ **Figure C2.1** Circular waves spreading out on a pond

When we think of waves the first example that comes to mind is probably that of waves on the surface of water, like those seen in Figure C2.1.

When the equilibrium of the surface is disturbed (for example by a falling drop, or touching it with a finger), it results in oscillations of the water surface at that point. Because of the forces between water molecules, the oscillations are transferred to neighbouring molecules a short time later, and then they spread outwards as a two-dimensional wave on the water surface.

A simpler, one-dimensional, example is shown in Figure C2.2: in this example the waves are produced by continuously shaking one end of the rope. Point A is the oscillating source of the wave energy, which travels to the other end, point B. All points on the rope, point P for example, oscillate up and down (as shown)..



■ **Figure C2.2** Creating a wave by shaking the end of a rope

◆ **Wave (travelling)** A wave that transfers energy away from a source. Sometimes called a progressive wave.

◆ **Propagation (of waves)** Transfer of energy by waves.

◆ **Medium (of a wave)** Substance through which a wave is passing.

These two examples are both **travelling waves**. Another kind of wave (a standing wave) is discussed in Topic C.4.

Scientists describe the motion of a wave away from its source as **propagation** of the wave. The matter through which the waves pass is called the **medium** of the wave.

All waves involve oscillations and they can be described as being either ‘mechanical’ or ‘electromagnetic’:

◆ **Wave (mechanical)** A wave which involves oscillating masses (including sound).

◆ **Wave (electromagnetic)** A transverse wave composed of perpendicular electric and magnetic oscillating fields travelling at a speed of $3.0 \times 10^8\,\text{m s}^{-1}$ in free space.

Mechanical waves involve the oscillations of masses.

Electromagnetic waves, such as light, involve the oscillations of electric and magnetic fields.

The first part of this topic will deal with mechanical waves. Electromagnetic waves are discussed later.

A mechanical travelling wave can be described as an oscillating disturbance that travels away from its source through the surrounding medium (solid, liquid or gas) transferring energy from one place to another. Most importantly, waves transfer energy without transferring the matter itself.

■ **Figure C2.3** Ocean waves transferring a large amount of energy at Brighton, England – there is no continuous net movement of the water itself

For example, ocean waves may ‘break’ and ‘crash’ on to a shore or rocks, transferring considerable amounts of energy (that they got from the wind), but there is no net, continuous movement of water from the ocean to the land. A wooden log floating on a lake will simply oscillate up and down as waves pass (unless there is a wind).

Examples of mechanical waves

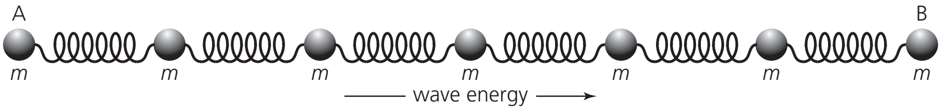
- waves on strings, ropes and springs
- waves on water
- sound (and similar waves in liquids and solids)
- earthquake waves.

Models of mechanical waves

SYLLABUS CONTENT

- ▶ Transverse and longitudinal travelling waves.

In order to understand more about the propagation of mechanical waves it is convenient to visualize the *continuous* medium in which they are travelling as being composed of *separate* (discrete) *particles* of mass, *m*, separated by springs representing the restoring forces that arise when the medium is disturbed from is equilibrium position. See Figure C2.4. The wave can be produced by shaking the end A, the wave then travels along the system to B.



■ **Figure C2.4** Wave model of masses and springs

Experiments can confirm that the speed of the wave along the system increases if the masses are smaller, or if the springs are stiffer.

There are two different ways in which A can be shaken: left–right–left–right, or up–down–up–down (as shown). This identifies the two basic kinds of mechanical wave: transverse and longitudinal.

◆ **Transverse wave** A wave in which the oscillations are perpendicular to the direction of transfer of energy.

◆ **Crest** Highest part of a transverse mechanical wave.

◆ **Trough** Lowest point of a transverse mechanical wave.

◆ **Longitudinal wave** Waves in which the oscillations are parallel to the direction of transfer of energy.

◆ **Compressions (in a longitudinal wave)** Places where there are increases in the density and pressure of a medium as a wave passes through it.

◆ **Rarefactions (in a longitudinal wave)** Places where there are reductions in the density and pressure of a gas as a wave passes through it.

Transverse and longitudinal mechanical waves

In a **transverse wave**, each part of the medium oscillates *perpendicularly* to the direction in which the wave is transferring energy.

The waves shown in Figure C2.1, C2.2 and C2.3 are transverse waves. The black line in Figure C2.5 represents the positions of the particles in a continuous medium which is transferring wave energy to the right. The arrows show which way the particles are moving at that moment. The red line represents their positions a short time later. Each particle is oscillating with the same amplitude and frequency, but each particle is slightly out of phase with its neighbour.

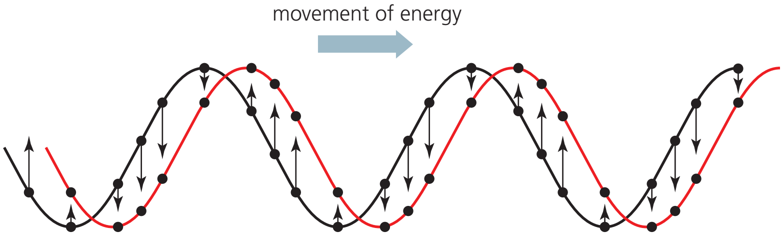


Figure C2.5 Movement of particles as a transverse wave moves to the right

The tops of transverse waves are often called **crests**, while the bottoms of the waves are called **troughs**.

Mechanical waves on strings, ropes and water surfaces are all transverse in nature.

In a **longitudinal wave**, each part of the medium oscillates *parallel* to the direction in which the wave is transferring energy.

Stretched springs are often used to demonstrate waves. They are more massive than strings and this reduces the wave speed, so that the waves can be observed more easily. Stretched 'slinky' springs are particularly useful for demonstrating longitudinal waves. See Figure C2.6, which shows the characteristic **compressions** and **rarefactions** of longitudinal waves on a 'slinky'. Longitudinal waves are sometimes called compression (or pressure) waves.

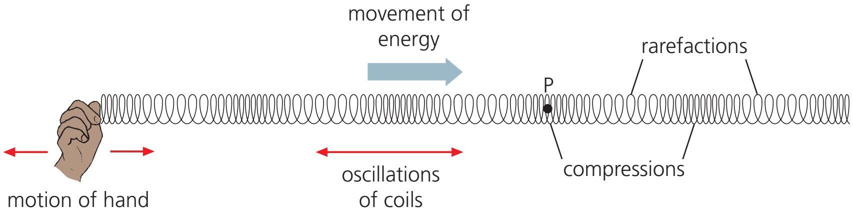


Figure C2.6 Oscillations of a spring transferring a longitudinal wave

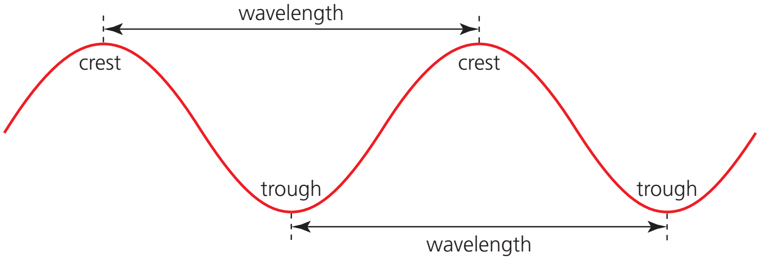
Sound travelling through air is a good example of a longitudinal wave (more details below). Longitudinal compression waves can travel through solids and liquids. Earthquakes are a combination of longitudinal and transverse waves. Transverse mechanical waves cannot travel through gases (or liquids) because of the random nature of molecular movements.

Terms used to describe all types of waves

SYLLABUS CONTENT

- Wavelength, λ, frequency, f, time period, T, and wave speed, v, applied to wave motion as given by:
$v = f\lambda = \frac{\lambda}{T}$

The concept of wavelength, λ, is central to the study of waves. See Figure C2.7.



■ **Figure C2.7** One wavelength of a transverse wave

◆ **Wavelength, λ** The distance between two adjacent crests of a wave. More precisely: the shortest distance between two points moving in phase.

◆ **Time period, T** The time taken for one complete wave to pass a point.

◆ **Wave speed, v** The speed at which energy is transferred by a wave.

One **wavelength**, λ, is the shortest distant between two crests, or two troughs. Or the shortest distance between two compressions or rarefactions in a longitudinal wave. More generally, it is defined as the shortest distance between two points moving in phase (SI unit: m).

Displacement, amplitude, time period and frequency have all been discussed before (Topics A.2 and C.1) and are defined in a similar way in the study of waves:

The *amplitude* of a wave is the maximum displacement of the medium from its equilibrium position.

We saw in Topic C.1 that the energy of an oscillation was proportional to its amplitude squared. So, speaking generally, waves with greater amplitude transfer more energy. (We will see in Topic C.3 that the *intensity* of a wave is proportional to its amplitude squared.)

The **time period** of a wave, T, is the time for one oscillation of a particle within the medium, or the time it takes for one complete wave to pass a particular point (unit: second).

The frequency of a wave, f, is the number of oscillations per second of a particle within the medium, or the number of waves to pass a particular point in one second (SI unit: hertz). The following equation is repeated from Topic C.1:

$$f = \frac{1}{T}$$

A wave travels forward one wavelength, λ, every time period, T.

Therefore:

wave speed, $v = \dfrac{\lambda}{T}$

Since $T = 1/f$, we can write:

$$\text{wave speed, } v = f\lambda \left(\text{or } v = \frac{\lambda}{T}\right)$$

WORKED EXAMPLE C2.1

Water waves are passing into a harbour. Five crests are separated by a distance of 9.6 m. An observer notes that 12 waves pass during a time of one minute. Determine:

- **a** the wavelength
- **b** the period
- **c** the frequency
- **d** the speed of the waves.

Answer

a $\lambda = \dfrac{9.6}{4} = 2.4\,\text{m}$

b $T = \dfrac{60}{12} = 5.0\,\text{s}$

c $f = \dfrac{1}{5.0} = 0.20\,\text{Hz}$

d $v = f\lambda = 0.20 \times 2.4 = 0.48\,\text{m}\,\text{s}^{-1}$

1 Consider Figure C2.1. Explain why the amplitude of the waves decreases as they spread away from the central point.

2 Consider Figure C2.2.
  a State the type of wave which is travelling along the rope.
  b If the wave speed is 1.7 m s^{-1}, calculate the wavelength produced by shaking the end seven times every 10 seconds.
  c If the rope was replaced by a thinner one, would you predict that the wave speed would increase, or decrease (under the same conditions)? Explain your answer.

3 Describe how the point P on the slinky spring shown in Figure C2.6 moves as the wave passes through it.

4 If you watch waves coming into a beach, you will notice that they get closer to each other.
  a State and explain how their wavelength is changing.
  b Suggest what has caused the waves to change speed.

5 After an earthquake, the first wave to reach a detector 925 km away arrived 149 s later. This type of wave is called a P wave (pressure wave).
  a Suggest whether this is a longitudinal or transverse wave.
  b Calculate the average speed of the wave (m s^{-1}).
  c Suggest why your calculation produces an ‘average’ speed.
  d If the wave had a period of 11.21 s, what was its wavelength?

LINKING QUESTION

- How can the length of a wave be determined using concepts from kinematics?

This question links to understandings in Topic A.1.

Tool 2: Technology

Generate data from models and simulations

Some time after a Primary (longitudinal) wave is received from an earthquake, a different kind of wave will be detected. This is called a Secondary (transverse) wave. If the delay between the detection of the two waves is measured and the speeds of both waves are known, the distance to the original earthquake can be determined.

Set up a spreadsheet that will calculate the distance to the source of an earthquake (dependant variable) for various time delays (independent variable). Assume speeds of waves are 5500 m s^{-1} and 3200 m s^{-1}.

Representing waves graphically

Waves can be represented by displacement–position or displacement–time graphs. They both have similar sinusoidal shapes.

Figure C2.8 shows how the displacements of particles (from their mean positions) vary with *distance* from a fixed point (position). x_0 is the amplitude of their oscillations. It may be considered as a ‘snapshot’ of the wave at one particular moment.

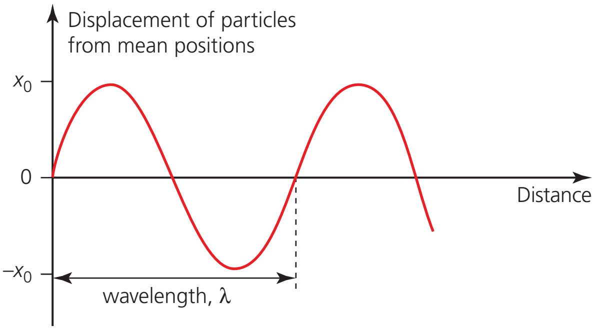


Figure C2.8 Displacement–distance graph for a wave

Figure C2.9 shows how the displacement of a certain particle (from its mean position) varies with time at one precise location. It could be considered as a video of that part of the medium.

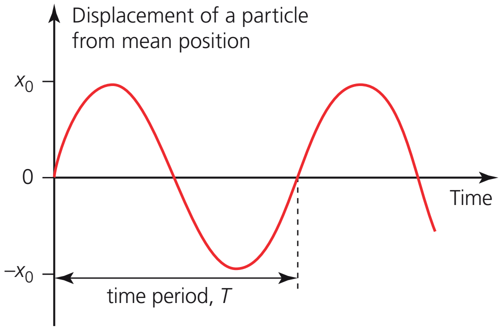


Figure C2.9 Displacement–time graph for a wave

Common mistake

Graphs like these can be used to represent *both* transverse and longitudinal waves. Because of their shape, it is a common mistake to think that they only represent transverse waves. The direction of the displacements shown on the vertical axes of these graphs are not specified, so they could be either

- in the direction of wave travel (longitudinal waves), or
- perpendicular to wave travel (transverse waves).

Pulses

A short duration oscillating disturbance passing through a medium may be described as a wave **pulse**. See Figure C2.10 for a simplified representation.

◆ **Pulse (wave)** A travelling wave of short duration.

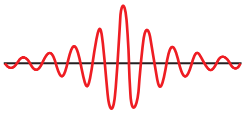


Figure C2.10 Wave pulse

6 Sketch a displacement–time graph for a transverse wave of frequency 4.0 Hz and an amplitude of 2.0 cm. Assume that the wave has its maximum positive displacement at time $t = 0$. Continue the graph for a duration of 0.5 s.

7 Figure C2.11 represents a longitudinal wave.

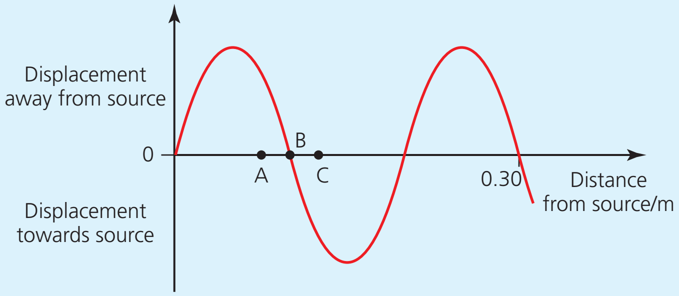


Figure C2.11 A longitudinal wave

a State its wavelength.

b Describe the instantaneous movement of a particle which is

  i at a distance A from the source

  ii at a distance C from the source.

c Is there a compression, a rarefaction, or neither, at position B?

8 A wave pulse is made on a water surface by touching it once with a fingertip. Sketch a possible displacement–position graph of the resulting disturbance spreading out on the surface.

Sound waves

SYLLABUS CONTENT

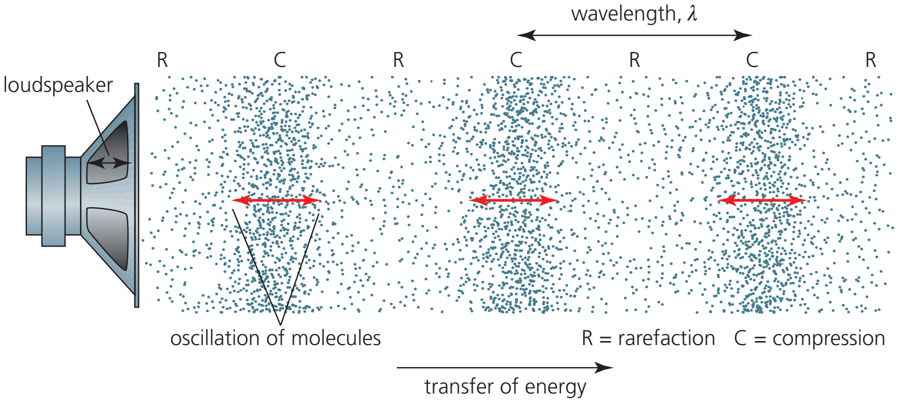
- The nature of sound waves.

◆ **Sound** Longitudinal waves in air or other media that are audible to humans.

◆ **Ultrasound** Frequencies of sound above the range that can be heard by humans (approximately 20 kHz).

A vibrating surface will disturb its surroundings and propagate longitudinal waves through the air. The human ear is capable of detecting this type of wave if the frequency falls within a certain range (approximately 20 Hz to 20 kHz). What we hear is called **sound**. Higher frequencies of the same type of wave, which we cannot hear, are called **ultrasound**. (Lower frequencies are called infrasound.)

Figure C2.12 shows how the surface of a loudspeaker can produce longitudinal waves in air. The random arrangement of molecules changes as the wave passes through the air. The compressions and rarefactions result in small periodic changes of air pressure.



■ **Figure C2.12** Arrangement of molecules in air as sound passes through

If the graphs shown in Figure C2.8 and C2.9 represented sound waves, the vertical axes could also be changed to represent variations of air pressure (above and below average air pressure).

Speed of sound

Sound is a mechanical wave involving oscillating particles and, as such, needs a medium to travel through. Sound cannot pass through a vacuum.

Generally, we would expect that sound will travel faster through a medium in which:

- the particles are closer together
- there are stronger forces between the particles.

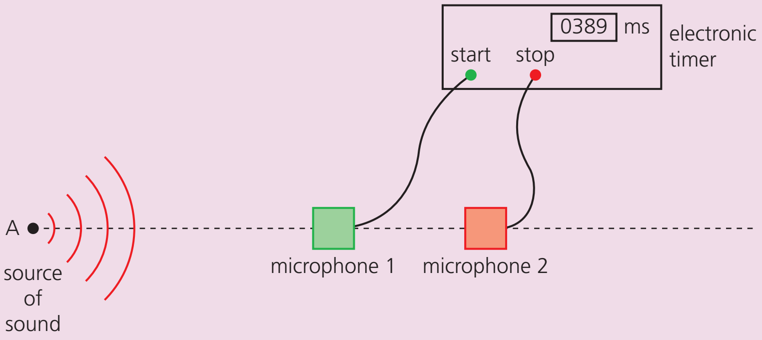
This means that sound usually travels faster in solids than liquids, and slowest in gases, such as air.

The speed of sound in the air around us increases slightly with temperature because then the molecules move faster.

Inquiry 1: Exploring and designing

Designing

Figure C2.13 shows an electronic method for determining the speed of sound.



■ **Figure C2.13** Laboratory experiment to determine the speed of sound

Design an experiment and a valid methodology using this apparatus to determine a value for the speed of sound. Suggest improvements to the design shown in Figure C2.13 so that the speed of sound in air is measured as accurately as possible.

Pitch and loudness of sound

The sounds that arrive at our ears at any one time usually include a range of waves with different frequencies and amplitudes. The oscillations are transferred to our eardrums and our brain interprets them as sounds of different **pitch** and **loudness** (volume).

Figure C2.14 shows pure sound waves of two different frequencies. We describe these sounds arriving at our ears as high pitched and low pitched.

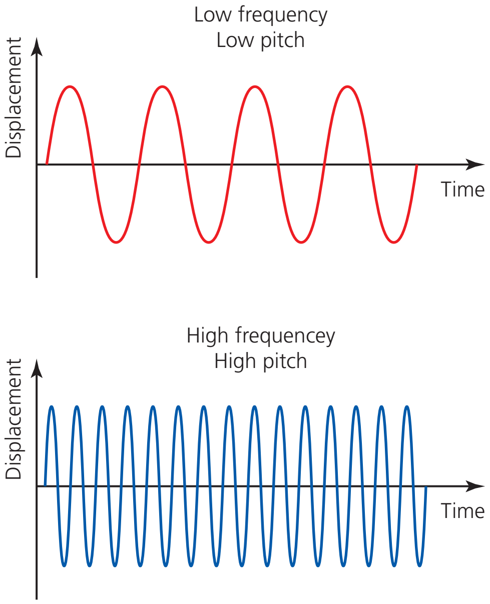
Figure C2.15 shows two waves of the same frequency, but different amplitudes. The wave of larger amplitude transfers more energy and we describe the effects of this as a louder sound.

◆ **Pitch** The sensation produced in the human brain by sound of a certain frequency.

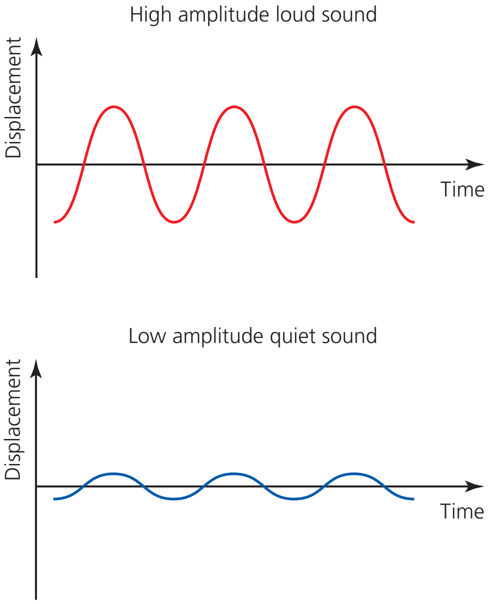
◆ **Loudness** A subjective measure of our ears' response to the level of sound received.

◆ **Logarithmic scale (on a graph)** Instead of equal divisions (for example, 1, 2, 3, …), with a logarithmic scale each division increases by a constant multiple (for example, 1, 10, 100, 1000 …).

◆ **Decibel** A measure of sound level.



■ **Figure C2.14** Sounds of different frequency / pitch



■ **Figure C2.15** Sounds of different amplitude / loudness

Tool 3: Mathematics

Logarithmic graphs and power laws

Sound intensity is the power that is carried perpendicularly by sound waves through unit area. It is easily measured by electronic meters, and apps for mobile devices are commonplace.

A normal human ear is capable of detecting sounds with a very wide range of intensities. This makes showing them all on a linear chart impossible. To get over this problem, we use a **logarithmic scale**. On a logarithmic scale (on a chart or a graph) each equal increment represents the fact that the quantity has been multiplied by the same factor (usually 10). As an example, we will consider the **decibel** scale. See Figure C2.16.

A student may wish to investigate the relationship between the intensity of sound (of a constant frequency) and the thickness of material placed between the source and the detector. The student may have no idea what this relationship will be.

Carry out calculations involving logarithmic and exponential functions

Sometimes there is no 'simple' relationship between two variables, or we may have no idea what the relationship may be. So, in general, we can write that the variables x and y are connected by a relationship of the form: $y = kx^p$, where k and p are unknown constants. That is, y is proportional to x to the power p.

Taking logarithms of this equation we get:

$$\log y = (p \times \log x) + \log k$$

Compare this to the equation for a straight line, $y = mx + c$.

If a graph is drawn of $\log y$ against $\log x$, it will have a gradient p and an intercept of $\log k$.

Using this information, a mathematical equation can be written to describe the relationship. Note that logarithms to the base 10 have been used in the above equation, but natural logarithms (ln) could be used instead.

The decibel scale is widely used to compare the intensity of a sound to a reference level. Each additional 10 on the scale represents an increase by a factor of 10 in sound intensity. So, for example, a sound of 50 dB intensity is 10× more intense than a sound of 40 dB. A sound of 60 dB intensity is 100× more intense than a sound of 40 dB, and so on.

Of course, sound intensities decrease with distances from their sources, which are not stated in Figure C2.16, so the numbers should be seen as just a rough guide.

Displaying all parts of the electromagnetic spectrum (later in this topic) is done with a logarithmic scale for the same reason.

Decibels	Example
0	Silence
10	Breathing, ticking watch
20	Rustling leaves, mosquito
30	Whisper
40	Light rain, computer hum
50	Quiet office, refrigerator
60	Normal conversation, air conditioner
70	Shower, toilet flush, dishwasher
80	City traffic, vacuum cleaner
90	Music in headphones, lawnmower
100	Motorcycle, hand drill
110	Rock concert
120	Thunder
130	Stadium crowd noise
140	Aircraft taking off
150	Fighter jet aircraft taking off
160	Gunshot
170	Fireworks
180	Rocket launch

■ **Figure C2.16** An approximate guide to sound levels in decibels

ATL C2A: Research skills

Evaluating information sources for accuracy, bias, credibility and relevance

Find three websites that enable you to check your hearing and follow their instructions. Compare the results and write a short review of your findings.

Were there any differences in the results for each website?

What might account for those differences?

Evaluate the sites in terms of their reliability.

WORKED EXAMPLE C2.2

a Calculate the wavelength of a sound of frequency 196 Hz if the speed of sound in air is $338\,\mathrm{m\,s^{-1}}$.

b If a longitudinal compression wave of the same frequency has a wavelength of 26.1 m in steel, determine the speed of the wave.

c Explain why the wave speed is greater in steel than in air.

Answer

a $\lambda = \dfrac{v}{f} = \dfrac{338}{196} = 1.72\,\mathrm{m}$

b $v = f\lambda = 196 \times 26.1 = 5.12 \times 10^3\,\mathrm{m\,s^{-1}}$

c Because the particles are closer together and there are stronger forces between them.

TOK

Knowledge and the knower

- What criteria can we use to distinguish between knowledge, belief and opinion?
- How do we distinguish claims that are contestable from claims that are not?
- How do our interactions with the material world shape our knowledge?

‘If a tree falls in a forest and no one is around to hear it, does it make a sound?

Figure C2.17 A fallen tree in a forest

This well-known philosophical question can be answered in different ways, depending on the perspective we take on what is meant by ‘sound.’

◆ **Audible range** Range of sound frequencies that can be heard by humans.

If we think of sound only as an effect in the human ear and brain, then the answer is clearly ‘no’, although there will still be longitudinal waves in the air. If we define sound as a hearable (**audible**) oscillation (regardless of whether anyone is there to hear it), then the answer is ‘yes’.

Consider how the knowledge questions above relate to this problem. You may also find the following guiding questions useful:

- Should we believe in things that we have not personally seen / observed / experienced?
- Can we assume that an unobserved event behaves in exactly the same way as an observed event?
- Does observation affect / change the event being observed?
- If the fall of a tree, and any consequential effects, are never observed, is this the same as saying that the tree never fell at all?

9 **a** Sketch a graph to show the air pressure variations (from normal) for a duration of 0.2 s at a certain point through which a sound wave of frequency 100 Hz is passing. Mark one time where there is a compression and one time where there is a rarefaction.

b Determine a value for the period of the wave and show it on the graph.

10 Outline an experiment using hand-held stopwatches to determine a value for the speed of sound in air.

11 Many people know that you can estimate the distance to a storm centre by counting the number of seconds between a flash of lightning and hearing the thunder: about one kilometre for every three seconds. Explain the physics behind this idea.

12 The ultrasound waves used in a medical scanner had a frequency of 9.6 MHz.

a If the wavelength was 0.16 mm, determine the speed of ultrasound waves in the body.

b Suggest three properties of ultrasound that make it useful for obtaining scans from inside the human body.

13 Figure C2.18 shows the use of ultrasound waves (**sonar**) to detect the depth of the ocean below a boat. Waves are produced in a **transducer** and a pulse is directed downwards. The transducer has a diameter of 3 cm. Some wave energy is reflected back from the seabed and then received and detected at the same transducer a short time later. The time delay is used to calculate the depth of the water.

- **a** If the speed of sound waves in sea water is 1520 m s^{-1}, calculate the depth of water if the delay between the pulses is 29 ms.
- **b** To limit the spreading of the waves emitted by the transducer it is required that wavelength is much smaller than the size of the transducer. Show that this is true if the waves have a frequency of 214 kHz.
- **c** Suggest why the system uses wave pulses rather than continuous waves.

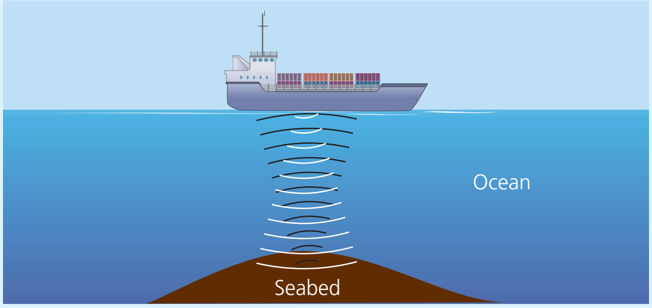


Figure C2.18 Boat using sonar

14 The speed of sound in helium gas is much greater than in air, which is mostly nitrogen (for the same temperature and pressure).
Use knowledge from Topic B.2 to discuss reasons for this difference.

◆ **Sonar** The use of reflected ultrasound waves to locate objects.

◆ **Transducer** Device that converts one form of energy to another. The word is most commonly used with devices that convert to or from changing electrical signals.

◆ **Vacuum** A space without any matter. Also called free space.

◆ **Free space** Place where there is no air (or other matter). Also called a vacuum.

Light waves

In Topic B.1 we described the range of thermal radiations (including light) emitted from various surfaces and 'black bodies' at different temperatures. The true nature of light was not discussed in B.1, but it was a major issue among scientists for hundreds of years.

In the seventeenth century, Isaac Newton believed that a beam of light consisted of particles ('corpuscles'), others thought that light could travel as waves. The wave nature of light was not demonstrated until 1801, when the English physicist Thomas Young showed that light could 'interfere'. This famous experiment and the nature of interference are explained and discussed in Topic C.3.

Light is a transverse electromagnetic wave but, unlike mechanical waves, it does not require a medium to travel through. Light can travel across a **vacuum**, sometimes called '**free space**'. Light travelling from the Sun, through space, to arrive at the Earth is an obvious example.

Visualizing the oscillations of light waves is more difficult than the models of *mechanical* waves that we discussed earlier in this topic. Figure C2.19 shows that light oscillations are high frequency periodic variations in the strength of electric and magnetic fields (which are perpendicular to each other). Electric and magnetic fields are discussed in Theme D.

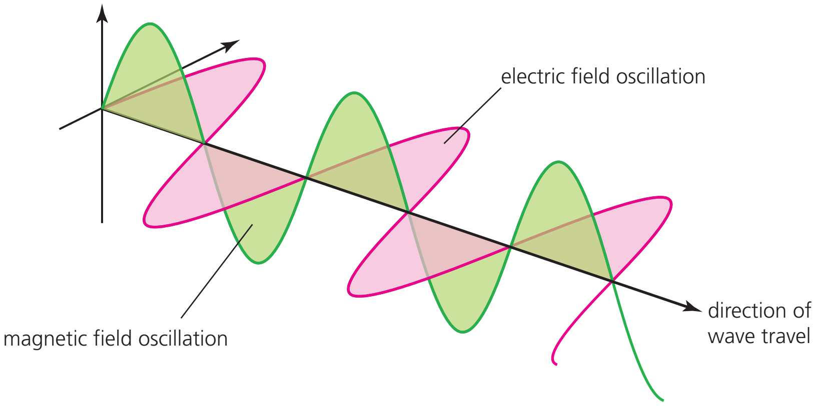


Figure C2.19 Light and other electromagnetic waves are combined electric and magnetic fields

■ **Figure C2.20** Spectrum of visible light

The speed of light in a vacuum is $3.00 \times 10^8\,\mathrm{m\,s^{-1}}$ (more accurately: $299\,792\,458\,\mathrm{m\,s^{-1}}$). It is given the unique symbol '*c*'.

In **transparent** materials light travels at slightly slower speeds. For example, light travels at almost the same speed in air ($299\,970\,500\,\mathrm{m\,s^{-1}}$) as in free space, but at $2.26 \times 10^8\,\mathrm{m\,s^{-1}}$ in water.

The **continuous spectrum** of visible **white light**, from red to violet, is a familiar sight (Figure C2.20). The different colours that we see are created by waves of different frequencies.

Red light has the lowest frequency, violet light has the highest frequency. 'White light' is not a precise scientific term, but it can be assumed to be the same as the light received in the black-body radiation from the Sun on a cloudless day.

◆ **Transparent** Describes a medium that transmits light without scattering or absorption.

◆ **Continuous spectrum** The components of radiation displayed in order of their wavelengths, frequencies or energies (plural: spectra).

◆ **White light** Light which contains all the colours of the visible spectrum with approximate equal intensity.

The fundamental property of a light wave is its *frequency*. If a light wave enters a different medium and then travels more slowly, its frequency cannot change, but its wavelength will decrease ($\lambda = v/f$). However, when we quote data for light waves, it is common to use wavelengths, rather than frequencies. This is because light wavelengths are easier to visualize and measure.

WORKED EXAMPLE C2.3

An orange light has a frequency of 4.96×10^{14} Hz. Determine its wavelength as it passes through

a air

b a type of glass in which the speed of light has reduced to $1.94 \times 10^8\,\mathrm{m\,s^{-1}}$.

Answer

a $\lambda = \frac{c}{f} = \frac{3.00 \times 10^8}{4.96 \times 10^{14}} = 6.05 \times 10^{-7}\,\mathrm{m}$

b $\lambda = \frac{v}{f} = \frac{1.94 \times 10^8}{4.96 \times 10^{14}} = 3.91 \times 10^{-7}\,\mathrm{m}$

Different animals, birds and insects are able to detect different ranges of frequencies. For example, bees are not good at detecting the colour red, but they are able to detect higher frequencies (ultraviolet).

Red light has the longest wavelength in the visible spectrum, approximately 7×10^{-7} m. Violet has the shortest wavelength, approximately 4×10^{-7} m.

Use data from the previous paragraphs.

15 a Calculate a typical value for the frequency of red light in air.

b What is the frequency of the same light in glass?

16 Estimate a value for the wavelength of yellow light:

a in air **b** in water.

17 The 'light year' is widely used as a unit of distance in astronomy.
How far does light travel (km) in free space in one year?

18 Briefly outline why light waves are described as electromagnetic waves.

TOK

Knowledge and the knower

- How do our interactions with the material world shape our knowledge?

Perception of colour

We may all agree that light waves have a certain frequency, and whether those waves can be detected in some way by a human eye. There is no ambiguity in that, and most people would agree on the 'colours of the rainbow'. However, how our brains process signals about the light waves detected by our eyes, and how we communicate our impressions of specific colours to other people can be problematic. 'That dress is green' can never be an indisputable scientific fact. 'Colour blindness' (see Figure C2.21) may be an unusual medical condition, but it highlights the fact that human brains can interpret signals in different ways.

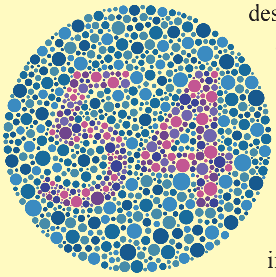


Figure C2.21 Test for colour blindness

Added to that, different people, societies and cultures are known to describe colours in different ways. If two people see, or describe, a colour differently, can one be 'right' and the other 'wrong'?

Finally, in terms of physics, it should be pointed out that if you say that a 'dress is green' you probably assume it is being seen under normal lighting conditions, with white light. The colour perceived will change if the lighting is changed. For example, if a red light was used, or it was seen through a yellow filter. Even looking at the dress at night under artificial lighting could change its appearance.

◆ **Ultraviolet** Part of the electromagnetic spectrum which has frequencies just greater than can be detected by human eyes.

◆ **Electromagnetic spectrum** Electromagnetic waves of all possible different frequencies, displayed in order.

◆ **Electromagnetic radiation** Waves which consist of combined oscillating electric and magnetic fields.

Electromagnetic waves

SYLLABUS CONTENT

- The nature of electromagnetic waves.

The extent of a visible spectrum such as that seen in Figure C2.20 is limited by:

- the inability of the human eye to detect higher or lower frequencies, and/or
- the ability of any particular source to produce a wider range of frequencies. Light is just a small part of a much wider continuous spectrum.

Just beyond the red end of the visible spectrum, there are waves which have longer wavelengths, called infrared. Infrared radiation was discussed in Topic B.1. Just beyond the violet end of the visible spectrum, there are waves of shorter wavelength, called **ultraviolet**.

The complete range (spectrum) of possible electromagnetic wavelengths extends from more than 100 000 km to less than 10^{-16} m. They have different origins and no single source produces all of these waves.

All electromagnetic waves travel at the same speed in vacuum, $c = 3.00 \times 10^8\,\mathrm{m\,s^{-1}}$.

They are all composed of oscillating electric and magnetic fields (Figure C2.19).

Together they are known as the **electromagnetic spectrum** (Figure C2.22). They are also described as **electromagnetic radiation**.

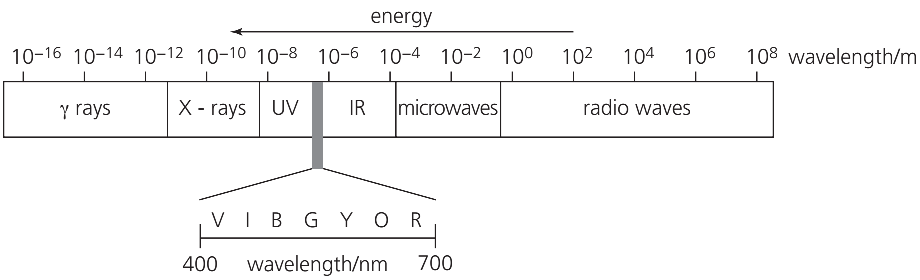


Figure C2.22 Electromagnetic spectrum

'Energy' refers to the energy carried by *photons* of the radiation, as explained later in the course.

Common mistake

Remember that the spectrum is continuous and the boundaries chosen between different named sections are somewhat arbitrary.

LINKING QUESTION

- How can light be modelled as an electromagnetic wave?

This question links to understandings in Topic D.2.

Inquiry 1: Exploring and designing

Exploring

Select sufficient and relevant sources of information

After deciding on a general area of interest, for an investigation you will often need to select and research other sources of information for background knowledge and any physics needed which is beyond the IB course (if appropriate). Your teachers should be an excellent source of advice and information and, obviously, the internet has multiple sources (of various quality). Physics books, science magazines and books from libraries can all be sources of information and inspiration.

Example 1: If you wish to investigate the effect that water vapour in the air has on the rate of evaporation from a water surface, you will need to learn about *humidity*.

Example 2: If you wish to investigate the world-wide use of solar heating of water, you will need to learn about the hours of sunlight in different locations, the variation in altitude of the Sun, comparative costs and so on.

Your intended investigation could be both interesting and unusual, but it needs to be realistic in terms of the apparatus that is available in your school, and the time available. So, it may be wise to check with teachers about whether an intended investigation is sensible under the circumstances.

Any sources of information should be acknowledged in the investigation report, including those which were researched but not used (with a reason given).

Task: Apart from sources on Earth, waves from all parts of the electromagnetic spectrum arrive at Earth from space. Use the internet to gather information about the origins of these waves and to what extent they are able to pass through the atmosphere and reach the Earth's surface.

The list in Table C2.1 shows some origins of electromagnetic waves and a selection of their uses.

Table C2.1 The different sections of the electromagnetic spectrum

Name	Typical wavelength / m	Origins (all are received from Outer Space)	Some common uses
radio waves	10^2	electronic circuits / aerials	communications, radio, television
microwaves	10^{-2}	electronic circuits / aerials	communications, mobile phones, ovens, radar
infrared (IR)	10^{-5}	everything emits IR but hotter objects emit much more than cooler objects	lasers, heating, cooking, medical treatments, remote controls
visible light	5×10^{-7}	very hot objects, light bulbs, the Sun	vision, lighting, lasers
ultraviolet (UV)	10^{-8}	the Sun, UV lamps	fluorescence
X-rays	10^{-11}	X-ray tubes	medical diagnosis and treatment, investigating the structure of matter
gamma rays	10^{-13}	radioactive materials	medical diagnosis and treatment, sterilization of medical equipment

LINKING QUESTION

- How are waves used in technology to improve society? (NOS)

This question links to understandings in Topics C.3, C.4, C.5, D.2, D.3 and D.4.

Top tip!

The fact that electromagnetic waves have some properties that could not be explained satisfactorily by their wave nature had very important consequences. A new 'particle' model for light, introduced at the start of the twentieth century, was the beginning of **quantum physics**. This is introduced in Topic E.2.

◆ **Quantum physics** Study of matter and energy at the subatomic scale. At this level quantities are quantized.

LINKING QUESTIONS

- How are electromagnetic waves able to travel through a vacuum?
- Can the wave model inform the understanding of quantum mechanics? (NOS)

These questions link to understandings in Topic E.2.

Nature of science: Experiments

Pure research

The first artificial electromagnetic (radio) waves

Heinrich Hertz (Figure C2.23) was the first to produce and detect artificial electromagnetic waves (1887 in Karlsruhe in Germany). He used high voltage electrical sparks. The electrical currents in sparks involve the necessary high frequency oscillating electric and magnetic fields. Although the distance involved was very small, it was the start of modern wireless communication. It was left to others (such as Guglielmo Marconi) to develop the technology for transmission over longer and longer distances – and then to design techniques to modify the amplitude, frequency or phase of the radio waves to transfer information, such as speech.

Figure C2.23 Heinrich Hertz

Tragically, Hertz died at the age of 36 in 1894. This was long before the far-reaching consequences of his discovery had been exploited.

Hertz had been trying to provide evidence for the electromagnetism theories of James Clerk Maxwell, and he has been widely quoted as saying that his discovery was ‘of no use whatsoever’. He was not alone in that opinion at the time.

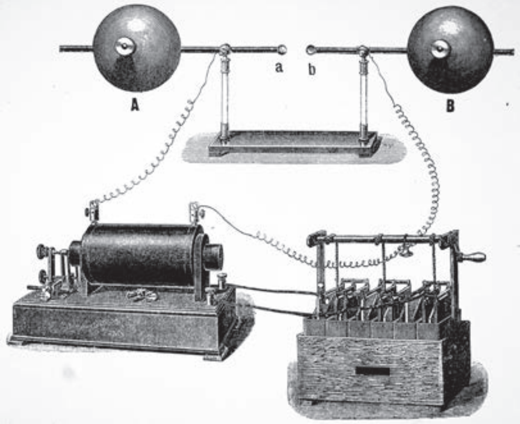


Figure C2.24 Hertz’s apparatus for the first artificial production of electromagnetic waves

‘Pure research’ is about extending knowledge and confirming theories, it is not about solving practical problems. But there are many historical examples of pure research leading to unexpected benefits of major significance – such as radio communication.

Of course, a large number of examples of pure research have *not* produced any worthwhile gains for society. An often-asked question is ‘should governments spend large amounts of money on open-ended research which has no obvious benefits (at that time)?’

In terms of laboratory investigations that you might carry out as a student: the common expectation is that they should have an ‘aim’, which may be answering a specific question. But maybe that is too restrictive?

19 Determine the frequency (in MHz) of a gamma ray which has a wavelength of 4.1×10^{-12} m.

20 A mobile phone network uses electromagnetic waves of frequency 1200 MHz.

- **a** Calculate their wavelength.
- **b** State which part of the electromagnetic spectrum contains these waves.
- **c** Use the internet to find out the frequency used in microwave ovens.
- **d** Suggest why our bodies are not warmed up by using mobile phones.

21 As you are reading this, which types of electromagnetic radiation are there in the room?

22 a State which types of electromagnetic radiation are considered to be dangerous.

- **b** What do they have in common?

23 Outline what properties of X-rays make them so useful in hospitals.

24 a Calculate how long it takes for a Bluetooth signal to travel from a mobile phone to a speaker which is 4.7 m away.

- **b** How much time (to the nearest minute) does it take light to reach the Earth from the Sun?
- **c i** How much time does it take a radio signal to travel to Mars from Earth?
  - **ii** Explain why your answer is uncertain. (Use the internet to obtain relevant data.)

C.3 Wave phenomena

Guiding questions

- How are observations of wave behaviours at a boundary between different media explained?
- How is the behaviour of waves passing through apertures represented?
- What happens when two waves meet at a point in space?

What are the basic behaviours of all waves?

- Reflection
- Refraction
- Diffraction
- Interference

Each of these properties will be discussed in this topic. But first we need to consider how we can represent travelling waves in two dimensions on paper, or screens.

◆ **Ripple tank** A tank of shallow water used for investigating wave properties.

◆ **Wavefront** A line connecting adjacent points moving in phase (for example, crests). Wavefronts are one wavelength apart and perpendicular to the rays that represent them.

Wavefronts and rays

SYLLABUS CONTENT

- Waves travelling in two and three dimensions can be described through the concepts of wavefronts and rays.

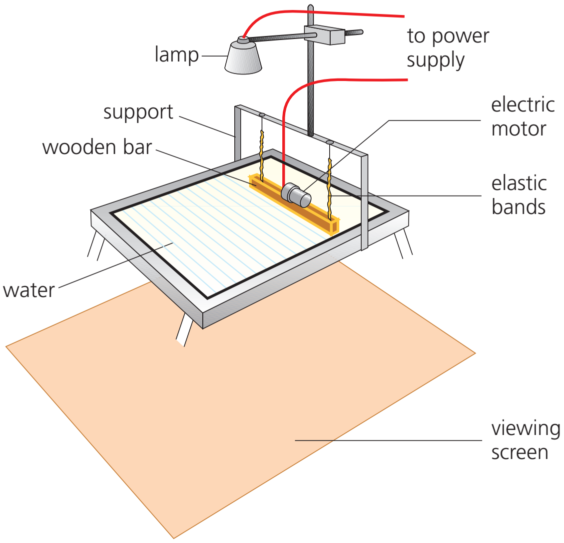


Figure C3.1 A ripple tank is used to investigate wave behaviour

Waves in two dimensions

Figure C3.1 shows a **ripple tank**: a common arrangement used to observe the behaviour of waves. Small waves (ripples) can be made by touching the surface of shallow water at a point, or with a wooden bar. Usually, a motor is attached to the bar to make it vibrate and produce continuous parallel waves at various frequencies. The light above enables the moving waves to be seen on the screen below the tank. (A *stroboscope* is often used to make the waves appear stationary.)

The 'waves' seen on the screen show the positions of wave *crests*. These lines are called **wavefronts**. They are one wavelength apart. More precisely:

A wavefront is a line joining neighbouring points moving in phase with each other.

The blue lines seen in Figure C3.2 represent the pattern produced by regular disturbances of the water at point P. The waves are spreading with equal speed in all directions (in two dimensions), so that the pattern is circular. Wave speed depends on the depth of the water, which is usually constant if the tank is horizontal.

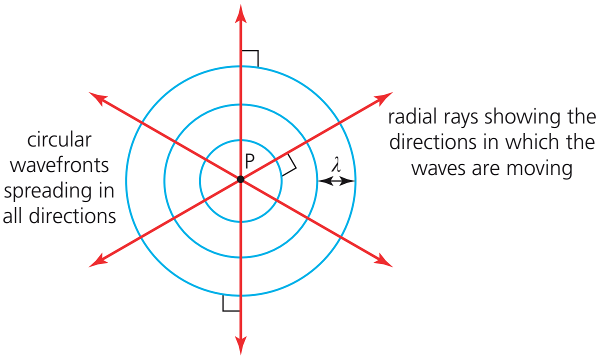
◆ **Ray** A line showing the direction in which a wave is transferring energy.

◆ **Radial** Diverging in straight lines from a point.

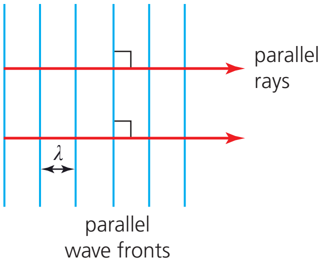
The red lines with arrows in Figure C3.2 are called **rays**. These rays can be described as **radial**, meaning that they are spreading out in straight lines from a point. Radial rays represent circular (or spherical) wavefronts.

Rays are lines showing the direction in which wavefronts are moving and energy is transferred. Rays and wavefronts are always perpendicular to each other.

The wavefronts that we will be considering in this course are either circular (or spherical), see Figure C3.2, or the other common possibility: parallel wavefronts, as shown in Figure C3.3. Parallel wavefronts can be made on the water surface in a ripple tank by using a wooden bar.



■ **Figure C3.2** Circular wavefronts and radial rays spreading from a point source



■ **Figure C3.3** Parallel wavefronts and parallel rays that are not spreading out

The movement of parallel wavefronts is represented by parallel rays.

Waves in three dimensions

◆ **Plane waves** Waves travelling in three dimensions with parallel wavefronts, which can be represented by parallel rays.

◆ **Visualization** Helping understanding by using images (mental or graphic).

Waves spreading from a point with constant speed in three dimensions can be represented by spherical wavefronts. If the source of waves is a very long way away (compared to the wavelength) the wavefronts will be (almost) straight and parallel. Such waves are described as **plane waves**. A common example: light waves from a distant point. The Sun is an obvious source of plane waves.

TOK

Knowledge and the knower, The natural sciences

- How do our interactions with the material world shape our knowledge?
- What is the role of imagination and intuition in the creation of hypotheses in the natural sciences?

Different ways of describing the same thing

We can use either the concept of wavefronts, or the concept of rays, to describe the movement of the same waves, but they are not normally used at the same time.

The use of wavefronts to describe waves on water surfaces, which are easily *visible*, is easily understood, but the associated concept of 'rays' seems unnecessary. However, when describing another wave phenomenon, the *invisible* passage of light through the air, why do we usually prefer the **visualization** of rays, although they have no physical reality? How might such visualizations extend, affect, or perhaps limit our understanding of the natural world?

1 a Describe and explain how parallel wavefronts in a ripple tank appear if they are moving perpendicularly into shallower water.

b Discuss how the wavefronts from a point source would appear if the tank was raised on one side. Explain your answer.

2 Figure C3.4 represents some light wavefronts passing from left to right through a lens.

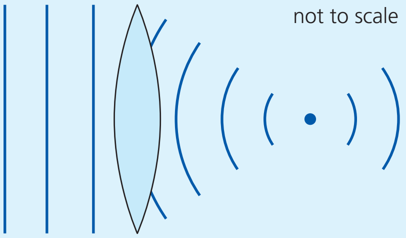


Figure C3.4 Light wavefronts passing from left to right through a lens

a Explain how you know that the source of light is a long way away.

b Make a sketch of the lens and show the path of five rays to represent the movement of the wavefronts.

c State which word we use to describe the effect of the lens on the light rays.

3 Figure C3.5 shows ocean waves as they approach the coast.
Suggest possible reasons why the separation and direction of the waves change.

Figure C3.5 Ocean waves refracting (and diffracting) as they approach a beach

Transmission, absorption and scattering of waves

SYLLABUS CONTENT

- Wave behaviour at boundaries in terms of transmission.

◆ **Transmission** Passage through a medium without absorption or scattering.

◆ **Opaque** Unable to transmit light (or other forms of energy).

The process in which waves are able to travel through a medium is described as wave **transmission**. For example, light and sound can be transmitted through air and water. During the transmission of waves, there is often *absorption* of energy: some or all of the wave energy is transferred to internal energy within the medium. Waves may also be randomly misdirected by interactions with irregularities within the medium. This is called scattering. Figure C3.6 illustrates these terms: waves are transmitted by medium 1, then enter medium 2, where they are each either transmitted, absorbed or scattered. In reality, all three processes can occur with the same waves.

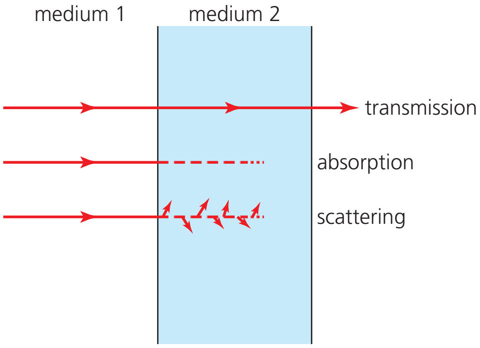


Figure C3.6 Transmission, absorption and scattering

A medium through which light can be transmitted, and through which we can see clearly, is described as being *transparent*. A medium through which light cannot be transmitted is described as **opaque**.

Wave power, intensity and amplitude

As waves spread out, and/or their energy is dissipated, the power that they transfer is reduced. We usually describe this as a reduction of wave intensity, a concept that was introduced in Topic B.1, and is defined again here:

$$\text{intensity, } I = \frac{P}{A} \quad \text{SI Unit: W m}^{-2}$$

Common mistake

Some books use the symbol A to represent amplitude, but this can cause confusion with the symbol for area. We will use the symbol A to represent area and the symbol x_0 to represent amplitude.

Speaking generally, we know that the energy transferred by a wave is proportional to its amplitude squared (Topic C.2). More precisely:

$$\text{intensity} \propto \text{amplitude}^2$$

Waves spreading from a point without absorption

There are reasons, discussed above, why waves may lose energy during transmission through a medium, but, if the waves were *spreading* out from a point source (that is, they are not plane waves), their intensity will decrease for that reason alone, without any absorption.

In two dimensions (surface waves)

As surface waves spread away from a point source the wavefronts will extend over greater and greater lengths. For example, if a circular wavefront increases its distance from its centre from r to $2r$, then its circumference will increase from $2\pi r$ to $4\pi r$. See Figure C3.7.

In each spreading wavefront the same amount of energy is spread over a greater length, so that the wave amplitude will decrease. If surface waves are a great distance from their source, the wavefronts will become almost parallel to each other, so that no loss of energy / power may be noticeable.

In three dimensions (such as light and sound)

We discussed the spreading of light and infrared waves from the Sun in Topic B.2. Refer back to Figure B2.1 in that chapter. The intensity of any waves spreading radially in three dimensions, without absorption, from a point source follows an inverse square relationship, which is repeated here:

$$\text{intensity, } I \propto \frac{1}{\text{distance}^2}$$

Figure C3.8 represents this kind of relationship in graphical form.

But note that, as before, if the waves are a great distance from their point source, the wavefronts will become almost parallel / plane waves, so that no loss of energy/power due to spreading may be noticeable.

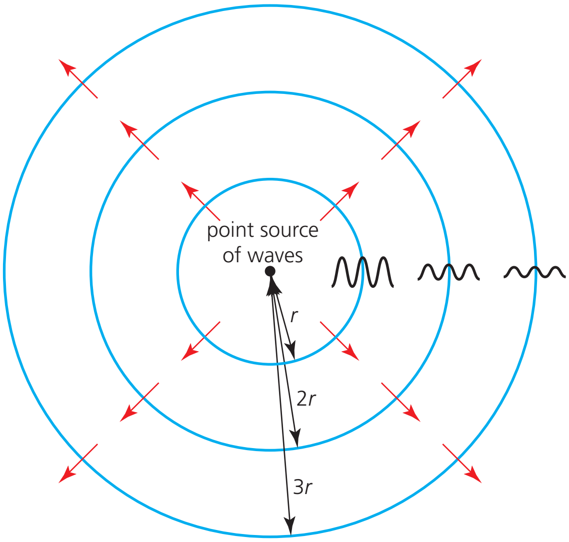


Figure C3.7 Amplitude of circular wavefronts decreases with the distance from source

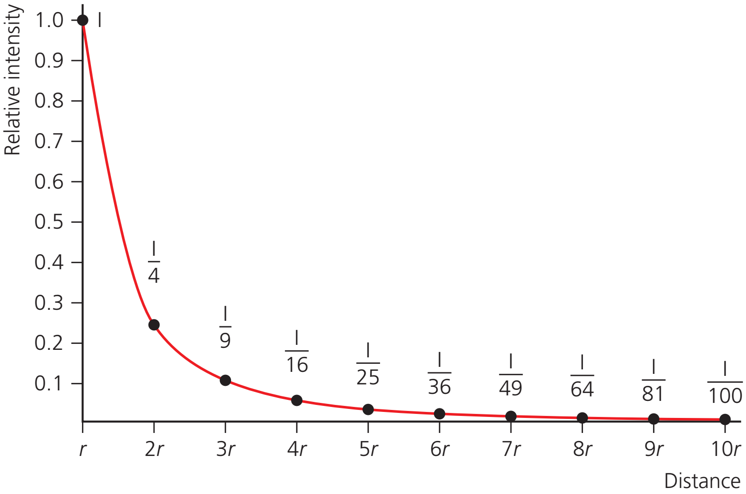


Figure C3.8 Inverse square relationship

Tool 3: Mathematics

Plot linear and non-linear graphs showing the relationship between two variables

Check the data shown in Table C3.1 to determine if there is an inverse square relationship between x and y:

- numerically
- graphically.

Table C3.1

x	y
1.34	9
0.96	17
0.81	24
0.70	32
0.64	38
0.59	45
0.55	52
0.52	58
0.49	66

4 Light and infrared radiation arriving perpendicularly on a solar panel have a total intensity of 780 W m^{-2}.

- **a** If the panel has dimensions of 50 × 80 cm, calculate the incident power.
- **b** Explain why this power will change during the course of the day.

5 A girl is reading a book at night using the light from a single lamp, which may be assumed to be a point source. If the lamp was originally 1.80 m away from the book, show that the intensity of the light on the book doubles if it is moved 0.53 m closer to the lamp.

6 Figure C3.9 shows some typical aerials used for transmitting (and receiving) signals to mobile phones.

Figure C3.9 Typical aerials for transmitting mobile phone signals

- **a** Explain why it desirable that the waves from these aerials are *not* emitted equally in all directions.
- **b** Suggest which feature of the aerials limits the vertical spreading of the waves.

7 The radiation from the Sun which reaches the top of the Earth's atmosphere has an intensity of 1360 W m^{-2} (see Topic B.1). It is approximately 40% visible light, 50% infrared and 10% ultraviolet.
The intensity reaching the Earth's surface is approximately 1000 W m^{-2}.
The approximate proportions reaching the Earth's surface are 44% visible light, 53% infrared and 3% ultraviolet.
Use this data to estimate the percentages of these three radiations which are:

- **a** transmitted by the Earth's atmosphere
- **b** absorbed / scattered by the Earth's atmosphere.

8 Why is the sky blue? (Research on the internet if necessary.)

9 It is considered to be a health risk to expose our ears to sounds of intensity greater than 10 mW m^{-2} for more than a few minutes.

- **a** Calculate the total power received on an eardrum of area 0.48 cm^2 from this intensity.
- **b** If the sound intensity 2.10 m from a loudspeaker at a rock concert was 0.44 W m^{-2}, estimate how far away you would need to be in order to reduce the intensity to 100 mW m^{-2}.
- **c** State two assumptions that you made in answering **b**. Discuss whether these assumptions are reasonable.

Reflection of waves and rays

◆ **Reflection (waves)** Change of direction that occurs when waves meet a boundary between two media such that the waves return into the medium from which they came.

SYLLABUS CONTENT

- ▶ Wave behaviour at boundaries in terms of reflection.

When a wave meets a boundary between different media some, or all, of the wave energy will be re-directed back into the first medium. This is called **reflection**.

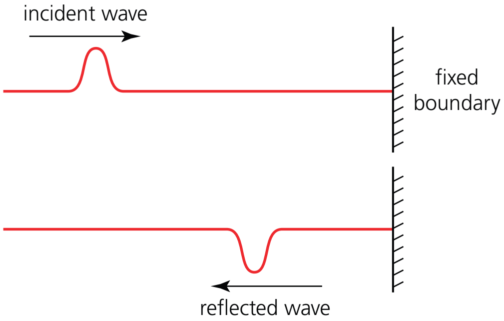


Figure C3.10 Reflection of a pulse off a fixed boundary

To develop understanding, we will first consider the simplest possible example: a single wave pulse on a rope or string, being reflected at a fixed boundary, as shown in Figure C3.10. A wave travelling towards a boundary is called an *incident wave*. The reflected wave is inverted from this type of boundary: there is a phase change of $\pi/2$.

If the incident waves are continuous, they may combine with the reflected waves to produce a standing wave, as discussed in the next topic, C.4.

Apart from fixed boundaries, where there is no possibility of transmission, waves may also reflect from a boundary where some transmission occurs. Usually, the wave will have different speeds in the two different media.

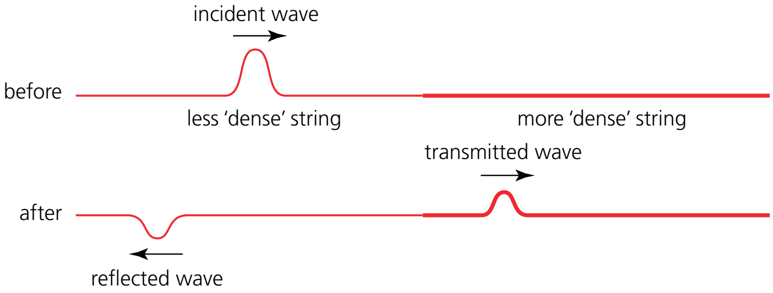


Figure C3.11 A pulse travelling into a 'denser' medium

Figure C3.13 Light reflected off and being transmitted by a window

In Figure C3.11 the 'denser' rope has a greater mass per unit length, so that the wave travels more slowly through it. The reflected wave is still inverted. The energy is shared between the reflected wave and the transmitted wave, so that both amplitudes are less than that of the incident wave.

Figure C3.12 shows the situation in which a wave pulse meets a boundary with a medium in which its speed would increase. Note that there is no phase change.

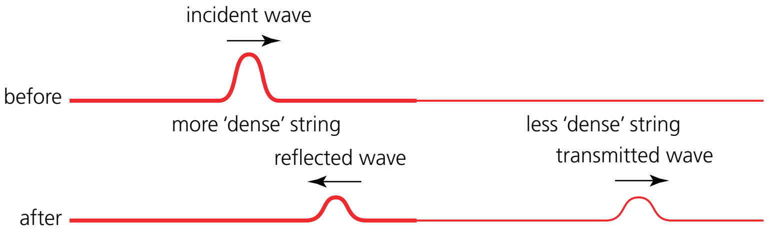


Figure C3.12 Longitudinal waves and pulses behave in a similar way to transverse waves

We will now extend the discussion of reflection to two and three dimensions.

Top tip!

When drawing wavefronts diagrams, make sure that they are *continuous* at the boundaries.

◆ **Ray diagrams** Drawings that represent the directions of different waves or particles as they pass through a system.

◆ **Incidence, angle** The angle between an incident ray and the normal.

◆ **Angle of reflection (rays)** Angle between a reflected ray and the normal.

◆ **Reflection, law of** Angle of incidence = angle of reflection.

Reflected wavefronts

When *parallel* wavefronts meet a plane (flat) boundary, some or all of them will be reflected so that the angle that the incident wavefront makes with the boundary is equal to the angle that the reflected wavefront makes with the boundary. See Figure C3.14.

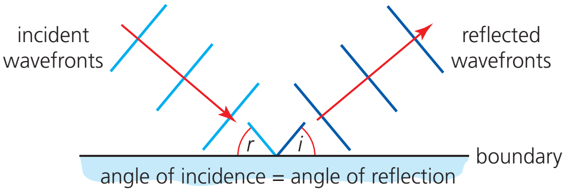


Figure C3.14 Reflection of parallel wavefronts from a straight boundary

When discussing the reflection of light, instead of wavefronts, it is more common to refer to rays, as shown in the **ray diagram** of Figure C3.15. A normal is a line perpendicular to the surface (at the point of incidence). The **angle of incidence** and the **angle of reflection** are measured between the ray and the normal (not the surface).

The **law of reflection**:

> The angle of incidence equals the angle of reflection.

Figure C3.16 shows radial light rays spreading from a point source (an 'object'). When they strike the plane mirror, the law of reflection can be used to determine the directions of the reflected rays. An eye looking into the mirror will see an image located as far behind the mirror as the object is in front.

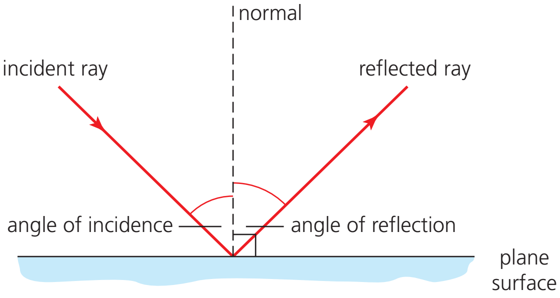


Figure C3.15 Reflection of rays from a plane surface

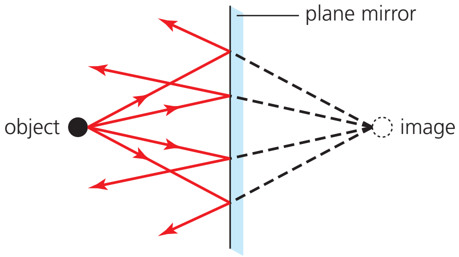


Figure C3.16 Image formed by reflection of rays in a plane mirror

Figure C3.17 represents the same situation using wavefronts instead of rays. The reflected waves appear to come from a point as far behind the reflecting surface as the actual source of the waves (the 'object') is in front.

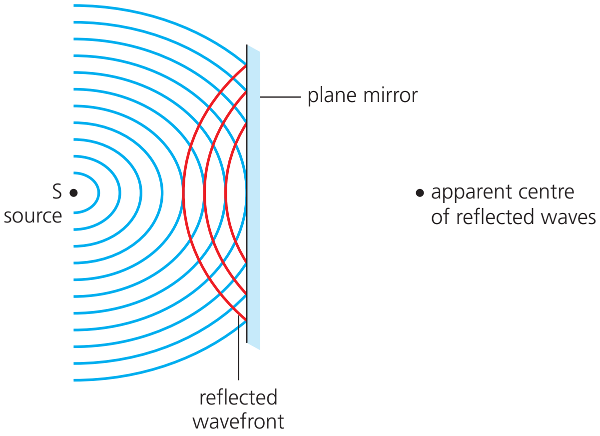


Figure C3.17 Reflection of circular wavefronts by a plane surface.

10 Figure C3.18 shows an idealized pulse on a string approaching a fixed end at a speed of $100\,cm\,s^{-1}$. Draw a sketch to show the position of this pulse 0.8 s later.

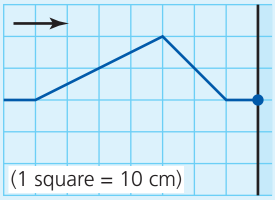


Figure C3.18 An idealized pulse on a string

11 Figure C3.19 shows a man looking at a wall. Make a rough copy and add to the wall the smallest mirror that would enable the man to see both the top of his head and his feet. Include light rays that explain your positioning of the mirror.

12 Predict if there will be a phase change when light waves reflect off a glass surface. Explain your answer.

13 Figure C3.20 shows light rays from the Sun being reflected to a focus. Describe the shape of the wavefronts

a arriving from the Sun

b being reflected to the focus.

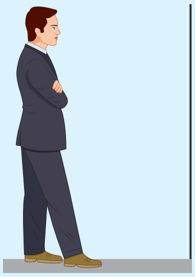


Figure C3.19 A man looking at a wall

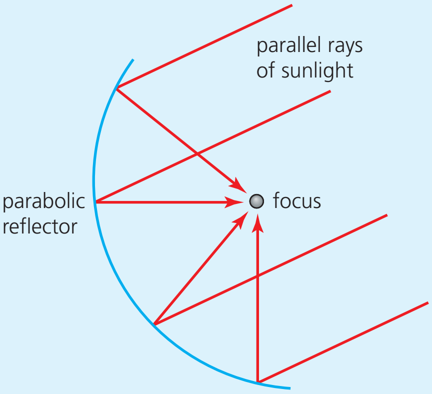


Figure C3.20 Light rays from the Sun being reflected to a focus

Inquiry 1: Exploring and designing

Exploring

Sound reflections in large rooms

Sound reflects well off hard surfaces like walls, whereas soft surfaces, such as curtains, carpets, cushions and clothes, tend to absorb sound. A sound that reaches our ears may be quite different from the sound that was emitted from the source because of the many and various reflections it may have undergone. Because of this, singing in the shower will sound very different from singing outdoors or singing in a furnished room. In a large room designed for listening to music (such as an auditorium, Figure C3.21), sounds travel a long way between reflections.

Figure C3.21 A large auditorium designed for effective transfer of sound to the audience

Since it is the reflections that are responsible for most of the absorption of the sound waves, it will take a longer time for a particular sound to fall to a level that we cannot hear. This effect is called *reverberation*. The longer reverberation times of bigger rooms mean that a listener may still be able to hear reverberation from a previous sound at the same time as a new sound is received. That is, there will be some 'overlapping' of sounds. Reflections of sounds off the walls, floor and ceiling are also an important factor when music is being produced in a recording studio, although some effects can be added or removed electronically after the original sound has been recorded.

Does your school or college have a performance space, such as a hall or a theatre, where you can make sounds and listen to them carefully when they arrive back at your ears? Or is there a performance space nearby you could visit?

1. Research online using search terms such as 'sound reverberation' to find out how professional designers adapt and design performance spaces to change reverberation. Then investigate your chosen performance space, inspecting it for installations which affect sound reverberation.
2. Discuss and suggest what measurements you could make to test the reverberation in the space.
3. Are there any improvements that could be made? State these and explain your reasoning.

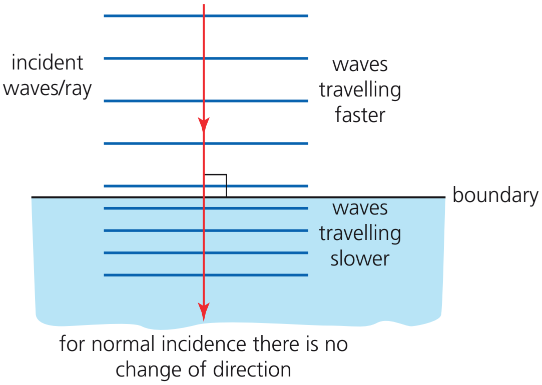


Figure C3.22 Waves slowing down as they enter a different medium

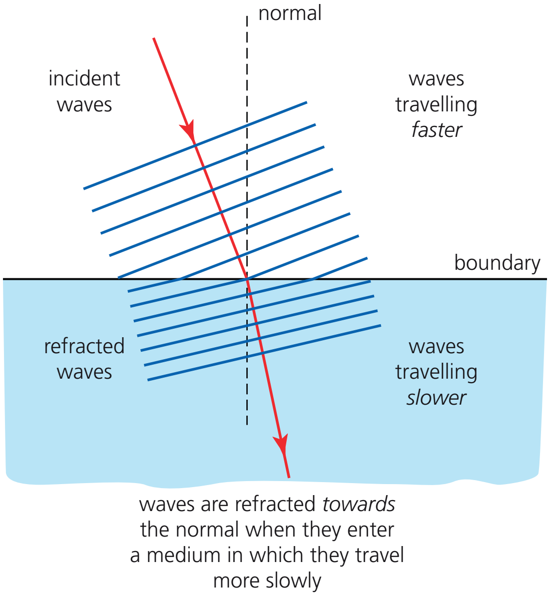


Figure C3.23 Waves refracting as they enter a denser medium

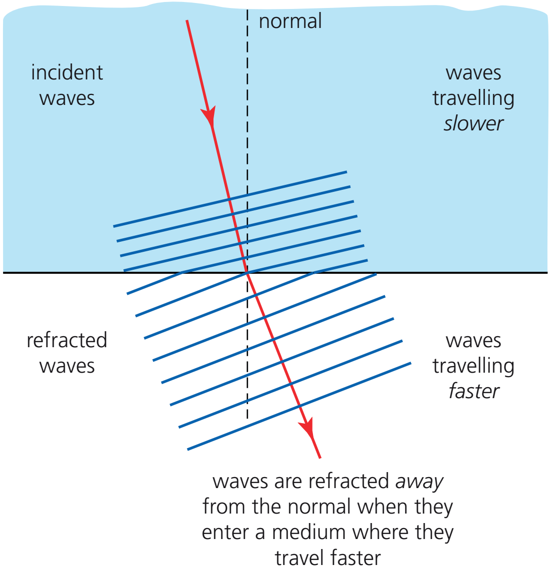


Figure C3.24 Waves refracting as they enter a less dense medium

Refraction of waves

SYLLABUS CONTENT

- Wave behaviour at boundaries in terms of refraction.
- Wavefront-ray diagrams showing refraction.
- Snell's law, critical angle and total internal reflection.
- Snell's law as given by:

$$\frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1} = \frac{v_2}{v_1}$$

where n is the refractive index and θ is the angle between the normal and the ray.

Waves will usually change speed when they travel into a different medium. Such changes of speed may result in a change of direction of the wave.

The speed of waves on the surface of water generally decreases as they move into shallower water.

Figure C3.22 shows parallel wavefronts arriving at a medium in which they travel more slowly. The wavefronts are parallel to the boundary and the ray representing the wave motion is perpendicular to the boundary.

In this case there is no change of direction but because the waves are travelling more slowly, their wavelength decreases, although their frequency is unchanged (consider $v = f\lambda$). Now consider what happens if the wavefronts are not parallel to the boundary, as in Figure C3.23.

Different parts of the same wavefront reach the boundary at different times and consequently, they change speed at different times. There is a resulting change of direction which is called **refraction**. The greater the change of speed, the greater the change of direction.

When waves enter a medium in which they travel more slowly, they are refracted towards the normal (assuming that the wavefronts are not parallel to the boundary).

Conversely, when waves enter a medium in which they travel faster, they are refracted away from the normal. This is shown in Figure C3.24; note that this is similar to Figure C3.23, but with the waves travelling in the opposite direction.

The refraction of light is a familiar topic in the study of physics, especially in optics work on lenses and **prisms**, but all waves tend to refract when their speed changes. Often this is a sudden change at a boundary between media, but it can also be a gradual, or irregular change. In Figure C3.25 differences in gas density produce irregular refraction and a blurred image above the fire. Stars twinkle in the night sky because of refraction in a shifting atmosphere.

◆ **Refraction** Change of direction that can occur when a wave changes speed (most commonly when light passes through a boundary between two different media).

◆ **Prism** A regularly shaped piece of transparent material (such as glass) with flat surfaces, which is used to refract and disperse light.

Figure C3.25 Irregular refraction over a fire

Refractive index

The amount of refraction that occurs when a light wave passes from one medium into another depends on the change of speed involved. This is represented numerically by the **refractive index, *n*,** of a medium:

$$\text{refractive index of a medium} = \frac{\text{speed of light in vacuum}}{\text{speed of light in the medium}}$$

$$n = \frac{c}{v}$$

Refractive index is a ratio of speeds, so it does not have a unit.

For example, the speed of light in water is 2.26×10^8, so that:

$$\text{refractive index of water} = \frac{3.00 \times 10^8}{2.26 \times 10^8} = 1.33$$

◆ **Refractive index, *n*** The ratio of the speed of waves in vacuum (or air) to the speed of waves in a given medium.

In practice, we will not usually be concerned with light travelling into, or out of a vacuum, but the speed of light in air is almost identical to the speed in a vacuum. This means that we can use the same refractive indices for light passing from air into, or out of, a medium.

Table C3.2 Refractive indices

Medium	Speed of light / $10^8\,\text{m}\,\text{s}^{-1}$	Refractive index
diamond	1.2	2.4
glass	1.8–2.0	1.5–1.7
plastic	1.9–2.3	1.3–1.6
lens in human eye	2.1	1.4
pure water	2.26	1.3
air	2.997	1.0
vacuum	2.998	-

As has been stated in Topic C.2, waves from all parts of the electromagnetic spectrum travel at exactly the same speed ($c = 3.00 \times 10^8\,\text{m}\,\text{s}^{-1}$) in free space (vacuum), but they all travel slower in other media. However, there are also some very small differences in the wave speeds in the same medium, for example, yellow light travels *very slightly* faster than green light in glass. For many applications this is not important, but it can result in the *dispersion* of white light into a spectrum (see later).

Snell's law

When we observe the refraction of light, the only direct measurements that can be made are of the angles involved. Figure C3.26 shows the paths of rays that could be seen in a standard experiment with a parallel-sided glass, or plastic, block. Two measurements are possible: θ_1, the angle that the incident ray makes with the normal and θ_2, the angle that the refracted ray makes with the normal. The rays are refracted *towards* the normal because the speed of light in glass is less than the speed of light in air. Because the block is parallel-sided, the same angles occur again as the ray emerges from the block. A ray of light entering the lower surface (as shown) will follow exactly the same path as a ray entering the upper surface, but in the opposite direction.

Some of the incident light will be reflected off the glass surface, but this has not been shown in the diagram.

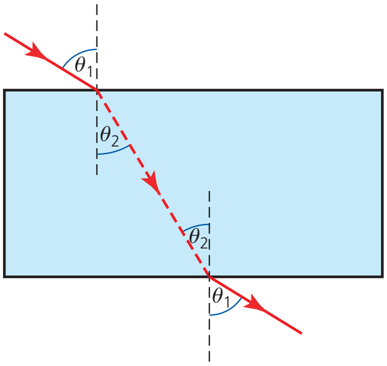
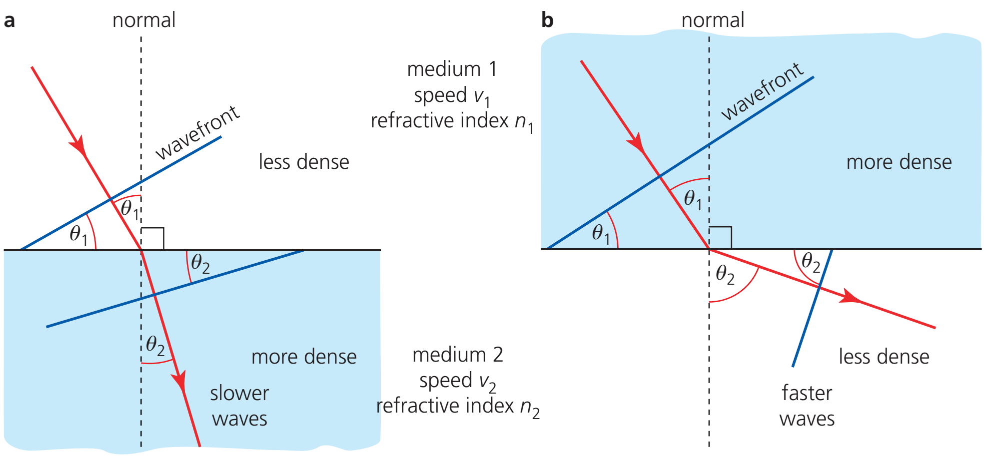


Figure C3.26 Light passing through a parallel-sided transparent block.

Figure C3.27 shows the situation in a little more detail and includes wavefronts.



■ **Figure C3.27** Light rays being refracted **a** towards the normal and **b** away from the normal

The Dutch scientist, Willebrord Snellius, was the first to show how the angles were related to the speeds of the waves and the refractive indices, as light passes from a medium of refractive index n_1, where its speed is v_1, to a medium of refractive index n_2, where the wave speed is v_2 (as seen in Figure C3.26).

Snell's law:

$$\frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1} = \frac{v_2}{v_1}$$

◆ **Snell's law (of refraction)** Connects the sines of the angles of incidence and refraction to the refractive indices in the two media (or the wave speeds). $\frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1} = \frac{v_2}{v_1}$.

If medium 1 is air, this reduces to:

$$\text{refractive index of medium 2, } n_2 = \frac{\sin\theta_1}{\sin\theta_2} = \frac{c}{v_2}$$

For light entering from air, refractive index, n = sine of the angle of incidence divided by the sine of the angle of refraction.

WORKED EXAMPLE C3.1

Consider Figure C3.26. If the angle of incidence on a plastic block was 48°, and the angle of refraction was 32°:

a determine the refractive index of the plastic, and

b calculate the speed of light in the plastic.

Answer

a $n_{\text{plastic}} = \frac{\sin\theta_1}{\sin\theta_2} = \frac{\sin 48}{\sin 32} = \frac{0.743}{0.530} = 1.40$

b $n_{\text{plastic}} = 1.40 = \frac{c}{v_{\text{plastic}}} = \frac{3.00 \times 10^8}{v_{\text{plastic}}}$

$v_{\text{plastic}} = 2.14 \times 10^8\,\text{m}\,\text{s}^{-1}$

WORKED EXAMPLE C3.2

A light ray travelling in water of refractive index 1.33 is incident upon a plane glass surface at an angle of 27° to the normal.

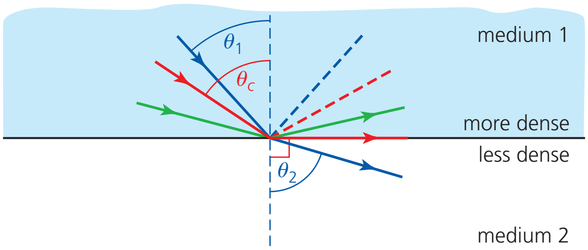
Calculate the angle of refraction if the glass has a refractive index of 1.63.

Answer

$\frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1} \Rightarrow \frac{1.33}{1.63} = \frac{\sin\theta_2}{\sin 27°}$

$\sin\theta_2 = 0.3704$

$\theta_2 = 22°$



■ **Figure C3.28** Total internal reflection occurs if the angle of incidence is greater than the critical angle θ_c

Critical angle and total internal reflection

Consider again Figure C3.24, which shows a wave / ray entering an optically less dense medium (a medium in which light travels faster). If the angle of incidence is gradually increased, the refracted ray will get closer and closer to the boundary between the two media. At a certain angle, the refracted ray will be refracted at an angle of exactly 90º along the boundary (see Figure C3.28). This angle is called the **critical angle**, θ_c, shown in red in the diagram.

For any angle of incidence some light will be reflected at the boundary, but for angles of incidence greater than the critical angle, *all* the light will be reflected back and remain in the denser medium. This is known as **total internal reflection**.

◆ **Critical angle** Largest angle at which a ray of light can strike a boundary with another medium of lower refractive index, without being totally internally reflected.

◆ **Total internal reflection** All waves are reflected back within the medium. Can only occur when a wave meets a boundary with another medium with a lower refractive index (in which it would travel faster).

◆ **Optically dense** If light travels slower in medium A, compared to medium B, then medium A is described as more optically dense.

We know that:

$$\frac{n_1}{n_2} = \frac{\sin\theta_2}{\sin\theta_1}$$

but at the critical angle, $\theta_1 = \theta_c$ and $\theta_2 = 90°$, so that $\sin\theta_2 = 1$, and then:

$$\frac{n_1}{n_2} = \frac{1}{\sin\theta_c}$$

Most commonly, the light will be passing from an **optically denser** material (medium 1) like glass, plastic or water, into air (medium 2), so that $n_2 = n_{\text{air}} = 1$, and so:

$$n_{\text{medium}} = \frac{1}{\sin\theta_c}$$

WORKED EXAMPLE C3.3

a Determine the critical angle for a water / air boundary. ($n_{\text{water}} = 1.33$)

b Determine the critical angle for a glass / water boundary. ($n_{\text{glass}} = 1.60$)

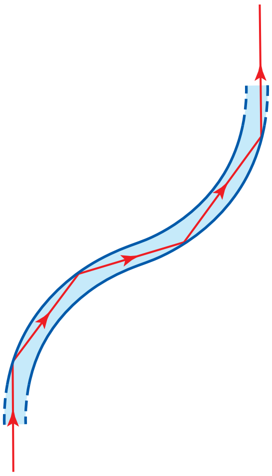
Answer

a $n_{\text{medium}} = \frac{1}{\sin\theta_c} \Rightarrow 1.33 = \frac{1}{\sin\theta_c} \Rightarrow \theta_c = 48.8°$

b $\frac{n_{\text{glass}}}{n_{\text{water}}} = \frac{1}{\sin\theta_c} \Rightarrow \frac{1.60}{1.33} = \frac{1}{\sin\theta_c} \Rightarrow \theta_c = 56.2°$

Applications of total internal reflection

One very important application of total internal reflection is in digital communication. Light passing into a glass fibre can be ‘trapped’ within the fibre because of multiple internal reflections and it will then be able to travel long distances, following the shape of the fibre (see Figure C3.29). The light can be modified to transmit digital information very efficiently.



■ **Figure C3.29** Total internal reflection along a glass fibre

ATL C3A: Communication skills

Clearly communicate complex ideas in response to open-ended questions

Most of the data transferred around the world is done using **optical fibres**. Choose one aspect of this important topic and use a variety of sources to access enough information that you can make an interesting three- to five-minute presentation to the rest of your group.

For example, you could choose one of the following: the choice of wavelength used, the use of binary signals, underwater cables, the *cladding* used in the fibres, the purity of the glass used, possible *bandwidths*, how far the signals can travel without the need for *regeneration*, how optical fibres are connected to copper wires, and so on.

◆ **Optical fibre** Thin, flexible fibre of high-quality glass that uses total internal reflection to transmit light along curved paths and over large distances.

Total internal reflection is also used in *endoscopes* for carrying out medical examinations. Light from a source outside is sent along fibres to illuminate the inside of the body. Other optical fibres, with lenses at each end, are used to bring a focused image outside for viewing directly, or via a camera and monitor (see Figure C3.30).

Figure C3.30 An endoscope can be used to inspect a patient's stomach

14 A ray of light travelling in air was incident at an angle of 47° to the surface of a plastic which had a refractive index of 1.58 Calculate:

- **a** the angle of refraction (to the normal) in the plastic
- **b** the speed of light in the plastic.

15 Parallel water waves travelling at $48\,\mathrm{cm\,s^{-1}}$ enter a region of shallow water with the incident wavefronts making an angle of 34° with the boundary. If the waves travel with a speed of $39\,\mathrm{cm\,s^{-1}}$ in the shallower water, predict the direction in which they will move.

16 Light rays travel at $2.23 \times 10^8\,\mathrm{m\,s^{-1}}$ in a liquid.

- **a** Determine the refractive index of the liquid.
- **b** Light rays coming out of the liquid into air meet the surface at an angle of incidence of 25°.
  Calculate the angle of the emerging ray to the normal in air.

17 A certain kind of glass has a refractive index of 1.55. If light passes into the glass from water (refractive index = 1.33) and makes an angle of refraction of 42°, what was the angle of incidence?

18 **a** Use trigonometry to show that the refractive index between two media is equal to the ratio of wave speeds (v_1/v_2) in the media.

- **b** Show that the refractive index for waves going from medium 1 into medium 2 is given by:
  ${}_1n_2 = n_2/n_1$

19 Explain why it is impossible for any medium to have a refractive index of less than one.

20 The refractive index of red light in a certain type of glass is 1.513.

- **a** If a ray of red light strikes an air / glass boundary at an angle of incidence of 29.0°, determine its angle of refraction.
- **b** A ray of violet light was incident at exactly the same angle (29.0°), but its angle of refraction was slightly less. Explain why.
- **c** If the angle of refraction for violet light was 18.5°, determine values for
  - **i** the refractive index of violet light in this glass
  - **ii** the speed of violet light in the glass.
- **d** We say that the red and violet light rays have been dispersed. Explain what that means. (See also next section.)

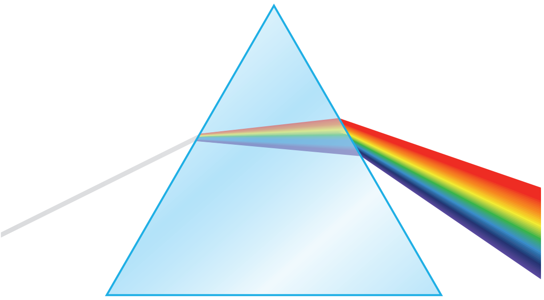
21 The speed of light in sea water is $2.21 \times 10^8\,\mathrm{m\,s^{-1}}$. Calculate the critical angle for light striking a boundary between sea water and air.

22 A certain kind of glass has a refractive index of 1.54 and water has a refractive index of 1.33.

- **a** In which medium does light travel faster?
- **b** Describe the circumstances which must occur for light to be totally internally reflected when meeting a boundary between these two substances.
- **c** Calculate the critical angle for light passing between these two media

Dispersion of light into a spectrum

The speeds of different colours (frequencies) of light in a particular medium (glass, for example) are not exactly the same. Red light travels the fastest and violet is the slowest. This means that different colours travelling in the same direction from the same source will not travel along exactly the same paths when they are refracted. When light goes through parallel-sided glass (like a window), the effect is not usually significant or noticeable. However, when white light passes into and out of other shapes of glass (like prisms and lenses), or water droplets, it can be **dispersed** (separated into different colours which spread apart). A triangular prism, as shown in Figure C3.31, is commonly used to disperse white light into a spectrum.



■ **Figure C3.31** A triangular prism used to produce a continuous spectrum of white light

◆ **Dispersion (light)** Separation into different wavelengths / colours (to form a spectrum, for example).

Diffraction of waves

SYLLABUS CONTENT

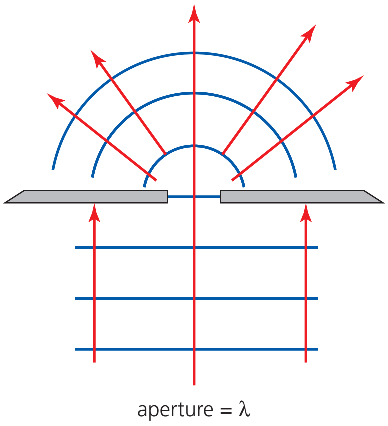
- ▶ Wave diffraction around a body and through an aperture.
- ▶ Wavefront-ray diagrams showing diffraction.

Waves of all types often encounter obstacles in their path, so it is important to understand how waves pass around (and through) such objects.

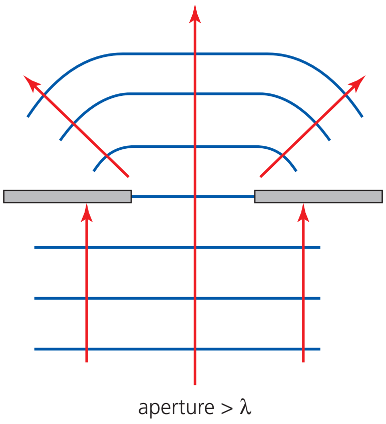
◆ **Aperture** A hole or gap designed to restrict the width of a beam of light (or other radiation).

◆ **Diffraction** The spreading of waves as they pass obstacles, or through apertures.

Waves will tend to spread around corners and as they pass through gaps (**apertures**). This important effect is known as **diffraction**. The simplest and most important example is shown in Figure C3.32. In this diagram the size of the aperture is the same as the wavelength, and the waves spread almost equally in all directions.



■ **Figure C3.32** Diffraction of plane waves by an aperture of width equal to one wavelength



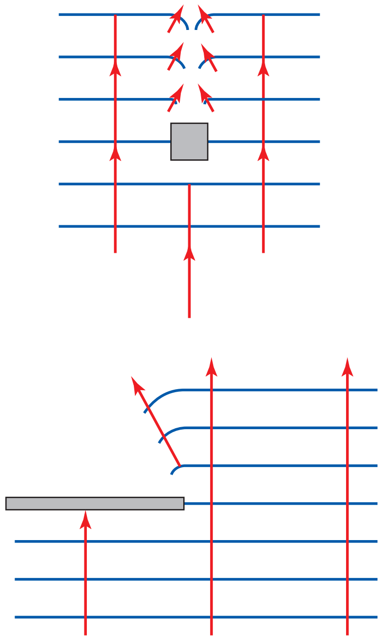
■ **Figure C3.33** aperture > wavelength

For apertures of greater width (compared to the wavelength), the effects of diffraction are less noticeable, as shown in Figure C3.33. Most of the wave energy continues travelling in its original direction. If aperture width is much greater than the wavelength, diffraction usually becomes insignificant.

Diffraction effects are most significant when the size of aperture or object ≈ wavelength.

It is important to realize that diffraction at an aperture also occurs when waves are emitted from a source or aerial, or received by an observer. Consider, for example, sound waves coming from a loudspeaker, or light waves entering an eye.

Figure C3.34 shows how waves can diffract around the edges of objects.



■ **Figure C3.34** Diffraction around objects

◆ **Laser** Source of intense, coherent, monochromatic light.

◆ **Monochromatic** Containing only one colour / frequency / wavelength (often, more realistically, a narrow range).

◆ **Aerial** A structure that receives or emits electromagnetic signals. Also called an *antenna*.

Examples of diffraction

Sound

A typical sound wave may have a wavelength of about 1 m. This is similar to the size of everyday objects and that is why we expect to be able to hear sources that we cannot see (because they are around a corner).

Light

Tool 1: Experimental techniques

Recognize and address relevant safety, ethical or environmental issues in an investigation

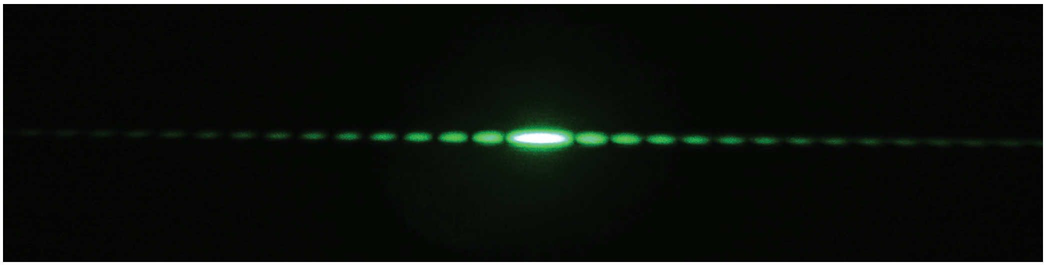
Laser light is very useful in light experiments, especially for observing diffraction. But, because of its intensity, a laser beam can be dangerous if it enters an eye.

Everybody should wear laser googles if they are available. If not, the beam should be kept horizontal and well below eye level. It should also *not* be directed towards a highly reflective surface.

Students should remain in the same places and, as soon as observations have been made, the laser should be quickly turned off.

As we have seen in Topic C.2, light has a very short wavelength (10^{-7}–10^{-6} m). This means that the diffraction of light around everyday objects will not normally be noticed. However, light *is* diffracted, and this can be observed with a very small aperture and a bright light. Figure C3.35 shows the diffraction of **laser** light through an aperture of width less than 0.1 mm. The laser light is **monochromatic**, which means there is only one colour (frequency).

The fact that light diffracts is important evidence of its wave-like nature.



■ **Figure C3.35** Diffraction pattern of light

An explanation of the pattern seen in Figure C3.35 is not important in this section, but is provided in the HL section towards the end of this topic.

Microwaves

A typical wavelength of the microwaves used in a mobile phone network is 0.1 m. Look again at Figure C3.9. The horizontal widths of the apertures emitting the waves encourages waves to be diffracted horizontally, parallel to the ground. The greater height of the aperture reduces the diffraction of waves vertically.

Microwaves are detected and transmitted by an aerial (antenna) inside the phone, which is also similar in size to the wavelength.

Conversely: a microwave beam of similar wavelength from an aircraft detection system at an airport (radar – discussed in Topic C.5) needs to be directional (not spreading out). Aircraft in different directions are located by rotating the aerial. This means that diffraction is not wanted, so the aerial's reflector is designed to be much bigger than the wavelength.

X-rays

X-ray wavelengths are comparable to the sizes of atoms and ions, and their separations, in solids. This means that X-rays are diffracted well by the regular arrangements of atoms / ions in most solids. By analysing X-ray diffraction patterns, we can learn about the structure of matter.

23 Suggest a reason why loudspeakers for producing lower pitched sounds are usually larger in size than speakers for higher pitched sounds.

24 If red light and blue light are passed through the same narrow slit, which will be diffracted more, and why?

25 Bluetooth technology uses a frequency of 2.4 GHz.

- **a** Calculate its wavelength.
- **b** Explain why this makes is suitable for, say, connecting a mobile phone to a Bluetooth speaker.

26 Sketch the wavefronts you would expect to see if parallel water waves of wavelength 1 cm were passing through an aperture of width 5 cm.

27
- **a** State a value for a typical X-ray wavelength.
- **b** Compare your answer to 0.28 nm, the approximate regular spacing of ions in a salt crystal.
- **c** Would you expect a distinct diffraction pattern to occur if X-rays were sent through water? Explain your answer.

Nature of science: Theories

Changing theories about diffraction

The first detailed observations of the diffraction patterns produced in the shadows by light passing through apertures were made more than 350 years ago but, at that time, the phenomenon defied any simple explanation because light itself was not understood, although many believe that a light beam consisted of some type of 'particles'. Many years later, after the wave theory of light became established, a theory of diffraction could be developed that involved the adding together of waveforms arriving at the same point from different places within the aperture. (Depending on the context of the discussion, the addition of waveforms can be variously described as superposition, interference or Fourier synthesis.) The more recent photon theory of light (Theme E) returns, in part, to a 'particle' explanation.

Superposition of waves

SYLLABUS CONTENT

- Superposition of waves and wave pulses.

◆ **Superposition (principle of)** The resultant of two or more waves arriving at the same point can be determined by the vector addition of their individual displacements.

When waves pass through each other at a point, we can add their displacements to determine the overall result at that place at that moment. This is called the **superposition** of waves.

The principle of superposition of waves: the overall displacement is the vector sum of the individual wave displacements.

When similar waves pass through each other they will usually have a wide range of different frequencies and amplitudes. Under these circumstances superposition effects are negligible.

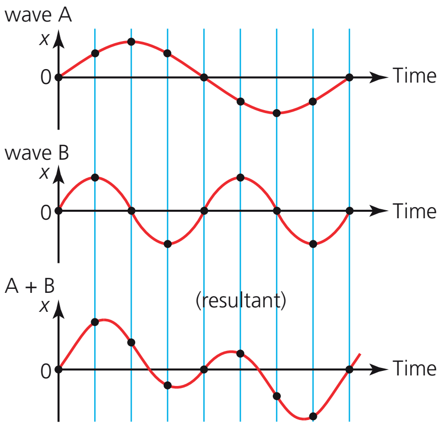


Figure C3.36 Adding wave displacements using the principle of superposition

But when waves have similar frequencies and amplitudes, the effect can be significant. Figure C3.36 shows a simple example of combining wave displacements to find a resultant.

We will use the principle of superposition in the next sub-topic (interference).

28 a Sketch a displacement–time graph for a sinusoidal oscillation of amplitude 4.0 cm and frequency 2.0 Hz. Start with a displacement of zero and continue for 0.75 s.

b On the same axes draw a graph representing an oscillation of amplitude 2.0 cm and frequency 4.0 Hz.

c Use the principle of superposition to draw a sketch of the resultant of these two waves.

29 Figure C3.37 shows two idealized square pulses moving towards each other. Draw the resultant waveform after

a 6.0 s **b** 7.0 s **c** 8.5 s **d** 12.0 s.

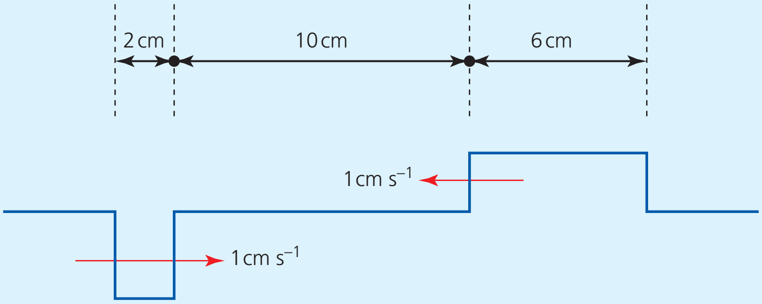


Figure C3.37 Two idealized square pulses moving towards each other

30 Two sinusoidal waves (A and B) from different sources have the same frequency and pass through a certain point, P, with the same amplitude.

a Sketch a displacement–time graph for the resultant waveform if A and B arrive at P in phase.

b Repeat for two waves that arrive at P exactly (π) out of phase.

◆ **Interference pattern (fringes)** Pattern observed when coherent waves interfere.

◆ **Coherent waves** Waves that have the same frequency and a constant phase difference.

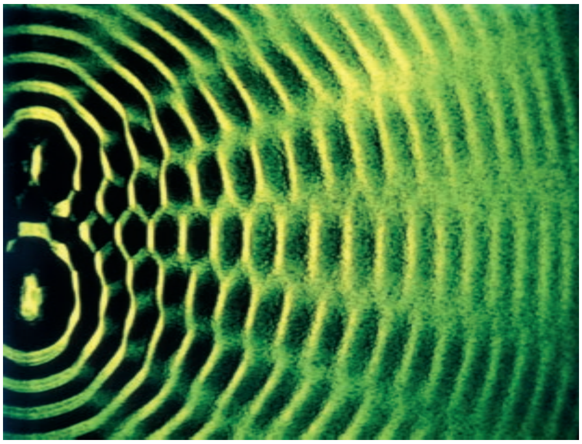


Figure C3.38 The interference of water waves on a ripple tank

Interference of waves

SYLLABUS CONTENT

- Double source interference requires coherent sources.

When the superposition of wavefronts produces a constant two- or three-dimensional pattern we describe it as an **interference pattern**. Most commonly, this effect occurs between two sources of waves which have the same single frequency and the same wave shape. Such sources (and the waves that they produce) are described as being **coherent**. Figure C3.38 shows an interference pattern produced by two sources of water waves on a ripple tank.

Constructive and destructive interference

SYLLABUS CONTENT

- The condition for constructive interference as given by: path difference = $n\lambda$
- The condition for destructive interference as given by: path difference = $\left(n + \frac{1}{2}\right)\lambda$.

◆ **Path difference** The difference in the distances from a particular point to two sources of waves.

◆ **Interference** Superposition effect that may be produced when similar waves meet. Most important for waves of the same frequency and similar amplitude. Waves arriving in phase will interfere **constructively** because their path difference = $n\lambda$. Waves completely out of phase will interfere **destructively** because their path difference = $(n + ½)\lambda$.

To explain an interference pattern like that seen in Figure C3.38, we need to consider the **path differences** between the waves arriving at various points from the two sources. Consider Figure C3.39, in which the straight lines are representing distances, not rays.

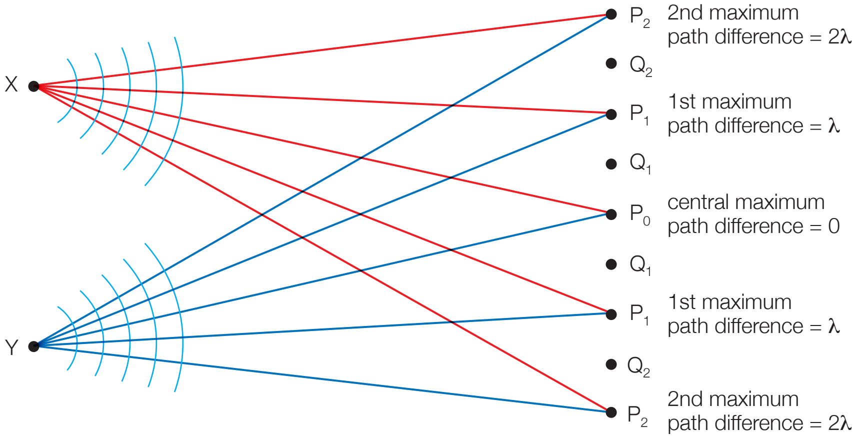
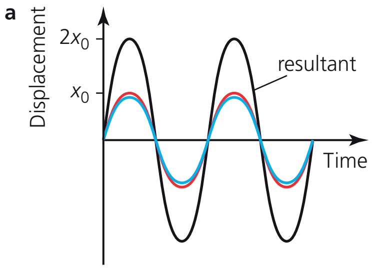
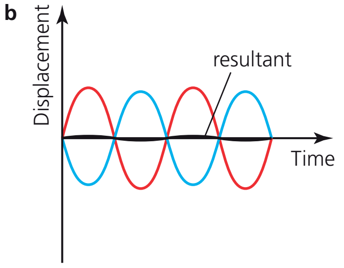


Figure C3.39 Interference and path difference



constructive interference (at P_0)



destructive interference (at Q_1)

Figure C3.40 a Constructive and **b** destructive interference

If the waves are emitted in phase from points X and Y, and both waves travel the same distance at the same speed to any point such as P_0, the waves will always be in phase at that position. Figure C3.40a shows the result, using the principle of superposition: the resulting wave has double the amplitude of the original waves. This is called **constructive interference**.

The path difference for the waves arriving at P_0, or any other point which is the same distance from X and Y, is zero. Constructive interference will also occur at any place where the waves always arrive in phase, which occurs where the path difference is 1λ, or 2λ, or 3λ … or $n\lambda$ (where n is a whole number), as shown by P_1 and P_2 in Figure C3.39.

The condition for constructive interference:

path difference = $n\lambda$

Destructive interference is shown in Figure C3.40b. This will occur at all points where the waves arriving from one source have travelled $\frac{1}{2}\lambda$, or $\frac{3}{2}\lambda$, or $\frac{5}{2}\lambda$, or $\left(n + \frac{1}{2}\right)\lambda$ more than the waves from the other source, as shown by the points Q_1 and Q_2 in Figure C3.39.

The condition for destructive interference:

$$\textbf{path difference} = \left(n + \frac{1}{2}\right)\lambda$$

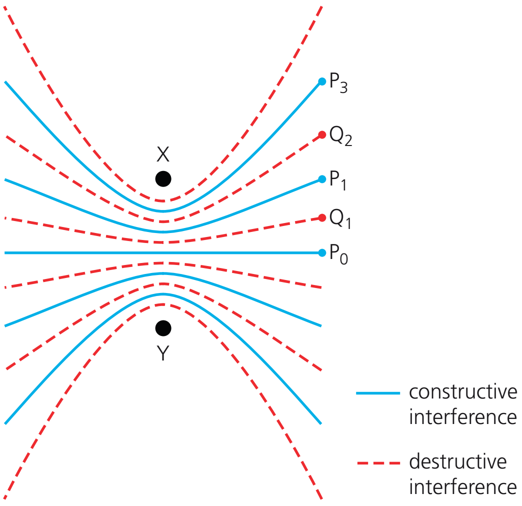


Figure C3.41 The interference pattern produced by coherent waves from two sources, X and Y

These conditions assume the usual situation: the waves are emitted in phase with each other. If the waves were exactly out of phase, these conditions would be reversed.

Constructive interference and destructive interference describe the extreme possibilities of wave superposition. At other locations the amount of interference varies between these extremes.

From Figure C3.40a, we can see that two sets of waves, each of amplitude x_0, result in a wave of amplitude $2x_0$. Since we know, from earlier in this topic, that wave intensity is proportional to amplitude squared, at places of constructive interference the intensity has quadrupled. This is possible because the intensity at places of destructive interference has been reduced.

Figure C3.41 shows the overall interference pattern produced as described in Figure C3.39. The right-hand half of Figure C3.41 can be compared to Figure C3.38.

Interference is a property *only* of waves, including electromagnetic waves like light. The interference of light cannot be explained, for example, by imagining that light consists of tiny particles. Thomas Young was the first to demonstrate that light could interfere (see below), thus demonstrating for the first time that light had wave properties.

Top tip!

When using two similar wave sources which are in phase, perfect destructive interference, resulting in waves of zero amplitude, is not possible. This is because one wave will always have a reduced amplitude because it has travelled further than the other wave to reach any particular point.

Nature of science: Theories

Competing theories

The nature of light has been widely debated for centuries. Different scientists developed different theories which seemed to partly contradict each other, and no single theory was able explain all the properties of light. This is not unusual in the development of scientific knowledge. There are many modern examples, including the consequences of climate change and the reasons for an ever-expanding Universe.

Clearly, we would prefer a single theory to help explain any particular phenomenon and to make useful predictions. However, if that is not possible at the present time, we can continue to use the best available, but less than perfect, competing theories which have proved to be useful. This is illustrated by a famous quote about the two theories of light, from William Henry Bragg: *'Physicists use the wave theory on Mondays, Wednesdays and Fridays and the particle theory on Tuesdays, Thursdays and Saturdays.'*

Young's double-slit interference experiment

SYLLABUS CONTENT

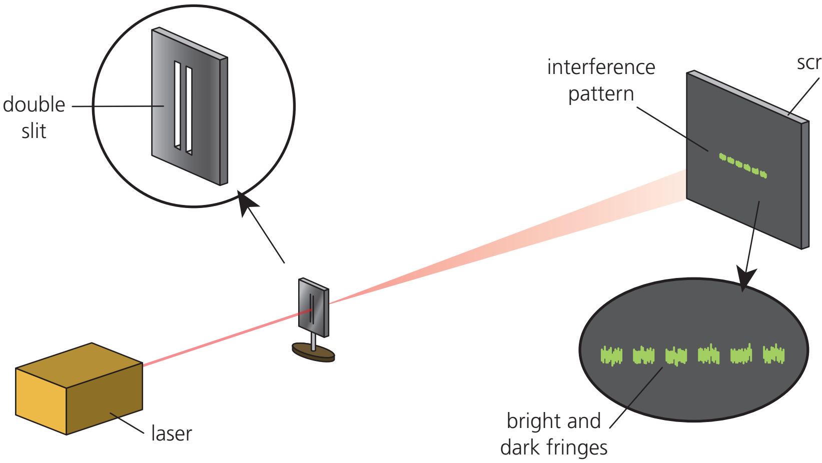
- Young's double-slit interference as given by: $s = \frac{\lambda D}{d}$, where s is the separation of fringes, d is the separation of the slits, and D is the distance from the slits to the screen.

The interference of light waves is not an everyday observation because:

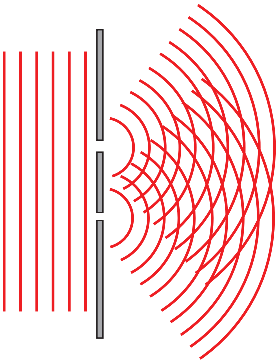
- Separate light sources are not coherent.
- Light has very small wavelengths, so that any interference pattern will be very small and difficult to observe.

◆ **Young's interference experiment** Famous experiment which provided the first evidence that light travelled as waves.

The interference of light can be demonstrated by passing monochromatic laser light through two narrow slits which are very close together (a modern version of **Young's interference experiment** (1801)). The resulting interference pattern can be seen on a distant screen, see Figure C3.42.



■ **Figure C3.42** Interference of light waves

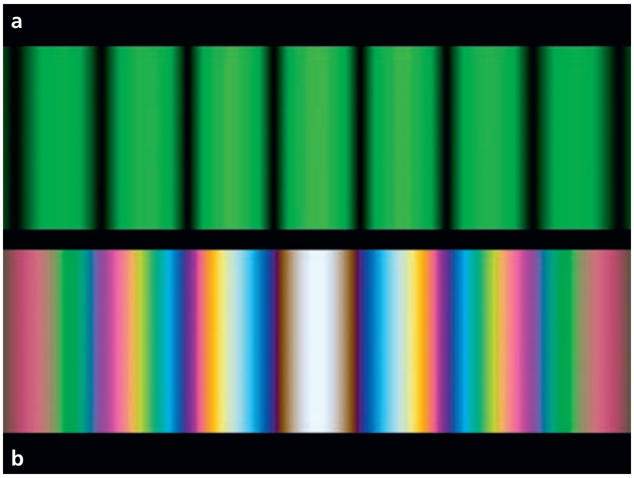


■ **Figure C3.43** Diffracted waves crossing over each other as they emerge from double slits

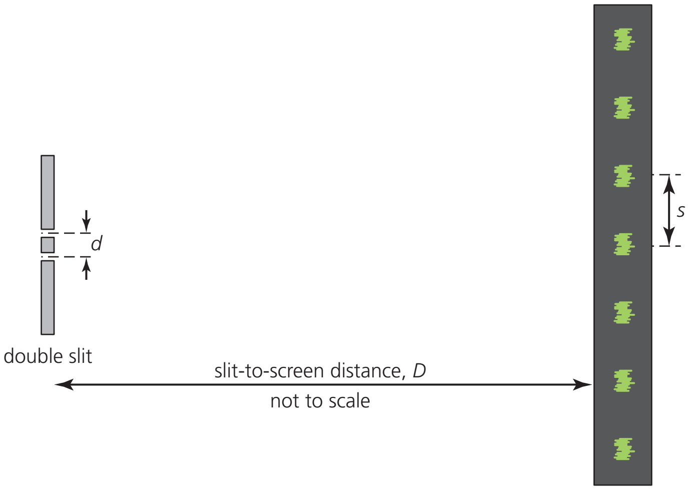
The double slits act as the necessary two coherent sources of waves. Figure C3.43 illustrates an example of the basic principle: each plane wavefront is diffracted into two beams as it emerges from the slits. The waves are coherent because they came from the same original wavefront. The diffracted waves then interfere.

A series of interference fringes is seen on the screen, as shown in the Figure C3.44a. The shape of the pattern depends on the shape of the apertures / slits. The closer the two slits, the wider the spacing of the interference pattern. Figure C3.44b shows the appearance of the fringes if white light is used with the same slits: fringes of different colours / wavelengths occur in slightly different places but overlap.

The wavelength of the light used can be determined from the geometry of the experiment (see Figure C3.45) by using the following equation, which is explained later in this topic for HL students.



■ **Figure C3.44** Interference patterns



■ **Figure C3.45** Geometry of the double-slit experiment

Separation of fringes in Young's experiment:

$$s = \frac{\lambda D}{d}$$

The separation, *s*, of the centres of the fringes is assumed to be constant, so that it is convenient to measure the total width of a number of fringes. (We are also assuming that $D >> s$ and that the light consists of plane wavefronts arriving in a direction which is perpendicular to the slits.)

WORKED EXAMPLE C3.4

In an experiment similar to that seen in Figure C3.42, the centres of the slits were separated by 0.48 mm and the screen was placed 1.96 m away from the slits. A student measured across nine equally spaced fringes and found that the centre of the first and the centre of the ninth fringe were separated by a distance of 2.25 cm.

a Determine the wavelength of the light used.

b How will the appearance of the fringes change if:

- **i** the screen is moved to a distance of 4.5 m from the slits
- **ii** the red light laser is replaced with a green light laser?

Answer

a $s = \frac{\lambda D}{d}$

$$\Rightarrow \frac{2.25\times 10^{-2}}{8} = \lambda \times \frac{1.96}{0.48 \times 10^{-3}}$$

$$\Rightarrow \lambda = 6.9 \times 10^{-7}\,\text{m}$$

b **i** The width and separation of the fringes will more than double, but their intensity (brightness) will be reduced.

ii Green light has a smaller wavelength than red light, so the fringes will be closer together.

The interference of light as it passes through two or more slits is discussed in more detail for HL students towards the end of this topic.

TOK

The natural sciences

- What kinds of explanations do natural scientists offer?

As we have seen many times, a simple equation (for example, $s = \lambda D/d$) can be used as a starting point to model scientific phenomena. Once that has been thoroughly understood, it can be adapted to more complicated situations. But is this kind of 'modelling' unique to science?

For example, would it be possible in principle to develop a mathematical model to describe the political and/or economic situation in Europe before the start of World War 1? And could such a model be used to predict what happened? Is there something fundamentally different between knowledge in physics and history, or is any historical situation just too complicated, or dependent on human behaviour, for mathematical analysis?

Interference of other types of waves

In theory, all types of waves can produce interference patterns. However, it may be impossible, or very difficult, to produce two *coherent* sources for some types of wave, because the waves are emitted in uncontrolled, random processes. This means that interference will *not* usually be observed with naturally produced waves. That is why the interference of light is not a common phenomenon.

So, to observe interference, we need to turn our attention to artificially produced waves. Apart from the interference of light, two other examples may be demonstrated in school laboratories.

Microwaves

Waves from this section of the electromagnetic spectrum are easily produced by electronic circuits and can have a wavelength of a few centimetres, which is ideal for demonstrations. See Figure C3.46.

■ **Figure C3.46** Interference of microwaves

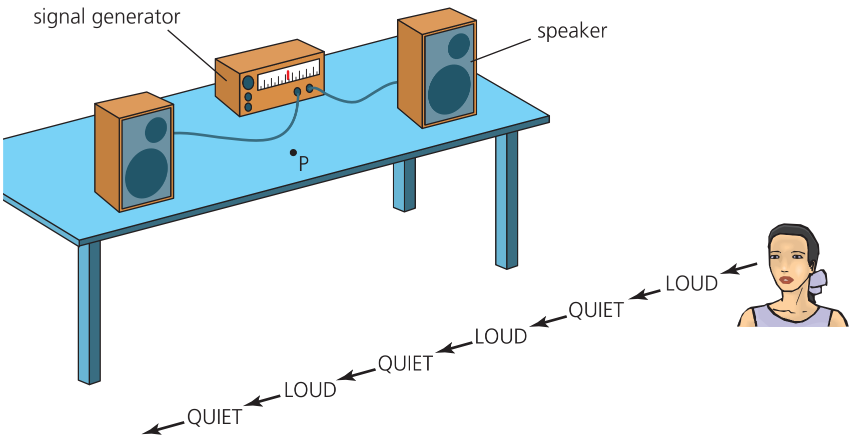
The two gaps between the aluminium sheets each have a width of about one wavelength (≈ 3 cm). They have the same effect as the double slits in Young's experiment: two diffracted, coherent waves emerge from the other side and then interfere. When the microwave detector is moved from side to side, constructive and destructive interference will be detected.

The equation $s \approx \lambda D/d$ *cannot* be used accurately in this situation, or with sound (as described below). This is because the assumptions made about the geometry of the light interference experiment (because of the very small wavelength of light) are not valid in the arrangement shown in Figure C3.46.

Sound

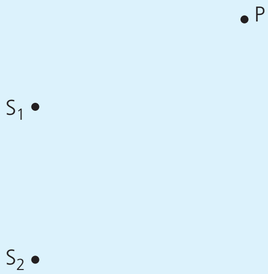
◆ **Signal generator** Electronic equipment used to supply small alternating currents of a wide range of different frequencies.

See Figure C3.47. The **signal generator** provides an oscillating electric current to the two loudspeakers, which then produce coherent longitudinal sound waves of the same frequency. As a listener walks past the speakers, as shown in the figure, the sound intensity rises and falls, because of constructive and destructive interference.



■ **Figure C3.47** Interference of sound waves

31 Figure C3.48 is one-quarter of the real size. It shows two coherent wave sources on a ripple tank and a point P. If the wavelengths are 2.5 cm, take measurements from the diagram to determine what kind of interference occurs at P.



■ **Figure C3.48** Two coherent wave sources on a ripple tank

32 Consider Figure C3.46.

a If constructive interference produces a maximum signal when the receiver is 57 cm from the centre of one slit, and 45 cm from the other, show why 3 cm and 4 cm are both possible values for the wavelength being used.

b Discuss how the actual wavelength can be quickly determined.

33 A teacher wants to demonstrate the interference of light to her class using green laser light of wavelength 5.32×10^{-7} m. She uses double slits of separation 0.50 mm. She would like the fringes to be at least 0.50 cm apart.

a Determine the closest distance she can place the screen to the slits.

b With the slits and screen still in the same places, explain how the teacher can change the experiment to produce fringes which are slightly further apart.

34 Consider Figure C3.47.

The centres of the speakers were 1.20 m apart. And the girl's closest distance to point P was 80 cm, which was where the sound was loudest. The next position where the sound was loud was 50 cm away in the direction shown.

a Are the sound waves from the two speakers emitted in phase, or out of phase?

b Use Pythagoras's theorem to determine the path difference and calculate the wavelength of the sound.

c Describe how the interference pattern would change if

i the sound frequency was increased.

ii the connections to one of the speakers was reversed.

35 The waves used to cook food in a microwave oven reflect repeatedly off the metal walls and can produce interference effects.

a Suggest how this could affect the way in which the food is cooked.

b Research into how microwave ovens are designed to overcome this problem.

36 Outline why the interference of light is not a common observation.

37 A boy stands halfway between two loudspeakers facing each other in a large open space. Both speakers are producing sounds of frequency 180 Hz. In this position he hears a loud sound.

a Explain why the sound level will decrease if he starts to walk towards either speaker.

b Discuss how the sound level will change if he walks from the mid-point along a line perpendicular to a line joining the speakers.

c How far must he walk directly towards one of the speakers before the sound level will rise to a maximum again? Assume that the speed of sound is 342 m s^{-1}.

d Explain why this experiment is best carried out in 'a large open space'.

Inquiry 1: Exploring and designing

Designing

Design and carry out an investigation into the properties of the waves emitted by a (television) remote control.

ATL C3B: Self-management skills

Breaking down major tasks into a sequence of stages

A good friend of yours is in the same physics class, but he has not been doing well and lately he seems to have become discouraged because he feels overwhelmed by work and deadlines. There are important physics examinations coming up in three weeks' time. Suggest ways in which you could advise him to prepare for the examinations.

A closer look at single-slit diffraction of light

SYLLABUS CONTENT

- Single-slit diffraction, including intensity patterns, as given by: $\theta = \frac{\lambda}{b}$, where b is the slit width.

Common mistake

Single-slit diffraction patterns and double-slit interference patterns are similar to each other and easily confused. The most obvious difference is that the central fringe of the single-slit diffraction pattern is brighter and wider than the other fringes.

Figure C3.32 showed a diffraction pattern produced by monochromatic light passing through a narrow slit. It can be produced by an experimental arrangement similar to that seen in Figure C3.42, but with the double slits replaced by a single slit.

In the introduction to diffraction, Figure C3.33 indicated the pattern of diffraction we would expect when waves pass through an aperture greater than the wavelength. Now, we need to explain why the diffraction pattern produced when light passes through a single narrow slit is different: it has fringes similar to that produced by double slits.

Firstly, it is important to understand that even a 'narrow' slit of width, for example, $b = 0.1\,\text{mm}$ is much greater than one wavelength, λ, of light. In fact:

$$\frac{b}{\lambda} \approx \frac{0.1 \times 10^{-3}}{5 \times 10^{-7}} = 200$$

Interference within a single wavefront

We need to imagine that a plane wavefront emerging from a single slit acts as if it was a series of point sources of 'secondary wavelets', maybe one for every wavelength. These ideas were famously first put forward as an explanation for the propagation of all waves by Christian Huygens in 1690.

These wavelets will be coherent and they will interfere.

Firstly, consider how secondary wavelets from the edge of the slit may interfere.

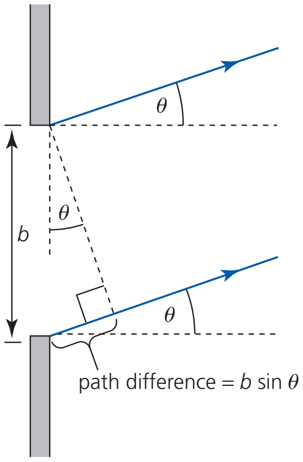


Figure C3.49 Path differences and interference

Figure C3.49 shows a typical direction, θ, in which **secondary waves** travel away from a single narrow slit of width, b.

If θ is zero, all the secondary waves will interfere constructively in this direction (straight through the slit) because there is no path difference between them. (Of course, in theory, waves travelling parallel to each other in the same direction cannot meet and interfere, so we will assume that the waves' directions are very nearly parallel.)

Consider what happens for angles increasingly greater than zero. The path difference, as shown in Figure C3.49, equals $b \sin \theta$ and this increases as the angle θ increases. There will be angles at which the waves from the two edges of the slit interfere constructively because the path difference has increased to become equal to a 1λ, 2λ, 3λ … and so on.

But if secondary wavelets from the edges of the slit interfere constructively, what about interference between all the other secondary wavelets? Consider Figure C3.50 in which the slit has been divided into a number of point sources of secondary wavelets. (Ten points have been chosen, but it could be *many* more.)

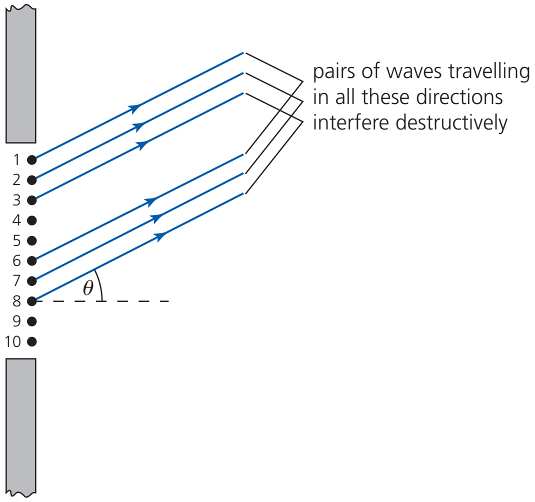


Figure C3.50 Secondary waves that will interfere destructively can be 'paired off'

◆ **Secondary waves** The propagation of waves in two or three dimensions can be explained by considering that each point on a wavefront is a source of secondary waves.

If the angle, θ, is such that secondary wavelets from points 1 and 10 would interfere constructively because the path difference is one wavelength, then secondary waves from 1 and 6 must have a path difference of half a wavelength and interfere destructively. Similarly, waves from points 2 and 7, points 3 and 8, points 4 and 9 and points 5 and 10 must all interfere destructively. In this way waves from all points can be 'paired off' with others, so that the first minimum of the diffraction pattern occurs at such an angle that waves from the edges of the slit would otherwise interfere constructively.

The first minimum of the diffraction pattern occurs when the path difference between secondary waves from the edge of the slit is equal to one wavelength. That is, if $b \sin\theta = \lambda$.

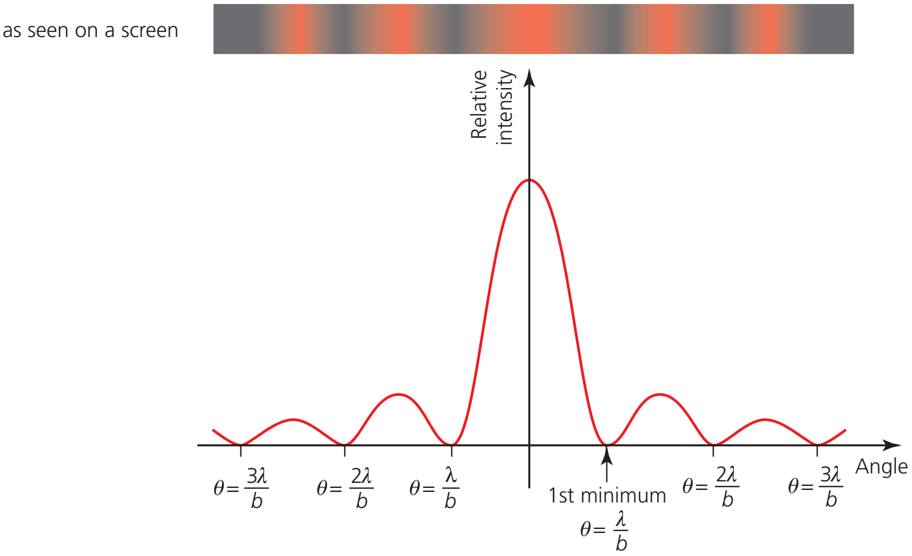
For the diffraction of light, the angle θ is usually small and approximately equal to $\sin\theta$, if the angle is expressed in radians. (This is valid for angles up to approximately 10°, 0.17 rad.). So that:

The angle for the first minimum of a **single-slit diffraction** pattern is:

$$\theta = \frac{\lambda}{b}$$

◆ Single-slit diffraction The simplest diffraction pattern is that produced by wavefronts interfering after they have passed through a narrow, rectangular slit. The first minimum occurs at an angle such that $\theta = \frac{\lambda}{b}$.

Similar reasoning will show that other diffraction minima will occur at angles, $\theta = 2\lambda/b, 3\lambda/b, 4\lambda/b$... and so on ($\theta_n = n\lambda/b$). These angles are represented on the intensity-angle graph shown in Figure C3.51.



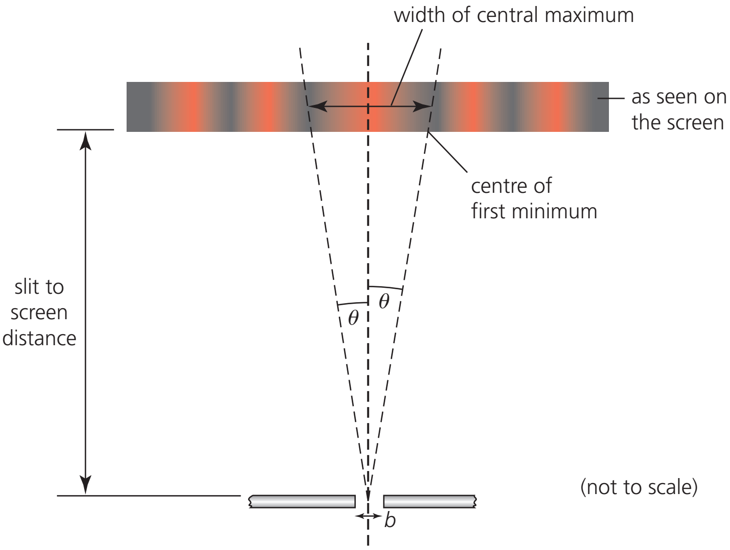
■ **Figure C3.51** Variation of intensity with angle for single-slit diffraction

Top tip!

As we have seen, a full explanation of diffraction involves discussing interference, and this can easily cause confusion, especially if they are thought to be two separate phenomena. To be clear: diffraction is the change of direction that occurs when waves pass gaps and obstacles. Diffracted wavefronts may then undergo superposition effects, which can lead to a pattern being formed which is usually called a 'diffraction pattern', although the principles of interference are used to explain it.

WORKED EXAMPLE C3.5

Monochromatic light of wavelength 663 nm (663×10^{-9} m) is shone through a gap of width 0.0730 mm.



■ **Figure C3.52** Monochromatic light shone through a gap

a Calculate the angle at which the first minimum of the diffraction pattern is formed.

b If the pattern is observed on a screen that is 2.83 m from the slit, determine the width of the central maximum.

Answer

a $\theta = \frac{\lambda}{b} = \frac{663 \times 10^{-9}}{7.30 \times 10^{-5}} = 9.08 \times 10^{-3}$ radians

b $\theta \approx \sin\theta = \frac{\text{half width of central maximum}}{\text{slit to screen distance}}$

half width of central maximum = $(9.08 \times 10^{-3}) \times 2.83 = 0.0257$ m (0.02569... seen on calculator display)

width of central maximum = $0.02569 \times 2 = 0.0514$ m

38 Electromagnetic radiation of wavelength 2.37×10^{-7} m passes through a narrow slit of width 4.70×10^{-5} m.

- **a** State in which part of the electromagnetic spectrum this radiation occurs.
- **b** Suggest how it could be detected.
- **c** Calculate the angle of the first minimum of the diffraction pattern.

39 Determine the wavelength of light that has a first diffraction minimum at an angle of 0.0038 radians when it passes through a slit of width 0.15 mm.

40 When light of wavelength 6.2×10^{-7} m was diffracted through a narrow slit, the central maximum had a width of 2.8 cm on a screen that was 1.92 m from the slit. Calculate the width of the slit.

41 **a** Sketch and label a relative intensity–angle graph for the diffraction of red light of wavelength 6.4×10^{-7} m through a slit of width 0.082 mm. Include at least five peaks of intensity.

- **b** Add to your graph a sketch to show how monochromatic blue light would be affected by the same slit.

ATL C3C: Thinking skills

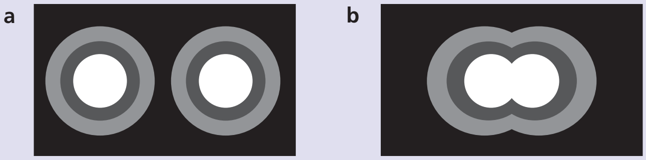
Providing a reasoned argument to support conclusions

Resolution

When light waves enter our eyes, they are diffracted and this affects our ability to see detail. The ability of our eyes (and/or optical instruments) to see objects as separate is called **resolution**. For example, our eyes will not be able to resolve the individual leaves seen on a distant tree.

This is easiest explained by considering two identical sources of light at night, for example car headlights a long way away. Considering the distance involved, we can assume that the light waves are effectively emitted from points sources, but when they are received by the eye, the images on the back of the eye (retina) are not points, but diffraction rings. (Rings are formed because the aperture in the eye is circular.) Consider Figure C3.53.

When the lights are close (Figure C3.53a), we can see two headlights, but when the lights are a long way away (maybe 5 km or more), the diffraction patterns overlap and cannot be distinguished – only one light is seen.



■ **Figure C3.53** Images of two point sources observed through circular apertures that are **a** easily resolvable and **b** just resolvable

Most of the wave intensity will be concentrated near the centre of diffraction rings, so that under most circumstances, diffraction effects are not significant. However, when we want to see fine details under a microscope, or make astronomical observations on distant objects, diffraction is the major factor limiting resolution.

We have seen that the amount of diffraction at an aperture can be represented by the ratio λ/b, so that the resolution of any telescope, or a microscope, can be improved by increasing the width of the aperture receiving the waves, or reducing the wavelength involved (often not possible).

Radio telescopes such as that seen in Figure C3.54 are much larger than optical telescopes. Explain why. Support your explanation with scientific reasoning.

■ **Figure C3.54** The Jodrell Bank radio telescope in England, UK

◆ **Resolution (optical)** The ability of an imaging system to identify objects as separate.

Two slits, multiple slits and diffraction gratings

SYLLABUS CONTENT

▶ Interference patterns from multiple slits and diffraction gratings as given by: $n\lambda = d\sin\theta$.

LINKING QUESTIONS

- What can an understanding of the results of Young's double-slit experiment reveal about the nature of light?
- What evidence is there that particles possess wave-like properties such as wavelength? (NOS)

These questions link to understandings in Topic E.2.

In this section we will discuss

- the advantages of passing light waves through more and more parallel slits
- how the pattern of single-slit diffraction affects interference between two or more slits.

Earlier in this topic we discussed the interference patterns seen when light waves pass through two narrow slits which are close together: for a given wavelength, λ, we saw that the spacing, s, of the pattern seen on a fixed screen depends on the separation of the slits, d, and the slits to screen distance, D $\left(s = \lambda\frac{D}{d}\right)$. See Figure C3.45.

Before we look at this in more detail, we need to obtain a more generalized equation which can be used for predicting where constructive interference occurs with any number of slits. Consider Figure C3.55, which shows two (almost) parallel rays from adjacent slits representing plane waves which interfere constructively when they have an angle θ to the original direction in which the waves were travelling. (For the moment, we will not consider any interference effects between wavelets from the same slit.)

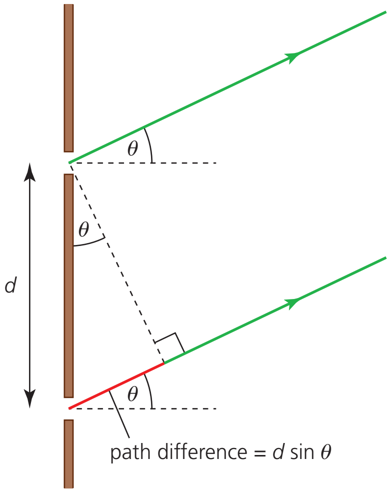


Figure C3.55 Explaining path difference = $d \sin\theta$

The extra distance travelled by the lower ray, the path difference, can be determined from the right-angled triangle: path difference = $d \sin\theta$.

Since the waves interfere constructively, we know that the path difference must also be equal to a whole number, n, of wavelengths. This leads to:

> Constructive interference occurs at angles such that:
>
> $$\boldsymbol{n\lambda = d \sin\theta}$$

In other words, constructive interference will occur at angles which have sines equal to λ/d, $2\lambda/d$, $3\lambda/d$, $4\lambda/d$ … and so on.

$$\sin\theta = \frac{n\lambda}{d}$$

For $n = 1$ and small angles:

$$\sin\theta = \frac{s}{D} = \frac{n\lambda}{d}$$

Common mistake

Students often confuse the equations $\theta = \lambda/b$ and $\sin\theta = n\lambda/d$.

$\theta = \lambda/b$ predicts the small angle for which diffraction at a single slit, of width b, produces an intensity minimum.

$\sin\theta = n\lambda/d$ predicts the angles at which interference of light from two or more slits, of separation d, produces intensity maxima.

(which can be re-arranged to give:

$$s = \frac{\lambda D}{d}$$

as used previously for two slits).

Similarly, destructive interference occurs at angles such that:

$$\left(n + \frac{1}{2}\right)\lambda = d \sin\theta$$

These equations can be used with any number of slits. Indeed, its most common application is with the very large number of slits of a *diffraction grating*, as discussed below. In such arrangements the angles are typically large enough that *sin* θ must be used, not θ in radians.

Note that the angles at which constructive and destructive interference occur depends on the separation of the slits, but *not* the number of slits.

Figure C3.56 represents these equations in the form of an intensity–$\sin\theta$ graph (for a small number of slits).

Figure C3.57 shows how what is seen on the screen relates to the various angles and values of n.

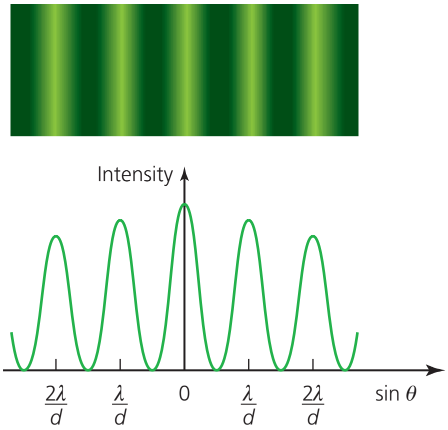


Figure C3.56 Variation of intensity with angle for multiple-slit interference

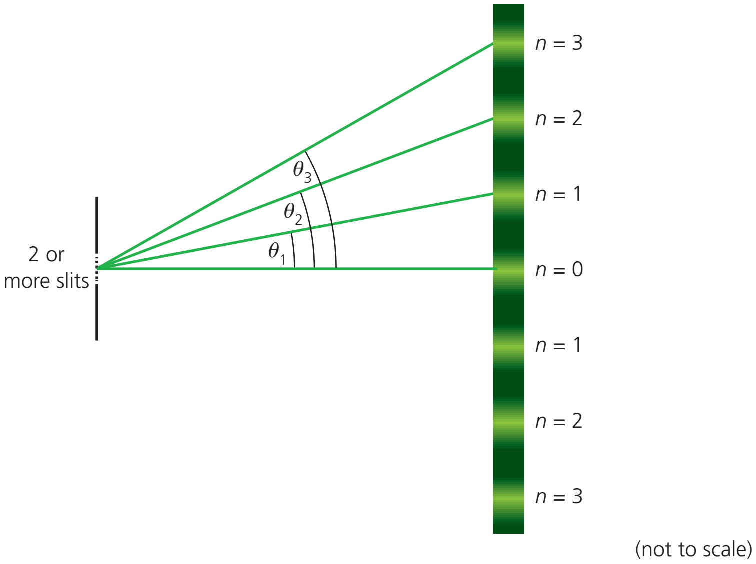


Figure C3.57 Separation and numbering of fringes seen on a screen

WORKED EXAMPLE C3.6

After passing green laser light through a few parallel slits of spacing 0.059 mm, an interference pattern was seen on a screen placed 3.12 m from the slits. It was similar in appearance to that seen in Figure C3.56. If the distance from the centre of the pattern to the centre of the third fringe was 8.42 cm

a Determine the:

i tan **ii** sin

of the angle, θ_3, between two rays from the slits going to the centre of the pattern and the centre of the third fringe.

b What is this angle in:

i degrees **ii** radians?

c Calculate a value for the wavelength of the laser light.

Answer

a i $\tan\theta_3 = \frac{8.42}{312} = 0.0270$

ii sine of the same angle = 0.0270 (same as tan)

b i 1.55°

ii 0.0270 (same as tan and sin)

c $n\lambda = d\sin\theta$

$3\lambda = (0.059 \times 10^{-3}) \times 0.0270$

$\lambda = 5.3 \times 10^{-7}\,\text{m}$

Because the angles are small the same answer for the wavelength could have been determined from

$$s = \frac{\lambda D}{d}$$

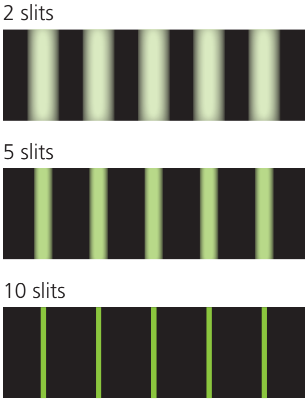


Figure C3.58 How an interference pattern changes as more slits are involved

Effect of having more slits (of the same width and separation)

It has already been stressed that the number of slits does not affect the directions in which constructive interference occurs. So, what is the advantage of having a light beam pass through more slits? Certainly, this should make the fringes brighter / more intense, but there is another, more important, reason for using **multiple slits**: the intensity peaks become ‘sharper’, more precisely located. This is shown in Figure C3.58, which compares the sharpness of the interference peaks obtained with different number of slits.

This effect is used particularly in diffraction gratings.

Diffraction gratings

A **diffraction grating** (see Figure C3.59) is a very large number of slits, very close together. A typical grating has 600 parallel slits every mm. We usually refer to this as 600 lines/mm. The separation of the centres of such lines, d, is $1.67 \times 10^{-6}\,\text{m}$, which is equivalent to about (only) three average wavelengths of light. This is much smaller than the width of the slits we have been discussing so far.

This very small separation of the slits, d, results in large values of $\sin\theta$ (consider $n\lambda = d\sin\theta$). If the light incident on the grating is also spread over a large number of lines, the intensity peaks will be well separated, intense and sharp. Figure C3.60 compares the intensity peaks produced by a diffraction grating to those produced by double slits. (The difference in intensity levels will be even greater than that shown.)

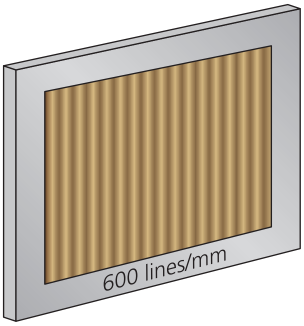
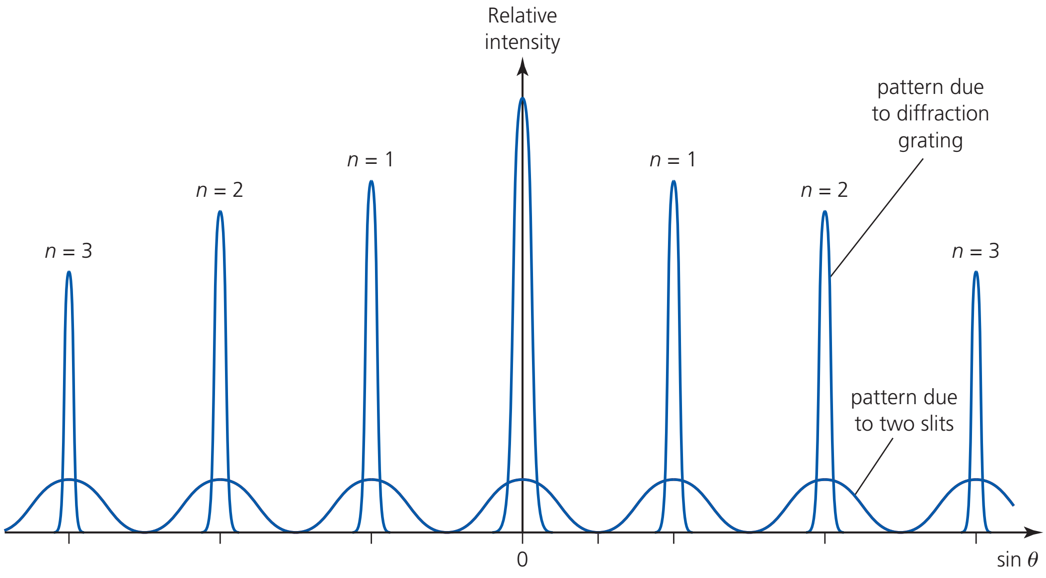


Figure C3.59 Diffraction grating

◆ **Multiple slits** By increasing the number of parallel slits (of the same width) on which a light beam is incident, it is possible to improve the resolution of the fringes / spectra formed.

◆ **Diffraction grating** A large number of parallel slits very close together. Used to disperse and analyse light.



■ **Figure C3.60** Comparing the maxima produced by double slits and a diffraction grating using monochromatic light

WORKED EXAMPLE C3.7

Use the data in the paragraphs above to calculate the angles (in degrees) for the first, second and third peaks from the centre, for light of wavelength 594 nm (5.94×10^{-7} m).

Answer

$n\lambda = d\sin\theta$

For $n = 1$, $\sin\theta = \frac{\lambda}{d} = \frac{5.94 \times 10^{-7}}{1.67 \times 10^{-6}} = 0.3557 \Longrightarrow \theta = 20.8°$

For $n = 2$, $\sin\theta = 2 \times 0.3557 = 0.7114 \Longrightarrow \theta = 45.3°$

For $n = 3$, $\sin\theta = 3 \times 0.3557 = 1.067 \Longrightarrow \sin\theta > 1$, which is not possible.

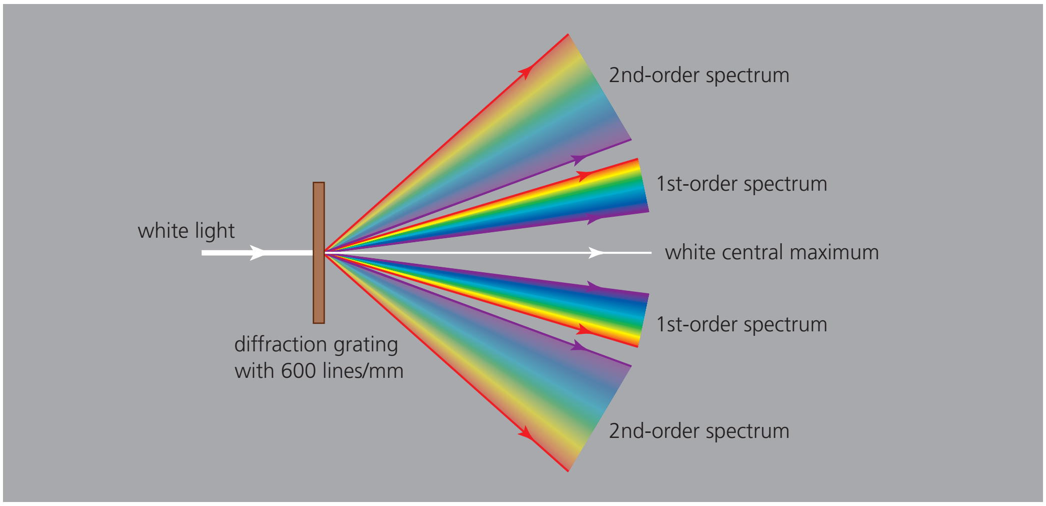
Only two peaks can be seen under these conditions. This is common when using gratings. One or two intense peaks are usually all that are required.

Observing spectra with diffraction gratings

Diffraction gratings are widely used for producing spectra and determining unknown wavelengths of light. (Prisms can also be used.)

Continuous white light spectrum

If white light is sent through a diffraction grating, different wavelengths/colours will be sent in slightly different directions. The light will be *dispersed* into spectra, as seen in Figure C3.61.



■ **Figure C3.61** White light passing through a diffraction grating

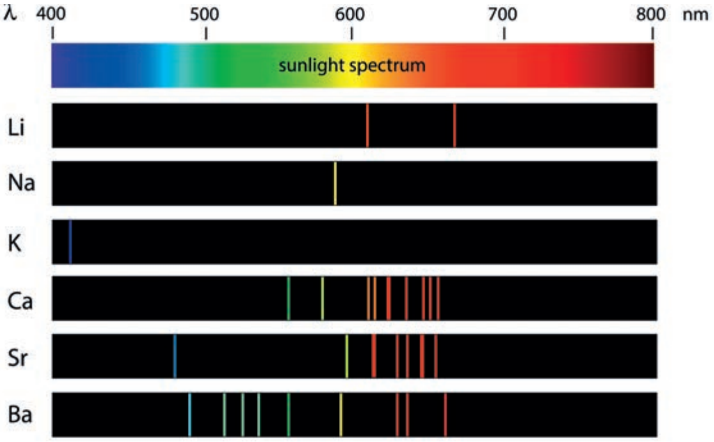
◆ **Spectral orders**
Considering the diffraction grating equation, $n\lambda = d\sin\theta$: different orders correspond to different values of n.

◆ **Spectrum, line**
A spectrum of separate lines (rather than a continuous spectrum), each corresponding to a discrete wavelength and energy.

We commonly refer to the **order of a spectrum**. All the wavelengths / colours with $n = 1$ are called the first-order spectrum, all the wavelengths / colours with $n = 2$ are called the second-order spectrum and so on.

Line spectra

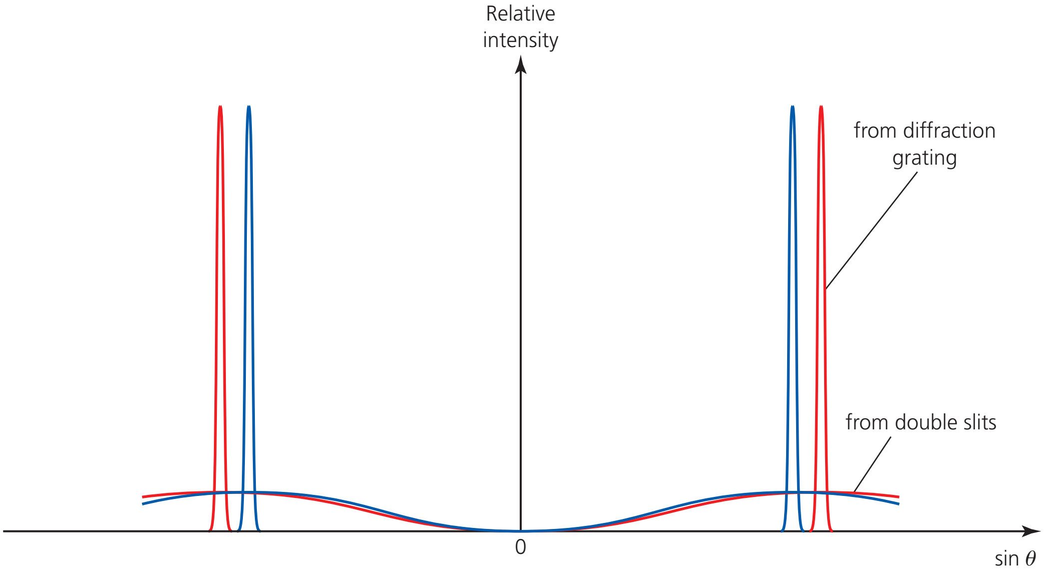
If atoms of a particular element, or some compounds, are given enough energy they will emit light in the form of a **line spectrum** (see Figure C3.62), not a continuous white light spectrum. This is discussed in much more detail in Topic E.1, and a full understanding is not expected here.



■ **Figure C3.62** Line spectra of various elements

Diffraction gratings offer an excellent way of producing and measuring line spectra. This is shown in Figure C3.63, which compares the ability of double slits and a diffraction grating to produce separate lines. The grating produces a much greater resolution.

Diffraction grating have very large numbers of lines very close together. This means that they are excellent at producing intense line spectra with high resolution.



■ **Figure C3.63** High resolution produced by diffraction gratings

Modulation by single-slit diffraction

SYLLABUS CONTENT

- ▶ The single-slit pattern modulates the double-slit interference pattern.

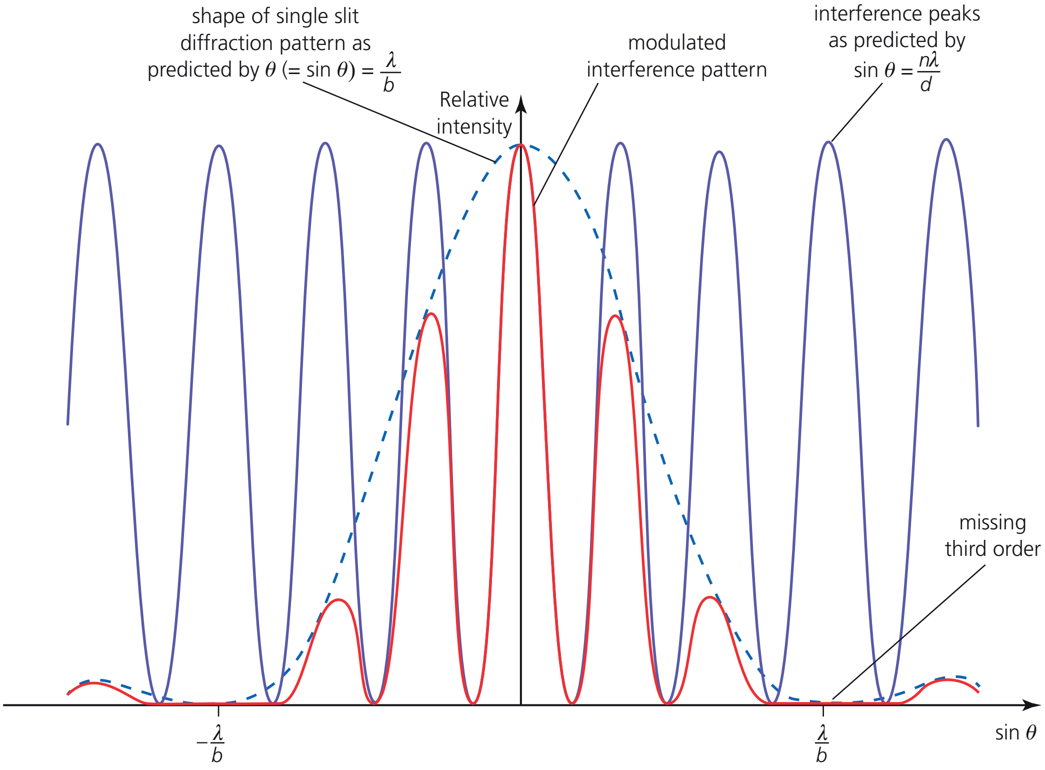
The explanation of the interference effects produced between light waves coming from multiple slits (including diffraction gratings) has so far ignored one very important factor: the effect of the single-slit diffraction pattern produced by the interference of secondary wavefronts within each and every slit. (We will assume that all slits are identical.)

The conclusions from a detailed superposition analysis of all wavefronts can be expressed in the following description:

◆ **Modulation** Changing the amplitude (or frequency) of a wave according to variations in a secondary effect.

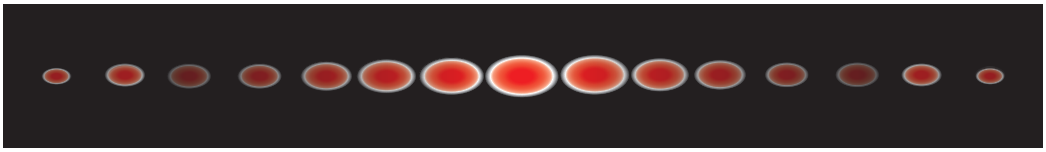
The intensity of the interference peaks produced by double slits, or multiple slits, is **modulated** by the shape of the single-slit diffraction produced by each individual slit.

This is explained by Figure C3.64. Note that *d* must always be larger than *b*, so that the spacing of the interference pattern must always be smaller than the spacing of the diffraction pattern from each slit. The shape and spacing of the single-slit diffraction pattern act as a guiding 'envelope' for the size of the interference peaks.



■ **Figure C3.64** How single-slit diffraction modulates multiple slit interference

This 'modulation' can result in some orders being suppressed or missing. For example, the third order in Figure C3.64 and the fourth and fifth orders in Figure C3.65.



■ **Figure C3.65** Missing orders because of the modulation effect of single-slit diffraction

HL ONLY

42 Light of wavelength 460 nm is incident normally on a diffraction grating with 200 lines per millimetre. Calculate the angle to the normal of the third-order maximum.

43 A diffraction grating is used with monochromatic light of wavelength 6.3×10^{-7} m and a screen a which is perpendicular distance of 2.75 m away.
Calculate how many lines per millimetre are on the diffraction grating if it produces a second-order maximum 68 cm from the centre of the pattern.

44 Monochromatic light of wavelength 530 nm is incident normally on a grating with 750 lines per millimetre. The interference pattern is seen on a screen that is 1.82 m from the grating.
Calculate the distance between the first and second orders seen on the screen.

45 A prism can also be used to produce a spectrum. Outline why red light is refracted less than blue light in a prism, but red light is diffracted more than blue light by a diffraction grating.

46 When using white light, explain why red light in the second-order spectrum overlaps with blue light in the third-order spectrum.

47 Sketch a relative intensity against $\sin\theta$ graph for monochromatic light of wavelength 680 nm incident normally on a diffraction grating with 400 lines per mm.

48 A diffraction grating produced two first-order maxima for different wavelengths at angles of 7.46° and 7.59° to the normal through the grating. This angular separation was not enough to see the two lines separately.
What is the angular separation of the same lines in the second order?

49 Figure C3.66 shows a centimetre ruler placed next to an interference pattern seen on a screen which was placed 3.40 m away from double slits.

a If the separation of adjacent slits was 0.64 mm, determine the wavelength of the light being used.

b Estimate the width of the individual slits.

c How would the pattern change if the light passed through 10 slits (of the same width and separation)?

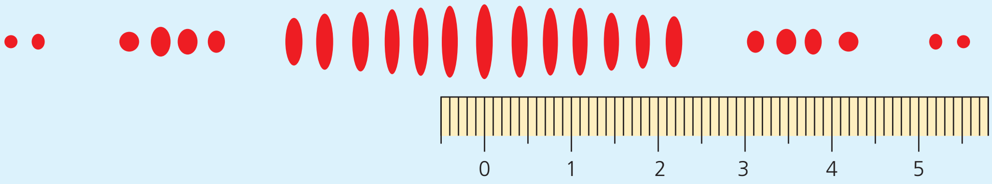


Figure C3.66 A centimetre ruler placed next to an interference pattern

C.4 Standing waves and resonance

Guiding questions

- What distinguishes standing waves from travelling waves?
- How does the form of standing waves depend on the boundary conditions?
- How can the application of force result in resonance within a system?

The nature of standing waves

SYLLABUS CONTENT

- The nature and formation of standing waves in terms of superposition of two identical waves travelling in opposite directions.
- Nodes and antinodes, relative amplitude and phase difference of points along a standing wave.

So far, the discussion of waves has been about travelling waves, which transfer energy progressively away from a source. Now we turn our attention to waves that remain in the same position.

Consider two travelling waves of the same shape, frequency, wavelength and amplitude moving in opposite directions, such as shown in Figure C4.1, which could represent transverse waves on a string or a rope.

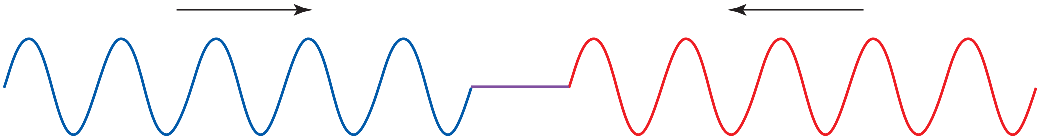


Figure C4.1 Two travelling waves of the same shape, frequency, wavelength and amplitude moving in opposite directions

As these waves pass through each other, they will combine to produce an oscillating wave pattern that does not change its position. Such patterns are called standing waves.

Standing waves usually occur in confined systems in which waves are reflected back upon each other repeatedly.

Typical examples of standing wave patterns (on strings) are shown in Figure C4.2. Note that a camera produces an image over a short period of time (not an instantaneous image) and that is why the fast-moving string appears blurred. This is also true when we view such a string with our eyes.

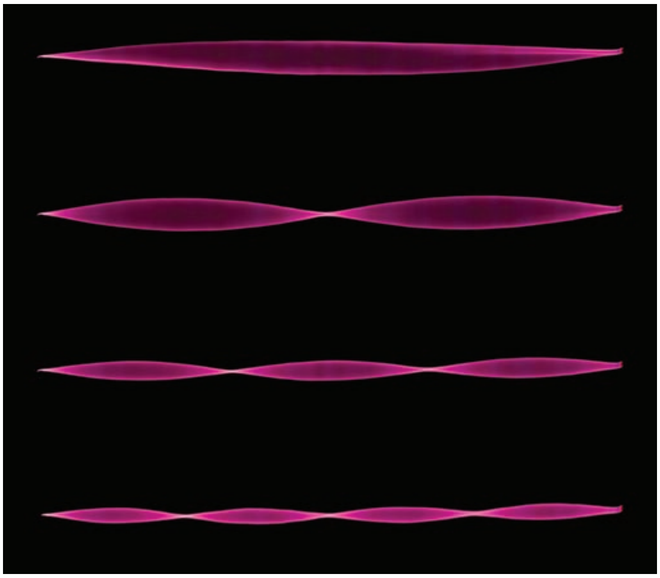


Figure C4.2 Standing waves on a stretched string

Simple standing wave patterns can be produced by shaking one end of a rope, or long stretched spring, at a suitable frequency, while someone holds the other end stationary. Patterns like those shown in Figure C4.2 require high frequencies (because a string is much less massive than a rope), but can be produced in a laboratory by vibrating a stretched string with a mechanical vibrator which is controlled by variable electrical oscillations from a signal generator. This apparatus can be used to investigate the places at which the string appears to be stationary and the frequencies at which they occur.

◆ **Standing wave** The kind of wave that can be formed by two similar travelling waves moving in opposite directions. The most important examples are formed when waves are reflected back upon themselves. The wave pattern does not move and the waves do not transfer energy.

◆ **Nodes** The positions in a standing wave where the amplitude is zero.

◆ **Antinodes** The positions in a standing wave where the amplitude is greatest.

Nodes and antinodes

A standing wave pattern remains in the same place.

Points where the displacement is always a minimum (often zero) are called **nodes**. Points where the amplitude is greatest are called **antinodes**.

Figure C4.3 represents the third wave from the photograph in Figure C4.2 diagrammatically. Note that the distance between alternate nodes (or antinodes) is one wavelength.

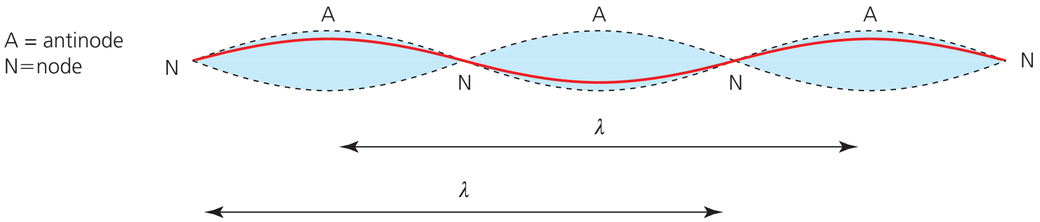


Figure C4.3 Nodes and antinodes in a standing wave

A system in which there is a standing mechanical wave has both kinetic energy and potential energy, but energy is not transferred *away* from the system in the form of a wave along the system. However, there will be energy dissipation within the system because of resistive forces, so that the amplitude of the standing wave will decrease, unless it is sustained by energy transferred from an external driving frequency (see later).

Nodes occur at places where the two waves are *always* exactly out of phase. At other places, the displacements will oscillate between zero and a maximum value which depends on the phase difference between the waves moving in opposite directions. At the antinodes the two waves are always exactly in phase. You should try using a computer simulation to illustrate this time-changing concept.

Between adjacent nodes all parts of the medium oscillate in phase with each other with the same frequency. Each position has a constant amplitude, but the amplitudes vary as shown in Figure C4.4.

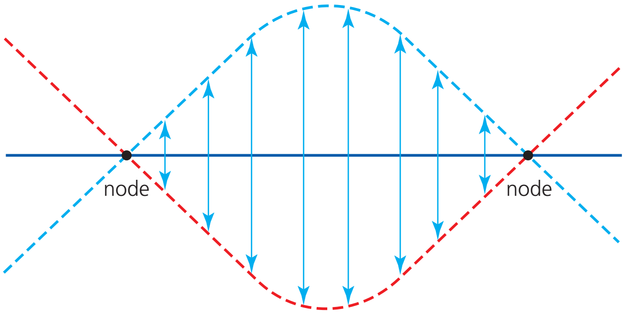


Figure C4.4 Variation of amplitude between nodes

Standing waves are possible with any kind of wave (transverse or longitudinal) moving in one, two or three dimensions. But, for simplicity, discussion will be confined to one-dimensional waves, such as transverse waves on a stretched string, or longitudinal sound waves in pipes.

Boundary conditions

Standing waves occur most commonly when waves are reflected repeatedly back from boundaries in a confined space. The frequencies of the standing waves will depend on the nature of the ends of the system. These are called the **boundary conditions**. For example, the ends of a string, or rope, may be fixed in one position or free to move; the ends of a pipe could be closed or open. When the ends are free to move, we would assume that standing wave has antinodes there. There will be nodes at fixed ends.

◆ **Boundary conditions** The conditions at the ends of a standing wave system. These conditions affect whether there are nodes or antinodes at the ends.

The boundary conditions of a standing wave system describe whether there are nodes or antinodes at the end of the system.

Comparing standing waves with travelling waves

Table C4.1 Comparison of standing waves and travelling waves

	Standing waves	**Travelling waves**
Wave pattern	standing (stationary)	travelling (progressive)
Energy transfer	no energy is transferred	energy is transferred through the medium
Amplitude (assuming no energy dissipation)	amplitude at any one place is constant but it varies with position between nodes: maximum amplitude at antinodes, zero amplitude at nodes	all oscillations have the same amplitude
Phase	all oscillations between adjacent nodes are in phase	oscillations one wavelength apart are in phase; oscillations between are not in phase
Frequency	all oscillations have the same frequency	all oscillations have the same frequency
Wavelength	twice the distance between adjacent nodes	shortest distance between points in phase

Standing wave patterns in strings

SYLLABUS CONTENT

- Standing waves patterns in strings.

If a stretched string *fixed at both ends* is struck or plucked, it can only vibrate as a standing wave with nodes at both ends. The amplitude of the standing wave will usually decrease quickly as energy is dissipated.

The simplest way in which it can vibrate is shown at the top of Figure C4.5. This is known as the **first harmonic**. It is usually the most important mode of vibration, but a series of other harmonics (**modes of vibration**) is possible and can occur at the same time as the first harmonic. Some of these harmonics are shown in Figure C4.5.

◆ **Harmonics** Different frequencies (modes) of standing wave vibrations of a given system. The frequencies are all multiples of the frequency of the **first harmonic**.

◆ **Modes of vibration** The different ways in which a standing wave can arise in a given system.

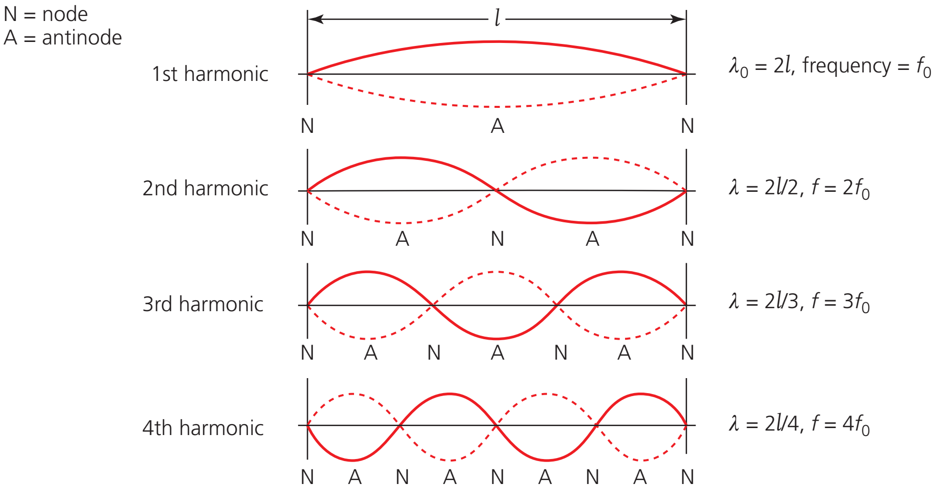


Figure C4.5 Modes of vibration of a stretched string fixed at both ends

The first harmonic is the standing wave with the lowest possible frequency, f_0 (greatest wavelength). Other harmonic are mathematical multiples of this frequency.

If a string is fixed at both ends, the wavelength, λ_0, of the first harmonic is $2l$, where l is the length of the string. The speed of the wave, v, along the string depends on the tension and the mass per unit length. If the wave speed is known, the frequency of the first harmonic, f_0, can be calculated using $v = f\lambda$ (from Topic C.2):

$$f_0 = \frac{v}{\lambda_0} = \frac{v}{2l}$$

The wavelengths of the harmonics are, starting with the first (longest), $2l, \frac{2l}{2}, \frac{2l}{3}, \frac{2l}{4}$ and so on. The corresponding frequencies, starting with the lowest, are f_0, $2f_0$, $3f_0$, $4f_0$ and so on.

The wavelength of the first harmonic can be found from the length of the system and the boundary conditions.

$v = f\lambda$ can then be used to connect frequency and wave speed.

WORKED EXAMPLE C4.1

A stretched rope with two fixed ends has a length of 0.98 m, and waves travel along it with a speed of $6.7\,\mathrm{m\,s^{-1}}$.

a Calculate:
- **i** the wavelength of the first harmonic
- **ii** the frequency of the first harmonic.

b Calculate:
- **i** the wavelength of the fourth harmonic
- **ii** the frequency of the fourth harmonic.

c Explain how your answers would change if the tension in the rope was increased.

Answer

a i $\lambda_0 = 2l = 2 \times 0.98 = 2.0\,\mathrm{m}$ (1.96 seen on calculator display)

ii $f_0 = \frac{v}{\lambda_0} = \frac{6.7}{1.96} = 3.4\,\mathrm{Hz}$

b i $\lambda = \frac{2l}{4} = \frac{\lambda_0}{4} = \frac{1.96}{4} = 0.49\,\mathrm{m}$

ii $f = 4f_0 = 4 \times 3.4 = 14\,\mathrm{Hz}$

c The wavelengths would remain the same, but the frequencies would increase because the wave would travel faster if the tension was increased.

Standing waves on strings and ropes are usually between fixed ends, but it is possible that one, or both, boundaries could be ‘free’. If the two ends are free (an unusual event), there will be antinodes at each end / boundary, so that the frequency of the first harmonic will be the same as for fixed boundaries, which has nodes at each end.

◆ **Oscilloscope** An instrument for displaying and measuring potential differences that change with time.

◆ **Waveform** Shape of a wave.

If there is a node at one end and an antinode at the other, the wavelength of the first harmonic will be greater and its frequency will be lower. An example of this situation would be a standing wave produced on a chain hanging vertically. (In that case, the wavelength would not be constant: see Question 4.)

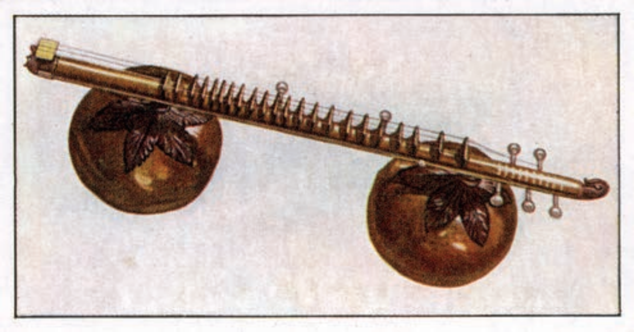
Inquiry 1: Exploring and designing

Designing

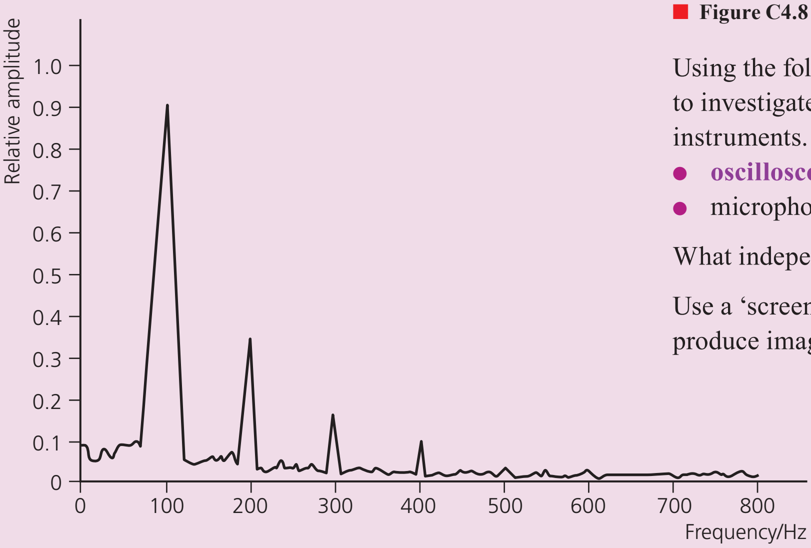
Musical instruments

An amazing variety of musical instruments have been used all around the world for thousands of years (see Figure C4.6). Most involve the creation of standing wave patterns (of different frequencies) on strings, wires, surfaces or in tubes of some kind. The vibrations disturb the air around them and thereby send out sound waves (musical notes).

When musical notes are played on stringed instruments, such as guitars, cellos and pianos, the strings vibrate mainly in their first-harmonic modes, but various other harmonics will also be present. This is one reason why each instrument has its own, unique sound. Figure C4.7 shows a range of frequencies that might be obtained from a guitar string vibrating with a first harmonic of 100 Hz.



■ **Figure C4.6** Vina, an Indian stringed instrument



■ **Figure C4.7** Frequency spectrum from a guitar string

The factors affecting the frequency of the first harmonic are the length of the string, the tension and the mass per unit length. For example, middle C has a frequency of 261.6 Hz. The standing transverse waves of the vibrating strings are used to make the rest of the musical instrument oscillate at the same frequency. When the vibrating surfaces strike the surrounding air, travelling longitudinal sound waves propagate away from the instrument to our ears.

■ **Figure C4.8** Creating standing waves on a cello

Using the following apparatus, design an experiment to investigate the sound waves produced by different instruments.

- **oscilloscope** or app on computer or phone
- microphone.

What independent variables will you change or control?

Use a 'screen capture' technique on the oscilloscope to produce images and compare the **waveforms** produced.

WORKED EXAMPLE C4.2

An experiment similar to that shown in Figure C4.9 can be used to determine the speed of a wave along a stretched string.

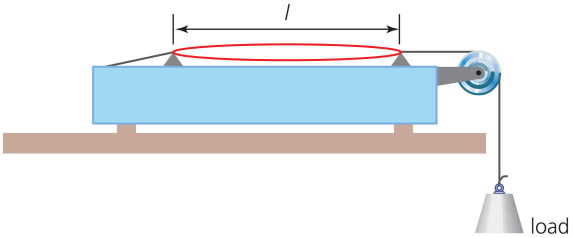
a If the length of the vibrating string was 79.4 cm and the first harmonic had a frequency of 140 Hz, calculate the wave speed.

b Determine the wavelength and frequency of the fifth harmonic.

c A student finds the following formula on the internet:

$$\text{wave speed} = \sqrt{\left(\text{tension} \times \frac{\text{length}}{\text{mass}}\right)}$$

If the mass of the vibrating string was 2.6 g, and the tension was 147 N, show that use of this formula confirms the answer to part **a**.



■ **Figure C4.9** Experiment to determine wave speed

Answer

a $v = f_0\lambda_0 = 140 \times (2 \times 0.794) = 222\,\mathrm{m\,s^{-1}}$

b $5f_0 = 5 \times 140 = 700\,\mathrm{Hz}$

$$\frac{\lambda_0}{5} = \frac{1.588}{5} = 0.318\,\mathrm{m}$$

c $$v = \sqrt{\left(147 \times \frac{0.794}{0.0026}\right)}$$

$$= 2.1 \times 10^2\,\mathrm{m\,s^{-1}}$$

This is within 5% of the answer to part **a**, so the two answers are consistent, within experimental uncertainties.

1 A string on a violin has a length of 32.8 cm and produces a note of 262 Hz (middle C).

a Calculate the speed of the wave on the string.

b State any assumption that you made.

c Suggest what will happen to the speed of the wave on the string and the frequency of the note if sometime later the same string has lost some tension.

2 The velocity of a transverse wave on a string of length 28 cm is 240 m s^{-1}.

a Calculate the frequency of the second harmonic of a standing wave on this string when both ends are fixed.

b Determine the wavelength of the sound that this produces in the surrounding air. Assume the speed of sound in air is 340 m s^{-1}.

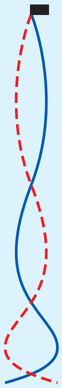
3 A teacher wishes to show his class standing waves on a thin string, similar to those seen in Figure C4.2. He uses a vibration generator set at a frequency of 384 Hz.

a Determine the length of string that is needed to produce the fourth harmonic if the speed of the wave is 285 m s^{-1}.

b Using the same apparatus, the teacher increases the length of the string to exactly 2.00 m. Predict the wavelength and frequency of the third harmonic in this new arrangement.

4 The top of a thin chain hanging vertically is shaken with increasing frequency until the standing wave pattern seen in Figure C4.10 is produced.

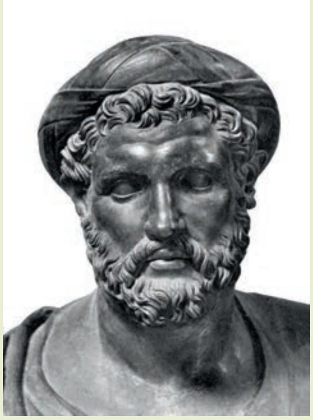
- **a** Suggest why the wavelength on this system decreases towards the lower end of the chain.
- **b** Describe the boundary conditions of this system.
- **c** State which harmonic is shown in the picture.
- **d** Explain why it is not possible to produce the second harmonic on this system.



■ **Figure C4.10** A standing wave on a hanging chain.

Nature of science: Patterns and trends

Pythagoras is often credited with being the first to realize that there was a mathematical relationship between musical notes and the dimensions of the instrument that produced them. That was about 2500 years ago. More generally, this may have been one of the first occasions when mathematical / 'scientific' reasoning was used to describe features of everyday life.



■ **Figure C4.11** Pythagoras

Standing wave patterns of air in pipes

SYLLABUS CONTENT

- ▶ Standing waves patterns in pipes.

Standing longitudinal waves of sound can be created easily in the air contained by a pipe / tube / column. The air may be set into motion by, for example. the simple action of blowing across the open end of the pipe. The sound produced by blowing across the top of an empty bottle is an everyday example. Musical wind instruments, like trumpets or flutes, use the same principle.

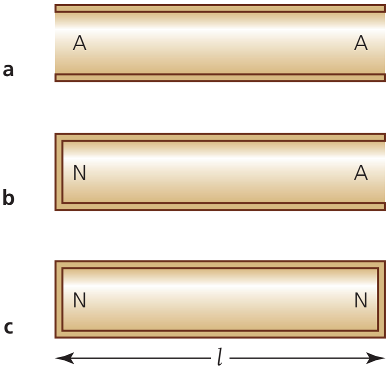
As with strings, in order to understand which wavelengths can be produced, we need to consider the length of the system and the boundary conditions.

The first harmonic is the standing wave with the greatest possible wavelength, λ_0, and lowest possible frequency, f_0. Other harmonics are multiples of this frequency.

After determining possible wavelengths, the equation $v = f\lambda$ can then be used to predict harmonic frequencies if the wave speed is known.

Figure C4.12 show the three possible combinations of boundary conditions.

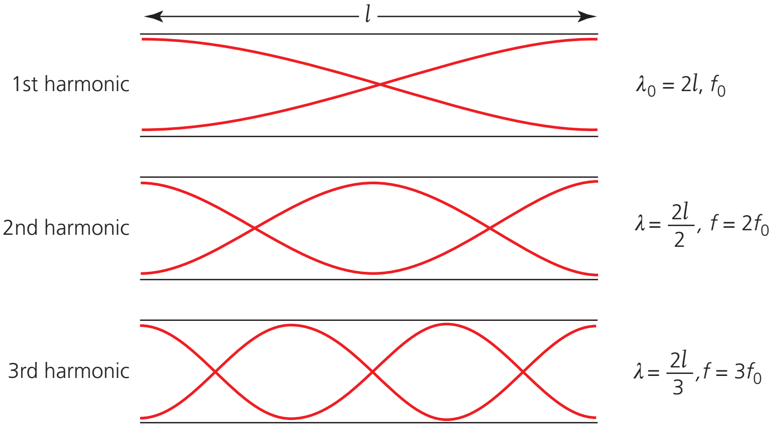
A pipe open at both ends must have antinodes, A, at the ends, and at least one node, N, in between. A pipe open at only one end has one antinode and one node as its boundary conditions.



■ **Figure C4.12** Nodes and antinodes at the ends of open and closed pipes

A pipe closed at both ends (an unusual situation) must have nodes at the ends and at least one antinode in between.

Figure C4.13 graphically represents the first three harmonics (vibration modes) for a pipe open at both ends.

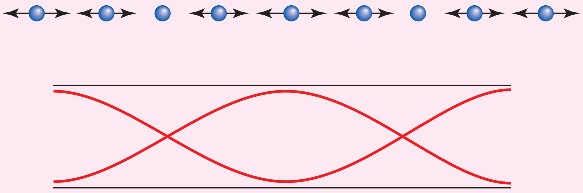


■ **Figure C4.13** The first three harmonics in a pipe open at both ends

The wavelength of the first harmonic (twice the distance between adjacent nodes or antinodes) is 2*l*.

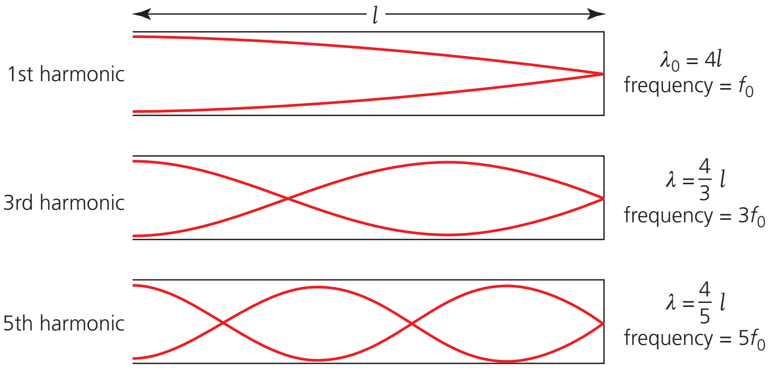
Common mistake

Note that the representations of standing longitudinal waves seen in Figure C4.13 and Figure C4.15 may cause confusion: the curved lines in the diagrams are an indication of the maximum *sideways* displacement of vibrating air molecules. They should not be mistaken for transverse waves, like those on a string. Figure C4.14 shows how the second harmonic shown in Figure C4.13 is representing the movements of some molecules.



■ **Figure C4.14** How the second harmonic can represent the movements of some molecules

For a pipe which is closed at one end, but open at the other, the possible standing waves are shown in Figure C4.15. The first harmonic has a longer wavelength and lower frequency than for a similar length pipe which is closed, or open, at both ends. Note that only odd-numbered harmonics are possible with these boundary conditions.



■ **Figure C4.15** Harmonics in a pipe open at one end

One way of demonstrating standing sound waves in air is by using a small loudspeaker attached to the end of a long, clear plastic tube, which is closed at the other end. See Figure C4.16. Electrical signals of different frequencies can be applied to the loudspeaker which then sends sound waves of the same frequencies down the tube. When the waves reflect back off the end of the tube, a standing wave can be set up only *if* the frequency is equal to the frequency of the one of the possible harmonics. Some powder can be scattered all along the pipe and when the loudspeaker is turned on and the frequency carefully adjusted, the powder is seen to move into separate piles. This is because the powder tends to move from places where the vibrations of the air are large (antinodes) to the nodes, where there are no vibrations. The tube may be considered as closed at both ends.

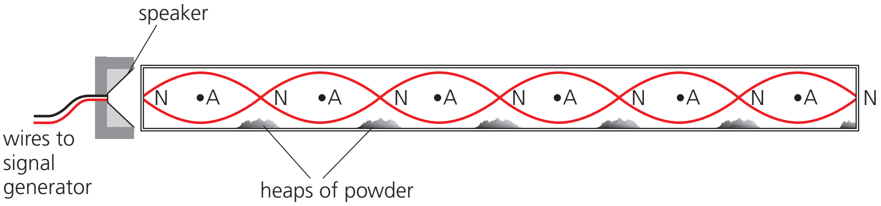


Figure C4.16 Demonstrating a standing wave with a loudspeaker

WORKED EXAMPLE C4.3

Consider Figure C4.16.

a State which harmonic is shown in Figure C4.16.

b Calculate the wavelength of this standing wave if the length of the tube is 73.5 cm.

c Determine the speed of the sound wave in air if the frequency used by the loudspeaker was 1410 Hz.

d Calculate the theoretical frequencies of the first two observable harmonics if the tube was left open at the end on the right-hand side.

Answer

a sixth

b 73.5/3 = 24.5 cm

c $v = f\lambda = 1410 \times 0.245 = 345\,\text{m}\,\text{s}^{-1}$ (345.45 seen on calculator display)

d See Figure C4.15. The wavelength of the first harmonic will be 4 × length of tube = 2.94 m.

$$f_0 = \frac{v}{\lambda} = \frac{345.45}{2.94} = 1.18 \times 10^2\,\text{Hz} \text{ (117.5 seen on calculator display)}$$

The second harmonic does not occur.

Frequency of third harmonic = $3f_0 = 3 \times 117.5 = 353$ Hz

Figure C4.17 shows another way of investigating standing waves in pipes. The length of the pipe is easily changed by raising or lowering it in the tall container of water. The pipe is always open at one end and closed at the other. Air in the pipe is disturbed by the pure, single frequency emitted from a **tuning fork** (see Figure C4.20) placed just above the open end of the tube, and the length of the pipe is adjusted until the sound of the standing wave becomes audible.

◆ **Tuning fork** Device designed to vibrate at only one precise frequency.

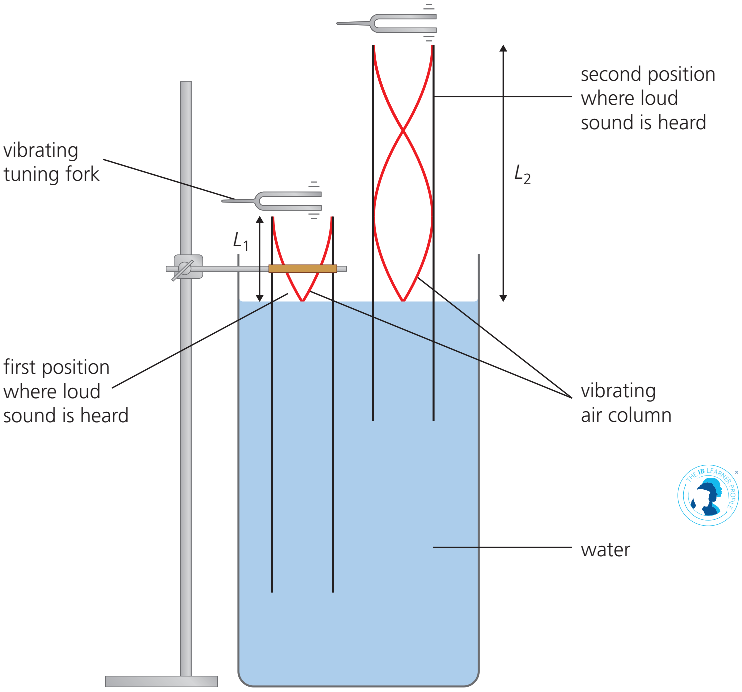


Figure C4.17 Demonstrating standing waves with a tuning fork

The length of the tube seen on the left of the Figure C4.17 has been adjusted to the shortest length for which a standing wave can be heard. This will be the first harmonic for a tube of that length. The position on the right corresponds to the third harmonic for a tube of the new length. (The speed, wavelength and frequency of the wave are unchanged.) There is no second harmonic possible with these boundary conditions (open at one end, closed at the other). Higher harmonics may be heard if the tube and container are long enough.

ATL C4A: Social skills

Working collaboratively

Suggest how physics and music students could work together to produce a presentation called 'The Science of Music' for the rest of their year group.

Consider how you could share your presentations online for future students.

For these questions, where necessary, assume that the speed of sound in air is 342 m s^{-1}.

5 a If the tuning fork in Figure C4.17 had a frequency of 384 Hz, show that the length of the pipe needed for the first harmonic to be heard is about 20 cm.

b How far will the pipe need to be raised until the next harmonic is heard?

6 Sketch the first three harmonics for air in a pipe which is closed at both ends.

7 A pipe has a length of 1.32 m and is closed at both ends. Determine the wavelength and frequency of its third harmonic.

8 The flute is the oldest of musical instruments. See Figure C4.18. There are a very large number of designs. It can be considered as a pipe which is open at both ends. Sound is produced when air is blown across an opening near the end of the pipe.

a Determine what length (cm) of pipe will produce a first harmonic of frequency 493 Hz.

b Explain the purpose of the holes along the length of the flute.

c A clarinet is a similar type of musical instrument to a flute, but the pipe is closed at one end. Compare the length of the pipes of a flute and clarinet that produce musical notes of the same frequency.

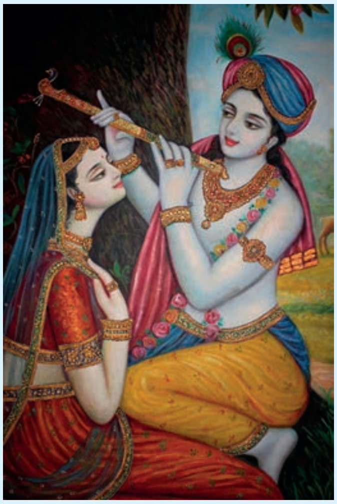


Figure C4.18 Radha listening to Krishna's flute

9 Apparatus similar to that shown in Figure C4.17 was used to investigate the variation of the speed of sound with air temperature. At a temperature of 30 °C, adjacent nodes were found to be separated by a distance of 23.3 cm when using a frequency of 750 Hz.

a Determine the speed of sound at this temperature.

b Give a molecular explanation of why this speed is greater than the 342 m s^{-1} used in other questions.

10 Water being poured into a bottle may produce many sounds, but there will usually be a noticeable increase in frequency of the sound as the bottle fills up. Explain this effect.

TOK

The natural sciences

- What is the role of imagination and intuition in the creation of hypotheses in the natural sciences?

When electrons were discovered in 1897, they were believed to be tiny solid particles. Twenty-seven years later, it was proved that electrons have wave properties (see Theme E). This led on to the theory that electrons could only be confined within atoms in the form of standing waves.

Most people will struggle to visualize electrons as three-dimensional standing waves in unimaginably small atoms. But, after about one hundred years, this still remains the accepted theory. There is no good reason for us to believe that the atomic-scale particles must behave in ways that are the same as masses which we can see with our eyes.

Natural frequencies of vibration

SYLLABUS CONTENT

- The nature of natural frequency.
- The effects of light, critical and heavy damping on the system.

◆ **Frequency, natural** The frequency at which a system oscillates when it is disturbed and then left to oscillate on its own, without influence from outside.

◆ **Vibration** Mechanical oscillation (usually of relatively small amplitude).

When many objects are struck briefly by an external force, they vibrate ‘freely’, or ‘naturally’ (although, for most objects, the vibrations may be insignificant and/or very short-lasting because energy is quickly dissipated into the surroundings). Such vibrations often disturb the air around them and send longitudinal waves into the environment, which may be heard as sound, if they have a suitable frequency.

When an object is disturbed and then left to oscillate without further interference, it is said to oscillate at its **natural frequency** (or frequencies).

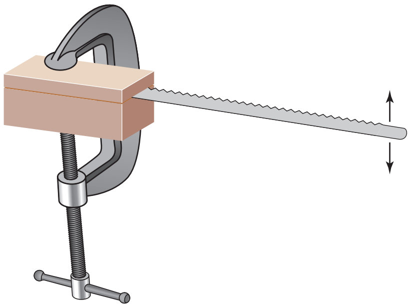


Figure C4.19 Vibrating hacksaw blade

Relatively small amplitude *mechanical* oscillations are commonly called **vibrations**.

The simplest examples of natural frequencies are those of a simple pendulum and mass on a spring, as discussed in Topic C.1. A further example is a clamped ruler, or hacksaw blade, as shown in Figure C4.19.

Vibrating objects will oscillate at a natural frequency(s) which depends on their dimensions and masses.

Tool 3: Mathematics

Linearize a graph

A student has read that the square of the time period of a hacksaw blade oscillator is proportional to the cube of the vibrating length.

What graph, involving frequency, should she draw in order to see if the relationship is correct?

An object made of only one material in a simple shape, such as a tuning fork (see Figure C4.20), may produce a single, ‘pure’, natural frequency. But most objects will have more complicated structures and a range of natural frequencies, although one frequency may dominate.

The *two-dimensional* standing wave patterns of a horizontal metal plate can be observed by placing small grains, such as fine sand or salt, onto a surface that is disturbed into vibration at the plate's natural frequencies. See Figure C4.21. (Oscillations can be easily maintained using a mechanical oscillator, driven by a signal generator, to vibrate the plate.)

Figure C4.20 This tuning fork produces a frequency of exactly 440 Hz (a musical A)

Figure C4.21 Demonstrating the oscillations of a metal plate

Damping

The motions of all macroscopic objects have resistive forces of one kind or another acting against them. Resistive forces will always act in the opposite direction to the instantaneous motion of an oscillator, and result in a reduction of its speed and kinetic energy.

Therefore, as with all other mechanical systems, useful energy is transferred from an oscillator into the surroundings (dissipated) in the form of thermal energy and maybe some sound. Consequently, an oscillator will move at slower and slower speeds, and its successive amplitudes will decrease in size. This effect is called damping.

Damping is the dissipation of energy from an oscillator due to resistive forces.

Top tip!

The magnitude of each successive peak of the graph shown in Figure C4.22 can be determined by multiplying the previous value by the same fraction. For example: $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8} \ldots$

(This is known as an exponential series, or a geometric sequence.)

It is common for the frequency (and time period) of a vibration to remain approximately constant during damping, as shown by the graph in Figure C4.22. This is because, although the displacements are reduced, the speeds and accelerations also decrease.

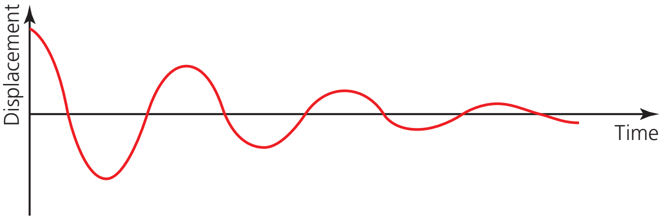


Figure C4.22 Decreasing amplitude (at constant frequency) of a damped oscillation

Tool 3: Mathematics

Propagate uncertainties in processed data

Successive amplitudes (in cm) of the **free vibration** of a hacksaw blade were measured to be 2.7, 2,4, 2.1, 1.8, 1.6, 1.4, 1.3, 1.1. The measurements were made to the nearest millimetre. Considering uncertainties in measurement, is it possible that this data fits an exponential pattern?

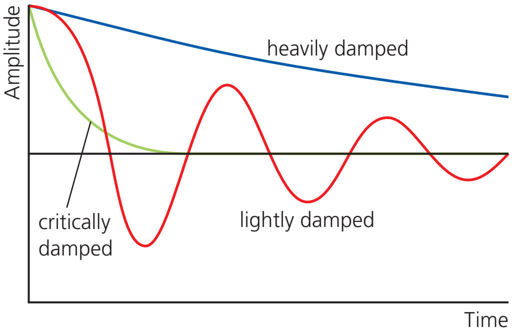


Figure C4.23 Light damping, heavy damping and critical damping

Damping can be investigated experimentally using simple apparatus like that shown in Figure C4.19, but with horizontal cards of different areas taped to the blade in order to increase air resistance.

The amount (degree) of damping in oscillating systems can be very different, as shown in Figure C4.23.

Some oscillations are *heavily damped* because of considerable frictional forces. In effect no oscillations occur because resistive forces are such that the object takes a long time (compared to its natural time period) to return to its equilibrium position.

Conversely, an oscillator may be *lightly damped*, so that it continues to oscillate, taking a relatively long time to dissipate its energy. A pendulum and a mass oscillating on a spring are good examples of lightly damped systems.

If an oscillation is opposed by resistive forces, such that it settles relatively quickly (compared to its natural time period) back into its equilibrium position, without ever passing through it, the process is described as **critical damping**. A car's suspension (see Figure C4.24) is an example of this kind of damping.

◆ **Vibration (free)** Vibration without any external influence.

◆ **Damping (critical)** When an oscillating system returns relatively quickly to its equilibrium position without oscillating.

Figure C4.24 A car's shock absorber

11 Describe how the natural frequency of
- **a** a simple pendulum,
- **b** a mass hanging vertically from a spring

can be increased.

12 A student was investigating the vibrations of a hacksaw blade, as shown in Figure C4.19. She displaced the end of the blade and then left it to vibrate freely, but she found that the vibrations were too quick for her to observe easily.
- **a** Suggest how she could decrease the frequency of the vibrations.
- **b** The blade then vibrated with a frequency of 1.0 Hz, while its amplitude was (almost) constant, at 0.50 cm, for several seconds.
  Sketch a displacement–time graph for the first two seconds of its motion.
- **c** The vibrations were then damped.
  Add a second line to your sketch for part **b**, to represent the damped oscillations.
- **d** Suggest how the student could have damped the vibrations.

13 Outline why the fine sand shown in Figure C4.21 moves into places which demonstrate the standing wave pattern on the plate.

14 Figure C4.25 shows an automatic door closer.
- **a** Describe its intended purpose.
- **b** What type of damping does it exhibit?
- **c** Give an example of where it might be used.

Figure C4.25 Automatic door closer

◆ **Vibration (forced)** Vibration affected by external periodic forces.

Forced vibrations

A **forced vibration** (oscillation) occurs when an external oscillating (periodic) force acts on a system. This may tend to make it oscillate at a frequency which is different from its natural frequency.

■ **Figure C4.26** How can we increase the amplitude of a swing?

We are surrounded by a range of oscillations. It is important to consider how these oscillations affect other things around them. In other words, what happens when an external oscillating force is continuously applied to a separate system?

To understand this, it is helpful to consider a very simple example: what happens when we keep pushing a child on a swing (see Figure C4.26)?

In this case it is fairly obvious: it depends on when we push the swing and in which direction. If we want bigger swings (increasing amplitudes), then we should push once every oscillation in the direction in which the child is moving at that moment. In more scientific terms, we would say that we need to apply an external force that has the same frequency as, and is in phase with, the natural frequency of the swing.

The most important examples of forced oscillations are those in which the frequency of the external force (**driving frequency**) is the same as the natural frequency. The child on the swing described above is a good example of this. When a regular periodic stimulus to a system results in an increasing amplitude the effect is called resonance.

◆ **Frequency, driving** The frequency of an oscillating force (periodic stimulus) acting on a system from outside. Sometimes called forcing frequency.

◆ **Resonance** The increase in amplitude that occurs when an oscillating system is acted on by an external periodic force that has the same frequency as the natural frequency of the system. The driving force must be in phase with the natural oscillations of the system.

◆ **Frequency–response graph** Graph used to show how the amplitude of a system's oscillations responds to different driving frequencies.

◆ **Resonant frequency** The frequency at which resonance occurs.

Resonance

SYLLABUS CONTENT

- ▶ The nature of resonance including the amplitude of oscillations based on driver frequency.
- ▶ The effect of damping on the maximum amplitude and resonant frequency of oscillation.

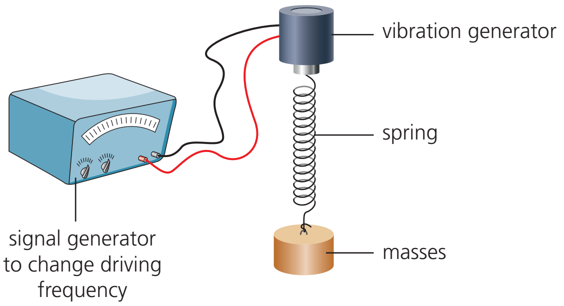
Resonance is the name given to the increase in amplitude and energy of an oscillation that occurs when a periodic external driving force has the same frequency as the natural frequency of a system.

The oscillations of the driving force must be in phase with the natural oscillations of the system.

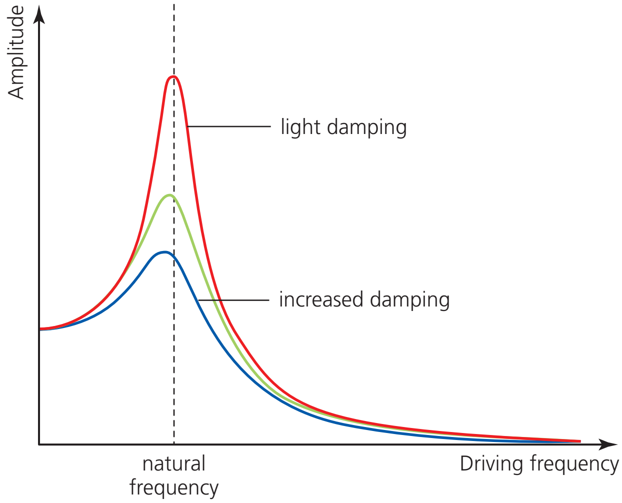
Simple quantitative laboratory experiments into the effects of resonance can be difficult to perform, but they can produce interesting results that show how varying the driving frequency affects the amplitude of an oscillating system. When the force is first applied, the oscillations may be erratic, but the system will settle into a regular pattern of movement with a measurable maximum amplitude.

Figure C4.27 shows a possible arrangement. The resonant frequency of the mass–spring system can be changed by using different springs and/or different masses. The driving frequency is provided by the vibration generator, which can be driven using different frequencies from the signal generator.

A typical **frequency–response graph** drawn from the results of an experiment like that shown in Figure C4.28 rises to a maximum amplitude at the **resonant frequency**. This occurs when the rate of energy dissipation (damping) has risen to a level that is equal to the power supplied from the source of the driving frequency.



■ **Figure C4.27** Investigating the resonance of a mass on a spring



■ **Figure C4.28** Typical frequency–response curves with different degrees of damping

The resonant frequency is at, or very close to, the natural frequency, but the sharpness and height of the peak also depend on the amount of damping in the system. The greater the damping, the greater the dissipation of energy and, therefore, the smaller the amplitude. The value of the resonant frequency reduces slightly with greater damping.

If there is a powerful input, or very little damping, amplitudes of vibration can become large and this may have destructive, or useful, consequences.

The energy of an oscillation is proportional to its amplitude *squared*.

There may be smaller resonance peaks at values of the driving frequency which are equal to the natural / resonant frequency divided by 2, 3, 4 and so on (not shown in the diagram).

LINKING QUESTIONS

- How does the amplitude of vibration at resonance depend on the dissipation of energy in the driven system?
- How can resonance be explained in terms of conservation of energy?

These questions link to understandings in Topic A.3.

Inquiry 2: Collecting and processing data

Processing data

Table C4.2 shows the results obtained in an experiment similar to that seen in Figure C4.27. The uncertainties in this experiment were significant.

Draw a graph of these results, including uncertainty bars, with a curve of best fit to determine a value for the resonant frequency. You should not assume that the largest recorded measurement is the peak of the graph.

■ Table C4.2 Results of resonance experiment

Applied frequency / Hz ± 2 Hz	Maximum amplitude / cm ± 0.5 cm
4	1.7
6	2.1
8	2.7
10	4.9
12	5.4
14	3.3
16	2.5
18	1.9

Examples of resonance

Some examples of resonance are useful but many more are unwanted, and we usually try to reduce their destructive effects. Avoiding resonance in all types of structure is a major concern for engineers and it provides an interesting combination of physics theory and practical engineering.

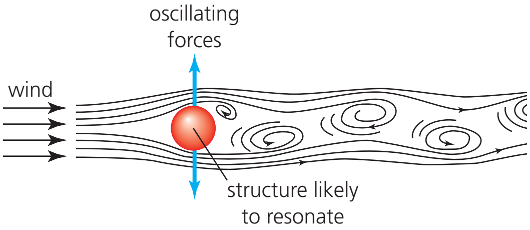
Unwanted resonance

Parts of almost all machinery (and their immediate surroundings) may vibrate destructively when their motors are operating at certain frequencies. For example, a washing machine may vibrate strongly when the spinner is running at a certain frequency, and parts of vehicles can vibrate when the engine reaches a certain frequency, or they travel at certain speeds.

Earthquakes may well affect some buildings more than others. The buildings that are most damaged are often those that have natural frequencies closest to the frequencies of the waves produced by the earthquake (Figure C4.29).

■ **Figure C4.29** Resonance may be one reason why some buildings collapse in an earthquake

Strong but steady winds, or currents, can also cause dangerous resonance in structures such as bridges and towers. This is often due to the effect of eddies and vortices as the wind or water flows around the structures. See Figure C4.30.



■ **Figure C4.30** A steady wind can cause oscillating forces because of the alternate way in which vortices can be formed

If you have ever crossed a small suspension bridge for walkers (such as the one in Figure C4.31), you will probably know how easy it is to set it vibrating with increasing amplitude by shaking it or stamping your feet at a certain frequency. This is because it would be too difficult or expensive to build such a simple bridge with a natural frequency that is very different from a frequency that people can easily reproduce, or to use a design that incorporated damping features.

■ **Figure C4.31** Walker on a suspension bridge in Nepal

The resonance of bridges has been well understood for many years and the flexibility of suspension bridges makes them particularly vulnerable. The famous collapse of the newly built Tacoma Narrows Bridge in the USA in 1940 is widely given as a simple example of resonance caused by the wind, although this is only part of a much more complex explanation. Videos of the collapse are easily found on the internet. In June 2000, the Millennium Bridge across the River Thames in London had to be closed soon after its opening because of excessive lateral (sideways) oscillations due to resonance (Figure C4.32).

In this case positive feedback was important. The slow oscillations of the bridge made people sway with the same frequency, and their motion simply increased the periodic forces on the bridge that were causing resonance. The problem was solved by adding energy-dissipating dampers, but it was about 18 months before the bridge could reopen.

To reduce the risk of damage from resonance engineers can:

- alter the shape of the structure to change the flow of the air or water past it
- change the design so that the natural frequencies are not the same as any possible driving frequencies – this will involve changing the stiffness and mass of the relevant parts of the structure
- ensure that there is enough damping in the structure and that it is not too rigid, so that energy can be dissipated.

■ **Figure C4.32** The Millennium Bridge in London was affected by resonance

Useful resonance

- The molecules of certain gases in the atmosphere oscillate at the same frequency as thermal radiation emitted from the Earth. These gases absorb energy because of resonance; this results in the planet being warmer than it would be without the gases in the atmosphere. This is known as the greenhouse effect, as discussed in Topic B.2.
- Microwave ovens use electromagnetic radiation in the microwave region. The microwave wavelength equals a vibrational frequency of water molecules, so that the molecules absorb the radiation.
- Your legs can be thought of as pendulums with their own natural frequency. If you walk with your legs moving at that frequency, energy will be transferred more efficiently and it will be less tiring (we tend to do this without thinking about it).
- Quartz crystals can be made to resonate using electronics – the resulting oscillations are useful in driving accurate timing devices such as watches and computers.
- The sound from musical instruments can be amplified if the vibrations are passed on to a supporting structure that can resonate at the same frequency. An obvious example would be the strings on a guitar causing resonance in the box on which they are mounted. Because the box has a much larger surface area it produces a much louder sound than the string alone.
- Magnetic resonance imaging (MRI) is a widely used technique for obtaining images of features inside the human body. Electromagnetic waves of the right frequency (radio waves) are used to change the spin of protons (hydrogen nuclei) in water molecules in the patient's body.

LINKING QUESTION

- How can the idea of resonance of gas molecules be used to model the greenhouse effect? (NOS)

This question links to understandings in Topic B.2.

15 Consider the oscillations of a mass on a spring, as shown in Figure C1.5.

a Use the formula from Topic C.1 to determine the natural frequency of a 740 g mass oscillating vertically on a spring which has a spring constant of $17.2\,\mathrm{N\,m^{-1}}$.

b Sketch a frequency–response curve to show how the amplitude of oscillation could change when the applied frequency from the vibration generator increases from 0.5 Hz to 2.0 Hz.

c Discuss how the results of the experiment will change if the mass was placed in water (in a beaker). Illustrate your answer by adding a second curve to your sketch for part **b** of this question.

16 a Estimate the natural frequency of your leg when it swings freely like a pendulum.

b If, when walking, your leg moves with the same frequency, predict your approximate speed, in $\mathrm{km\,h^{-1}}$.

17 a Use the internet to find out a typical frequency of waves generated by an earthquake.

b Suggest ways in which an architect / civil engineer can ensure that their designs do not resonate dangerously at that frequency.

18 It is claimed that an opera singer can shatter a wine glass using sound resonance. Research the internet for any video evidence of this effect. Quote your conclusions and sources.

19 A wing mirror on a car resonates at multiples of its natural frequency of 20 Hz.

a Sketch a graph to show the frequency response of the mirror as the rpm (revolutions per minute) of the car engine increase from 1000 to 4000.

b Suggest how the vibrations of the mirror could be reduced.

c Add a second curve to your graph to show the new frequency response.

20 Carbon dioxide gas in the Earth's atmosphere, CO_2, is an important cause of the greenhouse effect. In a simplified model, the molecule may be visualized as a simple harmonic oscillator, with the two oxygen atoms oscillating to and from a central carbon atom, as shown in Figure C4.33.

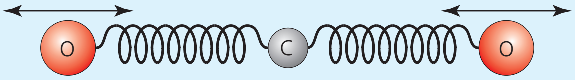


Figure C4.33 Oscillations of a carbon dioxide molecule

a If each oxygen molecule has a mass of $2.7 \times 10^{-27}\,\mathrm{kg}$, use the formula for the time period of a mass on a spring, with $k = 530\,\mathrm{N\,m^{-1}}$, to determine a value for a resonant frequency of the carbon dioxide molecule.

b In what part of the electromagnetic spectrum are waves of this frequency to be found?

C.5 Doppler effect

Guiding questions

- How can the Doppler effect be explained both qualitatively and quantitatively?
- What are some practical applications of the Doppler effect?
- Why are there differences when applying the Doppler effect to different types of waves?

What is the Doppler effect?

SYLLABUS CONTENT

- The representation of the Doppler effect in terms of wavefront diagrams when either the source or the observer is moving.

◆ **Doppler effect** When there is relative motion between a source of waves and an observer, the emitted frequency and the received frequency are not the same.

The **Doppler effect** is the name given to a phenomenon that is observed when there is relative motion between a source and an *observer* of waves. The term 'observer' is being used here to represent the person, or the device, receiving waves (of any type). The Austrian physicist Christian Doppler first identified in 1842 the effect that bears his name.

> When there is relative motion between a source of waves and an observer, the waves received by the observer will have a different frequency (and wavelength) than the waves emitted by the source. This is called the Doppler effect.

The easiest way to explain the Doppler effect is by drawing wavefronts. Figure C5.1a shows the common situation in which a stationary source, S, emits waves that travel towards a stationary observer, O, with the same speed in all directions. Figure C5.1b shows an observer moving directly towards a stationary source and Figure C5.1c shows a source moving directly towards a stationary observer.

Similar diagrams can be drawn to represent the situations in which the source and detector are moving apart. The Doppler effect may be better understood by observing computer simulations which show *moving* sources or observers.

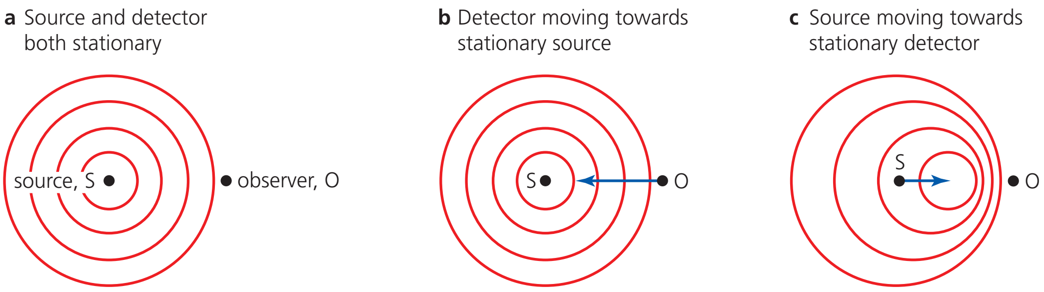


Figure C5.1 Wavefront diagrams to demonstrate the Doppler effect

The observer in Figure C5.1b will meet more wavefronts in a given time than if it remained in the same place, so that the received frequency, f', is greater than the emitted frequency, f.

Since $\lambda = \frac{v}{f}$ (from Topic C.2) and the wave speed, v, is constant, the received wavelength will be less than the emitted wavelength.

In Figure C5.1c, the distance between the wavefronts (the wavelength, λ) between the source and the observer is reduced, which again means that the received frequency will be greater than the emitted frequency. ($f = \frac{v}{\lambda}$ and the wave speed, v, is constant.)

The most common everyday examples of the Doppler effect are with sound, but the effect is usually only noticeable if the sound is loud and the movement is fast, from a moving vehicle, for example.

Doppler effect for sound waves

SYLLABUS CONTENT

- The nature of the Doppler effect for sound waves.

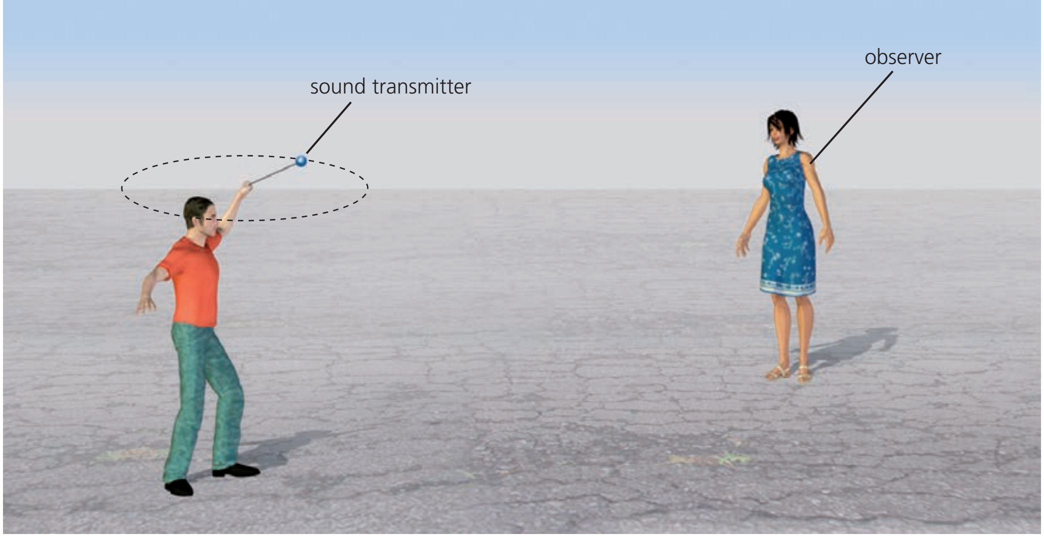


Figure C5.2 Demonstrating the Doppler effect with sound waves

Figure C5.2 shows a way in which the Doppler effect with sound can be demonstrated. A small source of sound (of a single frequency) is spun around in a circle. When the source is moving towards the observer a higher frequency is heard by the observer; when it is moving away, a lower frequency is heard.

The speed of sound through air depends only on the physical properties of the air. It does not vary with the motion of the source, or the observer. However, the speed with which sound *passes* a moving observer depends on their relative speeds. For example, a sound travelling away from a stationary source at 340 m s^{-1}, will pass an observer who is moving directly away from the source at 200 m s^{-1}, with a speed of 140 m s^{-1}.

The most common and easiest understood examples of the Doppler effect for sound include trains or cars which are moving quickly at a constant speed in an approximate straight line towards, or away from, a stationary observer.

Common mistake

Ambulances and police sirens are often given as examples of the Doppler effect, which they are. But be aware that they also emit sounds of varying frequency and loudness which should not be confused with the Doppler effect itself.

■ **Figure C5.3** These bats in Malaysia use the Doppler effect to navigate

Many types of bats use the Doppler effect. See Figure C5.3.

The Doppler effect for sound is covered in more mathematical detail later in this topic at the Higher Level.

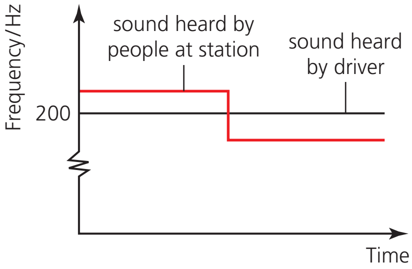
WORKED EXAMPLE C5.1

A train travelling at a constant speed approaches a station where it will not stop. The driver of the train sounds a warning horn, of frequency 200 Hz, as the train approaches and then passes through the station. Sketch graphs (on the same axes) to show how the frequency heard by:

a the driver varies with time

b the people at the station varies with time.

Answer

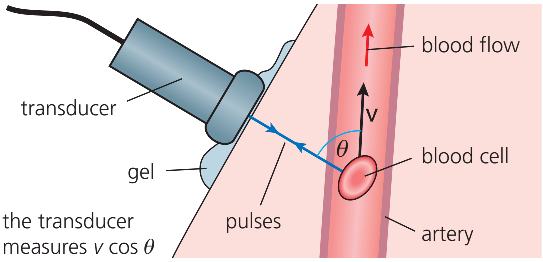


■ **Figure C5.4** There is a sudden frequency change as the train passes the people at the station

Top tip!

When the motion of the source is directly towards, or away from, the observer (or the other way around), there is *sudden* change of frequency when they pass. But the change of frequency is gradual if they do not pass close to each other.

The measurement of the rate of blood flow in an artery is an application of the Doppler effect for sound (ultrasound), which is shown in Figure C5.5.



■ **Figure C5.5** Measuring blood flow rate using the Doppler effect

Pulses of ultrasonic waves are sent into the body from the transducer and are reflected back from blood cells flowing in an artery. (A transducer is a general term used to describe any device which converts another form of energy into, or from, electrical energy.) The received waves have a different frequency because of the Doppler effect, and the measured change of frequency can be used to calculate the speed of the blood flowing in the artery. This information can be used by doctors to help to diagnose many medical problems. Because the waves usually cannot be directed along the line of blood flow, the calculated speed will be the component ($v\cos\theta$).

1 Draw diagrams similar to Figure C5.1b and c to represent the wavefronts when:

- **a** an observer is moving directly away from a stationary source
- **b** a source is moving directly away from a stationary observer.

2 A hospital patient had a Doppler ultrasound scan to check the blood flow in an artery, as shown in Figure C5.5. The transducer used a frequency of 6.0 MHz, and a healthy blood flow rate was expected to be about 10 cm s^{-1}.

- **a** State whether the detected frequency will be higher, or lower, than 6.0 MHz. Explain your answer.
- **b** **i** Explain why a calculation based only on the frequency change predicts that the blood speed in an artery is lower than its true value.
  - **ii** If the patient has a medical problem affecting blood flow rate, predict what effect this will have on the frequency measurements.
- **c** Suggest what properties of ultrasound make it useful for this medical examination.

Tool 3: Mathematics

Sketch graphs, with labelled but unscaled axes, to qualitatively describe trends

A ‘sketch’ graph should be drawn neatly, using a ruler where appropriate. The axes should be clearly labelled, with zeros and negative values indicated (if applicable). Usually there is no requirement to add scales, or numerical values to the axes, but important features should be labelled.

3 Sketch a frequency–time graph to show how the sound heard by the observer in Figure C5.2 varies during one oscillation of the sound transmitter.

4 Figure C5.6 shows the shape of waves spreading from the front of a boat (its bow).

- **a** Describe the overall shape of the wave fronts.
- **b** Explain why bow waves similar to these can cause damage to the surroundings if the boat is close to shore.

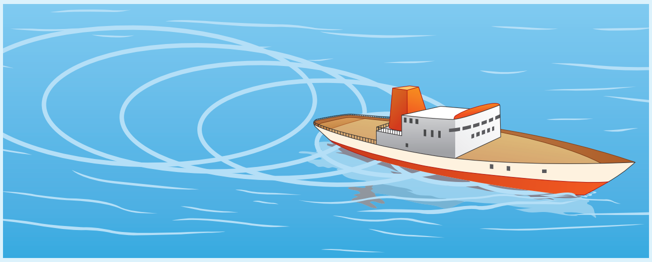


Figure C5.6 The shape of waves spreading from the front of a boat

Nature of science: Global impact of science

Shock waves: breaking the sound barrier

As an object, like an aircraft, flies faster and faster, the sound waves that it makes get closer and closer together in front of it. When an aircraft reaches the speed of sound, at about 1200 km h^{-1}, the waves superpose to create a ‘shock wave’. This is shown in Figure C5.7.

When an aircraft reaches the speed of sound it is said to be travelling at ‘Mach 1’ (named after the Austrian physicist, Ernst Mach). Faster speeds are described as ‘supersonic’ and twice the speed of sound is called Mach 2, and so on. As Figure C5.8 shows, the shock wave travels away from the side of the aircraft and can be heard on the ground as a ‘sonic boom’.

For many years some engineers doubted if the sound barrier could ever be broken. The first confirmed supersonic flight (with a pilot) was in 1947. Now it is common for military aircraft to travel faster than Mach 1. Concorde and Tupolev 144 were the only supersonic passenger aircraft in regular service, but their use has been discontinued. There are plans to introduce a new supersonic passenger aircraft before the year 2030.

Choose search terms to find out about the planned designs for future passenger / commercial supersonic aircraft.

Explore the reasons why previous supersonic airliners were discontinued. Were these reasons technological, scientific, economic, environmental, political? Explain your reasoning.

It is possible to use a whip to break the sound barrier. If the whip gets thinner towards its end, then the speed of a wave along it can increase until the tip is travelling faster than sound (in air). The sound it produces is often described as a whip ‘cracking’.

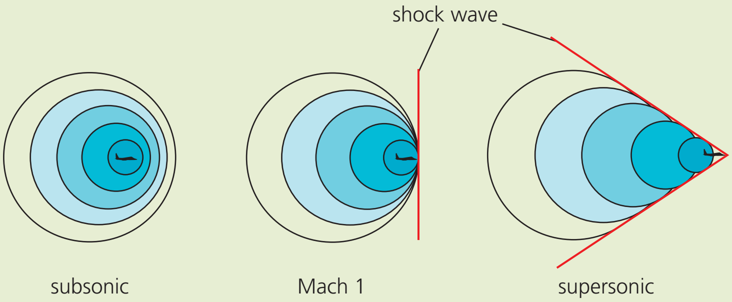


Figure C5.7 Creating a shock wave in air

Figure C5.8 Aircraft breaking the sound barrier

LINKING QUESTION

- What are the similarities and differences between light and sound?

This question links to understandings in Topic C.2.

Doppler effect for light and other electromagnetic waves

SYLLABUS CONTENT

- The nature of the Doppler effect for electromagnetic waves.
- The relative change in frequency, or wavelength, observed for a light wave due to the Doppler effect where the speed of light is much larger than the relative speed between the source and the observer, as given by: $\frac{\Delta f}{f} = \frac{\Delta\lambda}{\lambda} \approx \frac{v}{c}$.

Apart from mechanical waves like sound, the Doppler effect also occurs with electromagnetic waves, but the situation is more complicated because the speed of electromagnetic waves, as measured by any observer, is always the same: it is unaffected by the speed of the source, or the speed of the observer (this is a relativistic effect – see Topic A.5 – however, this theory is not of concern here).

The following equation for the change (shift) in frequency, Δf (= received frequency – emitted frequency), or shift in wavelength, $\Delta\lambda$, can be used if the relative speed between source and observer, v, is very much less than the speed of the electromagnetic waves, c ($v << c$). This is usually a valid assumption because of the very high value of c.

$$\frac{\Delta f}{f} = \frac{\Delta\lambda}{\lambda} \approx \frac{v}{c}$$

Using the Doppler effect to determine speeds

Microwaves of known frequency, f, are easily produced and transmitted as a beam which can be directed at a moving object. A small fraction of the radiation will be arrive back at the source after reflecting off the object.

◆ **Radar** A system which uses microwaves to detect the distance, direction and speed of moving objects.

The change in frequency, Δf, due to the Doppler effect can be used to determine the speed of an object.

Figure C5.9 Radar dish

Examples include:

- Speed 'guns' are used by police for checking the speed of moving vehicles.
- **Radar** is used for monitoring aircraft, or boat movements. (See below for an explanation of radar.)
- Radar is used for tracking the movement of storms.

In all these examples, the microwaves travel from the source to the object and then back to the source. For this reason, a factor of 2 should be added to the right-hand side of the equation above.

Radar (**RA**dio **D**etection **A**nd **R**anging) is a system used for determining the direction, distance and speed of an aircraft (or other object). Pulses of microwaves are sent from a rotating aerial (see Figure C5.9 for an example). After a very small fraction of a second, some microwaves which were reflected off the aircraft are received back at the aerial. The time delay can be used to determine the distance to the aircraft and the orientation of the aerial provides information about the direction to the aircraft.

The use of the Doppler effect with the radar also enables a direct calculation of aircraft speed and has the advantage of being able to ignore reflections from objects which are not moving or are only moving slowly.

WORKED EXAMPLE C5.2

a Calculate the change of frequency which will be detected by a police speed gun (see Figure C5.10) using a frequency of 20.6 GHz, when it is directed at a car moving directly away at a speed of 130 km h^{-1} (36.1 m s^{-1}).

■ **Figure C5.10** A police speed check

b State what change of frequency would be detected if the vehicle was moving directly *towards* the speed gun, at the same speed.

Answer

a $\frac{\Delta f}{f_0} = \frac{2v}{c}$

$$\Delta f = 2 \times 36.1 \times \frac{20.6 \times 10^9}{3.00 \times 10^8} = 4.96 \times 10^3 \text{ Hz (decrease)}$$

b The same magnitude as in part **a**, but the frequency would increase, rather than decrease.

ATL C5A

Research and communication skills

Use a variety of internet websites to learn about *Doppler weather radar*.

Evaluate your sources for reliability.

Plan a short presentation of the key information for other IB physics students.

5 The speed limit in a town is 50 km h^{-1} (13.9 m s^{-1}). In a safety check by the police, using the Doppler effect, a frequency increase of 2.85 kHz was detected from a car moving along a straight road.

a State whether the car was moving towards, or away from, the check point.

b If the speed gun used a frequency of 32.8 GHz, show that the car was travelling slower than the legal limit.

6 An airport radar system using microwaves of frequency 2.72 GHz sends out a pulse of waves that is reflected off an aircraft that is within its control area.

A reflected signal is received back at the airport 1.374×10^{-4} s later, at a frequency of 1050 Hz more than the emitted wave. At that time the aerial was pointing exactly north.

a Determine the distance between the aircraft and the radar aerial.

b **i** Use the equation:

$$\frac{\Delta f}{f_0} = \frac{2v}{c}$$

to determine a value of v for the aircraft.

ii Was the aircraft getting closer to, or further away from, the airport?

■ **Figure C5.11** Air traffic control uses the Doppler effect

c Your answer to part **b** should be less than the true value of the aircraft's speed ($120\,\mathrm{m\,s^{-1}}$) because it was not flying directly to the aerial at the airport at that moment. v was the component of its velocity towards the airport.

Make a sketch showing the positions of the airport and the aircraft, and the direction to north. Add vector arrows to represent the true speed of the aircraft and its component towards the airport.

7 Suggest how it might be possible for a military aircraft to avoid being detected by radar.

Tool 3: Mathematics

Propagate uncertainties in processed data

A motorist accused of driving over the speed limit might claim that their speed was within the 'margin of error' of the equipment used by the police. Consider again the data provided in Question 5.

If the value of the microwave frequency used was believed to be accurate to ± 0.2 GHz, and the change of frequency accurate to ± 0.05 kHz, what was the uncertainty in the determination of the car's speed (in $\mathrm{km\,h^{-1}}$)?

LINKING QUESTION

- What gives rise to emission spectra and how can they be used to determine astronomical distances?

This question links to understandings in Topics B.1 and E.5.

Doppler effect with light received from distant stars and galaxies

SYLLABUS CONTENT

- ▶ Shifts in spectral lines provide information about the motion of objects like stars and galaxies in space.

Top tip!

Everything emits thermal radiation, and we have seen in Topic B.1 that this emitted radiation is in the form of a *continuous spectrum,* with a wide range of frequencies. In Topic C.2 we saw that a continuous light spectrum can be displayed on a screen using a prism to disperse the radiation.

Individual atoms and simple molecules emit electromagnetic radiation of certain precise frequencies, rather than a continuous range of frequencies. Topic E.1 explains how this is connected to changing energy levels within the atoms. When elements (in the form of gases) are given enough energy, the spectra of the light that they emit are seen as a series of bright lines on black backgrounds – called line spectra. See Figure C5.12. Each line corresponds to a precise frequency.

Figure C5.12 The principal lines of the line spectrum of hydrogen

When a continuous spectrum passes through a gas, the atoms in the gas will absorb the same frequencies as they would emit when given energy. This results in a spectrum with black absorption lines, as seen in Figure C5.13.

Atoms of the elements present in the outer layers of a star absorb light of certain frequencies from the continuous spectrum of radiation emitted from the star. Each different element produces a unique set of lines and frequencies, and this can be used to identify the element emitting the radiation.

◆ **Star** Massive sphere of plasma held together by the forces of gravity. Because of the high temperatures, thermonuclear fusion occurs and radiation is emitted.

◆ **Galaxy** A very large number of stars (and other matter) held together in a group by the forces of gravity.

◆ **Expansion of the universe** The redshift of light (similar to the Doppler effect) from distant galaxies provides evidence of an expanding universe.

◆ **Redshift, (Doppler effect)** Increase in wavelengths of electromagnetic radiation due to the fact that the distance between the observer and the source is increasing.

◆ **Blueshift** Decrease in wavelength of electromagnetic radiation (from a 'nearby' star) due to the fact that the observer and the source are moving closer together.

◆ **Recession speed** The speed with which a galaxy (or star) is moving away from Earth.

The line spectra received from **stars** within distant **galaxies** are received at lower frequencies than light from similar elements on Earth. This can be considered to be a Doppler effect and it suggests that the source and the observer are moving apart from each other.

Measurements of the Doppler effect in line spectra from stars in distant galaxies show that most galaxies are moving apart from each other – the Universe is **expanding**.

The lines on the upper spectrum seen in Figure C5.13 are produced from an element on Earth. The lower spectrum is received from the same element in a distant galaxy. Although the continuous spectra are identical, the black lines, which are characteristic of a certain element(s), have all been 'shifted' towards the red end of the spectrum, although their overall pattern is unchanged. This change to lower frequencies (larger wavelengths) is called a **redshift**. A change to higher frequency is called a **blueshift**. (Blueshifts only occur for a few stars relatively close to the Earth for reasons that need not be understood.)

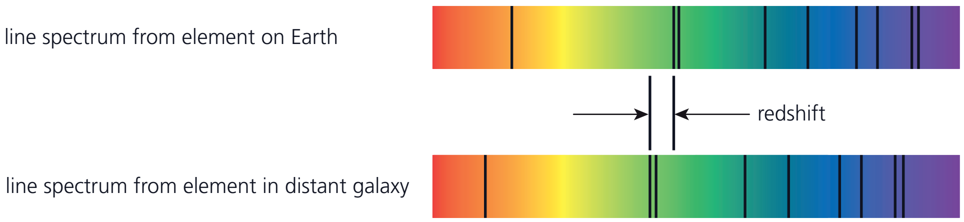


Figure C5.13 Redshift in line spectra

The equation:

$$\frac{\Delta f}{f} = \frac{\Delta \lambda}{\lambda} \approx \frac{v}{c}$$

can be used to determine the speed with which a star or galaxy is moving away from the Earth.

WORKED EXAMPLE C5.3

A line in the hydrogen spectrum has a wavelength of 4.86×10^{-7} m. When detected on Earth from a distant galaxy, the same line has a wavelength of 5.21×10^{-7} m. Determine the speed with which the galaxy is moving away from Earth. This is commonly called its **recession speed**.

Answer

$$\Delta\lambda = (5.21 \times 10^{-7}) - (4.86 \times 10^{-7}) = 3.5 \times 10^{-8}\,\text{m}$$

$$\frac{\Delta\lambda}{\lambda} \approx \frac{v}{c}$$

$$\frac{3.5 \times 10^{-8}}{4.86 \times 10^{-7}} = \frac{v}{3.00 \times 10^{8}}$$

$$v = 2.16 \times 10^{7}\,\text{m}\,\text{s}^{-1}$$

Inquiry 3: Concluding and evaluating

Figure C5.14 shows a sketch graph which summarizes how the amount of redshift detected varies with the distance of the galaxies or stars from Earth. There are significant experimental uncertainties involved, which are not shown, so that many points are not close to the line of best fit.

1 What conclusion can be drawn from this graph (for positive values of redshift)?

2 There are a few points below the horizontal axis. They are not errors. Suggest a possible explanation.

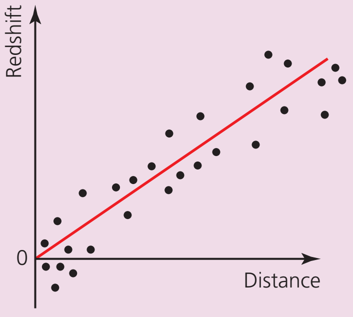


Figure C5.14 Variation of redshift

◆ **Big Bang model** Currently accepted model of the Universe, in which matter, space and time began at a point 13.7 billion years ago. The Universe has been expanding ever since.

Measurements on a large number of galaxies confirm that those with the greatest speeds are those which are furthest away. There is an obvious conclusion: they are further away *because* they are travelling faster. Moving back in time, they all started at the same place and time. This is the central concept of the **Big Bang model**. The Universe was created about fourteen billion years ago and has been expanding ever since.

Astronomers believe that all space itself is expanding, rather than stars and galaxies moving apart from each other into pre-existing space. This means that the full explanation of the Doppler effect with light from distant galaxies is not the same as the Doppler effect used to describe wave effects confined to Earth.

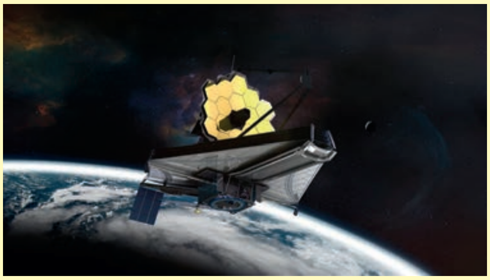
TOK

Knowledge and the knower

- How do the tools that we use shape the knowledge that we produce?

Nearly everything that we know about the Universe has been deduced only from electromagnetic radiation arriving at the Earth from space. This radiation has been detected by various types of telescopes and analysed using the techniques of **spectroscopy**. Astronomy is based solely on observation; we can choose what to observe, but we cannot design and carry out the type of laboratory experiments that characterize much of the rest of science.

Our knowledge of the Universe is limited by the instruments that we design to collect and analyse the radiation reaching the Earth.



■ **Figure C5.15** The James Webb Telescope was launched in December 2021

◆ **Spectroscopy** The analysis of spectra using instruments called spectroscopes or *spectrometers*.

8 A certain line on the helium spectrum has a well-known wavelength, but when observed from a distant galaxy it has a redshift of 1.85×10^{-8} m away from that value.
If the galaxy is receding from Earth at a speed of $7.84 \times 10^{6}\,\mathrm{m\,s^{-1}}$, determine the original wavelength of the wave.

9 **a** Calculate the size of the redshift in frequency of a wave of frequency 6.17×10^{16} Hz (from the hydrogen spectrum) received from a galaxy which has a recession speed of:

- **i** $2.20 \times 10^{6}\,\mathrm{m\,s^{-1}}$
- **ii** 10% of the speed of light.

b Determine the frequencies that will be detected on Earth.

10 Galaxies contain billions of stars all orbiting their common centre of mass (the centre of the galaxy).
If we are able to observe the stars in a galaxy ('side-on') suggest how the Doppler effect can be used to determine their rotational speeds around the centre.

LINKING QUESTIONS

- How can the Doppler effect be utilized to measure the rotational speed of extended bodies?
- How can the use of the Doppler effect for light be used to calculate speed? (NOS)

These questions link to understandings in Topics A.4.

TOK

Knowledge and technology

- To what extent are technologies, such as the microscope and telescope, merely extensions to the human senses, or do they introduce radically new ways of seeing the world?

Doppler first identified the effect that bears his name about 180 years ago, but he could not have foreseen the many useful applications that his discovery would lead to. This is mainly because the relatively small changes of frequency involved require a high level of technology.

Equations for use with the Doppler effect for sound (or other mechanical waves)

SYLLABUS CONTENT

- The observed frequency for sound and mechanical waves due to the Doppler effect as given by:

  moving source, $f' = f\left(\frac{v}{v \pm u_s}\right)$, where u_s is the velocity of the source

  moving observer, $f' = f\left(\frac{v \pm u_o}{v}\right)$ where u_o is the velocity of the observer.

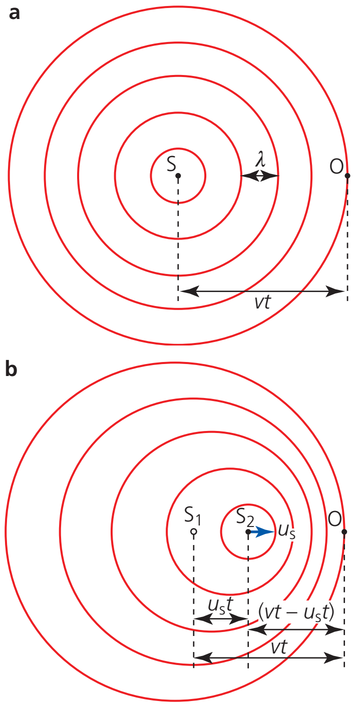


Figure C5.16 a Waves between a stationary source and a stationary observer **b** Waves between a moving source and stationary observer

Figure C5.16a shows waves of frequency f and wavelength λ travelling at a speed v between a stationary source S and a stationary observer O. In the time, t, that it takes the first wavefront emitted from the source to reach the observer, the wave has travelled a distance vt. The number of waves between the source and observer is $\frac{t}{T} = ft$.

The wavelength, λ, equals the total distance divided by the number of waves:

$\lambda = \frac{vt}{ft} = \frac{v}{f}$ as we would expect from Topic C.2.

Figure C5.16b represents exactly the same waves emitted in the same time from a source moving towards a stationary observer with a speed u_s. In time, t, the source has moved from S_1 to S_2. The number of waves is the same as Figure C5.16a, but because the source has moved forwards a distance, u_st, the waves between the source and the observer are now compressed within the length $vt - u_st$.

This means that the observed (received) wavelength, λ', equals the total distance divided by the number of waves:

$$\lambda' = \frac{vt - u_s t}{ft} = \frac{v - u_s}{f}$$

The observed (received) frequency, f', is given by:

$$f' = \frac{v}{\lambda'} = \frac{vf}{v - u_s}$$

If the source is moving away from the observer, the equation becomes:

$$f' = \frac{vf}{v + u_s}$$

In general, we can write:

$$\boldsymbol{f' = f\left(\frac{v}{(v \pm u_s)}\right)}$$

This is the equation for the Doppler effect from a *moving source* (speed u_s) detected by a stationary observer.

u_s is added when the source is moving away from the observer, and subtracted when the motion is towards the observer.

The equation for the frequency detected by a *moving observer* from a stationary source is:

$$\boldsymbol{f' = f\left(\frac{v \pm u_o}{v}\right)}$$

u_o is added when the observer is moving towards the source, and subtracted when the motion is away from the source.

These equations assume that the motion involved is in the same direction as a straight line joining the source and the observer.

Remember that these equations cannot be applied to electromagnetic waves ($c >> v$).

WORKED EXAMPLE C5.4

a A source of sound emitting a frequency of 480 Hz is moving directly towards a stationary observer at 50.0 m s^{-1}.

If it is a hot day and the speed of sound is 350 m s^{-1}, calculate the frequency received.

b What frequency would be heard on a cold day when the speed of sound was 330 m s^{-1}?

c Explain why the speed of sound is less on a colder day.

Answer

a $f' = f\left(\frac{v}{v - u_s}\right) = \frac{(480 \times 350)}{(350 - 50)} = 560\,\text{Hz}$

b $f' = \frac{(480 \times 350)}{(330 - 50)} = 566\,\text{Hz}$

c Sound is transferred though air by moving air molecules. On a colder day the molecules will have a lower average speed.

11 Calculate the frequency which will be received by an observer moving with a speed of 24.2 m s^{-1} directly away from a stationary source of sound waves of frequency 980 Hz. (Assume the speed of sound to be 342 m s^{-1}.)

12 In a Doppler ultrasound measurement, as shown in Figure C5.5, blood was flowing at a rate of 9.7 cm s^{-1} along the artery.

- **a** If the angle $\theta = 75°$, calculate the speed that the transducer should detect.
- **b** If the transducer emits waves of frequency 5.87400 MHz, and the speed of the ultrasound waves is 1540 m s^{-1}, determine the frequency received by the blood cells.

13 A car is travelling along a straight road at a constant speed of 31 m s^{-1}. The car emits a sound with a constant frequency of 224 Hz. A pedestrian on a footbridge over the motorway watches the car approach, travel directly under them, and then move away from the bridge. (Assume the speed of sound to be 342 m s^{-1}.)

- **a** Determine the frequencies heard by the pedestrian:
  - **i** before the car passes the footbridge
  - **ii** after the car passes the footbridge.
- **b** Another person watches and hears the same car but is many metres away from the side of the road. Describe and explain how the sound reaching this person is different from the sound heard by the pedestrian on the footbridge.

14 An ultrasound wave of frequency 30.0 kHz is directed at an approaching car. The wave reflects off the car and is received back at the stationary emitter with a frequency of 32.7 kHz.

- **a** Using an equation highlighted on page 409, calculate the velocity of the car. (Assume that the speed of sound is 335 m s^{-1}.)
- **b** Compare your answer to the answer obtained by using:

$$\frac{\Delta f}{f} \approx \frac{\text{relative speed of source}}{\text{speed of waves}}$$

15 A boat is travelling directly towards a jetty at a speed of 37 cm s^{-1} and creating waves on the water surface of original wavelength 59 cm.

If the water waves travel at a speed of 94 cm s^{-1}, determine the frequency and wavelength of the waves reaching the jetty.

D.1 Gravitational fields

Guiding questions

- How are the properties of a gravitational field quantified?
- How does an understanding of gravitational fields allow for humans to explore the Solar System?

■ **Figure D1.1** Johannes Kepler

The fact that (most) objects tend to fall towards the ground is, of course, a common and unsurprising observation. But a satisfactory explanation was not achieved until the work of Isaac Newton in the seventeenth century. Until then, Aristotle's ideas, from more than two thousand years earlier, were the accepted wisdom: falling objects were just returning to their 'natural' places.

A scientific understanding of gravity requires accurate observations of objects moving large distances where gravitational effects are variable and unaffected by air resistance: the planets and the Moon.

The Danish nobleman Tycho Brahe was renowned for his remarkably accurate astronomical measurements at a time before the invention of the telescope (late sixteenth century). But they were just that: empirical observations, without explanation.

Johannes Kepler (Figure D1.1) worked with Brahe's data and analysed his measurements mathematically, particularly those concerning the planet Mars. Kepler's three laws of planetary motion were a key development in the history of astronomy.

Kepler's laws of planetary motion

SYLLABUS CONTENT

- ▶ Kepler's three laws of orbital motion.

Kepler's laws were developed to describe the motions of the planets in the Solar System, but they can be applied to any group of bodies orbiting a common centre under the effects of gravity. (To *orbit* means to continuously move, revolving around another, larger, object.)

Kepler's laws were empirical (based on observations) and it was not until 70 years later that Newton was able to provide the underlying explanation.

Tool 3: Mathematics

Construct and use scale diagrams

An **ellipse** is the name that we give to the type of complete curve for which for all points: the sum of the distances from two fixed points is always the same. See Figure D1.2.

The two fixed points, shown as F_1 and F_2 in Figure D1.2, are each called a focus of the ellipse (plural: foci). If F_1 and F_2 are at the same point, in the centre, the ellipse becomes a circle.

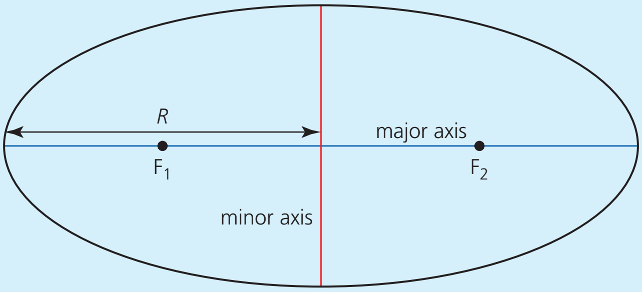


Figure D1.2 An ellipse

◆ **Ellipse** Closed curve consisting of points whose distances from each of two fixed points (**focuses, foci**) always add up to the same value.

Kepler's first law

Kepler showed that the planets of the Solar System move in *elliptical* paths. See Figure D1.3.

The planets orbit in elliptical paths, with the Sun at one of the two foci.

It should be noted that the orbits of most of the planets of our Solar System are nearly circular. Figures D1.2 and D1.3 have exaggerated the shape (*eccentricity*) of the ellipses.

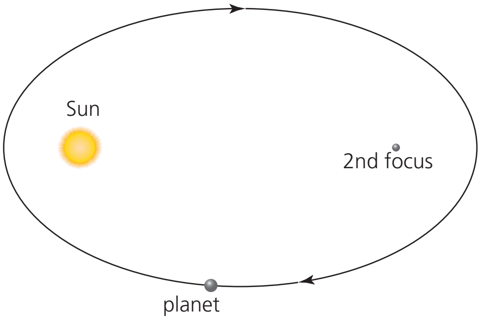


Figure D1.3 Elliptical path of a planet.

Kepler's second law

Kepler's second law expresses the fact that planets move faster when they are closer to the Sun.

A line joining a planet and the Sun sweeps out equal areas in equal times.

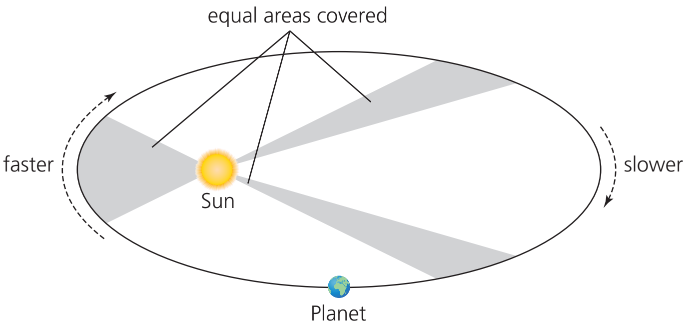


Figure D1.4 Equal areas in equal times

Kepler's third law

In effect, the third law provides a mathematical relationship between half the length of the major axis (R in Figure D1.1) and the planet's speed. In practice, because many orbits are close to being circular, we can usually assume that R is the average distance (orbital radius) to the Sun:

The square of a planet's orbital time period, T, is proportional to the cube of its average orbital radius, R:

$$T^2 \propto R^3$$

or

$$\frac{R^3}{T^2} = \text{constant}$$

The Earth has an average distance from the Sun of **1.50×10^{11} m**. (This is often called one *astronomical unit*, AU.)

WORKED EXAMPLE D1.1

If the average distance between the Sun and the planet Venus is 1.08×10^{11} m, calculate the time period of Venus's orbit.

Answer

$\frac{T^2}{R^3}$ = constant, the same for Venus as the Earth

If T_V is the period of Venus's orbit:

$$\frac{365.25^2}{(1.496 \times 10^{11})^3} = \frac{T_V{}^2}{(1.08 \times 10^{11})^3}$$

$T_V = 224$ (Earth) days

LINKING QUESTION

- How is uniform circular motion like – and unlike – real-life orbits?

This question links to understandings in Topic A.4.

1 The Earth's orbit around the Sun is not perfectly circular.
  a Use the internet to determine their maximum and minimum separation, and when these events occur.
  b Discuss what effect this has on the climate at any particular location (if any).

2 Figure D1.5 shows the orbits of a planet and a comet around the Sun.

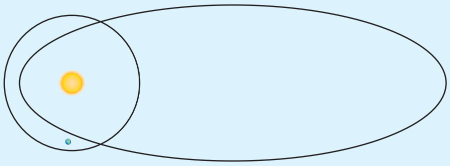


Figure D1.5 The orbits of a planet and a comet around the Sun

  a State how you know which is which.
  b At what positions will the comet be travelling:
    i fastest
    ii slowest?

3 Determine the value of T^2/R^3 in SI units for the planets of the Solar system.

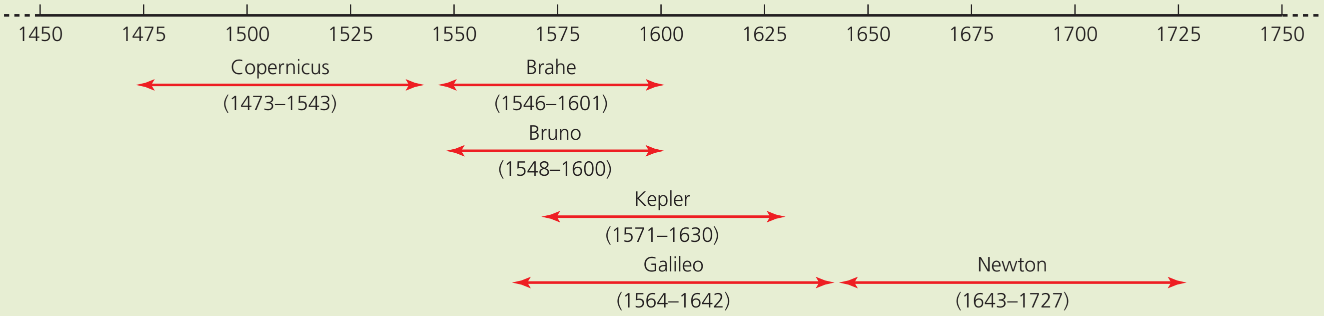
4 Calculate the period of the planet Mars, which has an average distance of 228 million kilometres from the Sun.

5 One of the planets of the Solar System has a period of approximately 84 years.
  a Determine its distance from the Sun.
  b Express your answer to part **a** in astronomical units, AU.
  c Find out which planet it is.

6 a Research how long it takes for the Moon to orbit the Earth.
  b The centre of the Moon is an average distance of 384 000 km from the centre of the Earth. Calculate a value of T^2/R^3 from this data.
  c Use Kepler's third law to determine the orbital radius of an Earth satellite which takes exactly one day to compete its orbit. (A satellite with an orbit of this radius can appear to remain 'stationary' above the equator.)

Nature of science: Science as a shared endeavour

'The shoulders of giants'



■ **Figure D1.6** Time lines of some famous early astronomers

Isaac Newton's law of gravitation is central to this topic and the name Newton appears prominently in physics text books. Famously, he is quoted as saying *'If I have seen further, it is by standing on the shoulders of giants'*. Figure D1.6 shows a time line of Newton's most famous predecessors in the study of astronomy in the sixteenth and seventeenth centuries.

■ **Figure D1.7** Isaac Newton

■ **Figure D1.8** Copernicus

Nicolas Copernicus, a Polish astronomer (Figure D1.8), is considered by many to be the founder of modern astronomy. In 1530 he published a famous paper stating that the Sun was the centre of the universe and that the Earth, stars and planets orbited around it (a *heliocentric* model). At that time, and for many years afterwards, these views directly challenged 'scientific', philosophical and religious beliefs. It was then generally believed that the Earth was at the centre of everything (a *geocentric* model). That profound and widespread belief dated all the way back to Ptolemy, Aristotle and others nearly 2000 years earlier. It should be noted, however, that Aristarchus in Ancient Greece is generally credited with being the first well-known person to propose a heliocentric model.

In Italy, the astronomer Giordano Bruno took the heliocentric model further with revolutionary suggestions that the universe was infinite and that the Sun was not at the centre. The Sun was, Bruno suggested (correctly), similar in nature to the other stars. He was burned at the stake in 1600 for these beliefs – at the time, considered by some to be heresy. About 30 years later, one of the greatest scientific thinkers of all time, Galileo Galilei, was placed on trial by the Roman Catholic Church under similar charges. Many years earlier he had used the newly invented telescope to observe the moons of Jupiter and had reasoned that the Earth orbited the Sun in a similar way, as had been proposed by Copernicus. Under pressure, he publicly renounced these beliefs and was allowed to live the rest of his life under house arrest. All this has provided the subject of many books, plays and movies.

About 700 years before the time of Newton, during the Islamic Golden Age, Abd al-Rahman al-sufi and other Muslim astronomers identified stars and constellations with impressive accuracy (building on the work of Ptolemy, centuries earlier). Abd al-Rahman al-sufi's 'Book of fixed stars' has an important place in the history of astronomy.

Newton’s universal law of gravitation

SYLLABUS CONTENT

- Conditions under which extended objects can be treated as point masses.
- Newton’s universal law of gravitation as given by: $F = G\frac{m_1m_2}{r^2}$ for bodies treated as point masses.

Isaac Newton was the first to realize that if the force of gravity makes objects (like apples) fall to the Earth and also keeps the Moon in orbit around the Earth, then it is reasonable to assume that the force of gravity acts between *all* masses. This is why it is called *universal* gravitation. Newton believed (correctly) that the size of the gravitational force between two masses increased with the sizes of the masses, and decreased with increasing distance between them – following an *inverse square relationship*.

Universal gravitation and the inverse square law

Newton knew that the distance between the Earth and the Moon was equal to 60 Earth radii, and he was able to prove that the centripetal acceleration of the Moon towards the Earth was equal to $g/60^2$ (using $a = v^2/r$ from Topic A.2). See Worked example D1.2 and Figure D1.9.

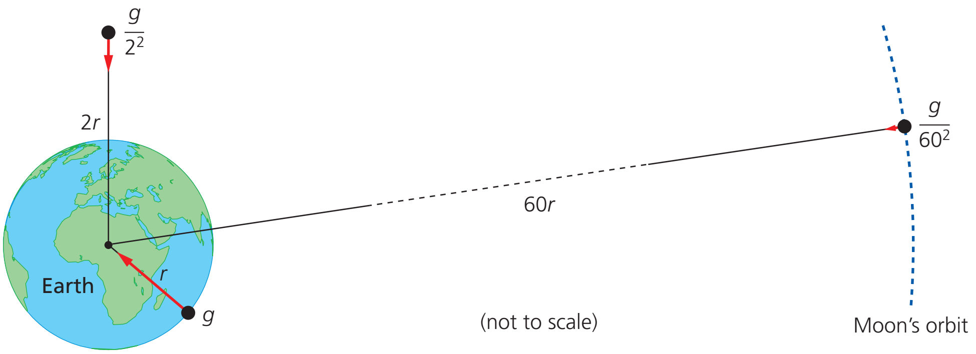


Figure D1.9 How the acceleration due to gravity varies with distance from the Earth

WORKED EXAMPLE D1.2

The average distance between the Earth and the Moon is 384 000 km and the Moon takes 27.3 days to orbit the Earth.

a Calculate the average orbital speed of the Moon. Assume that its orbit is circular.

b Determine the centripetal acceleration of the Moon towards the Earth.

c Compare your answer for **b** to $g/60^2$, with $g = 9.81\,\text{m s}^{-2}$. Comment on the difference.

Answer

a $v = \frac{2\pi r}{T} = \frac{(2 \times \pi \times 3.84 \times 10^8)}{(27.3 \times 24 \times 3600)} = 1.02 \times 10^3\,\text{m s}^{-2}$ (1022.90... seen on calculator display)

b $a = \frac{v^2}{r} = \frac{(1.0229 \times 10^3)^2}{3.84 \times 10^8} = 2.725 \times 10^{-3}\,\text{m s}^{-2}$

c $\frac{9.81}{60^2} = 2.725 \times 10^{-3}\,\text{m s}^{-2}$

The two answers are the same. This is very good evidence that gravitational accelerations (and forces) are represented by inverse square laws.

Newton's law

◆ **Newton's universal law of gravitation** There is a gravitational force between two point masses, m_1 and m_2, given by $F = G\frac{m_1 m_2}{r^2}$, where r is the distance between them and G is the universal gravitation constant.

◆ **Gravitational forces** Fundamental attractive forces that act across space between all masses. Gravitational force reduces with an inverse square law with increasing distance between point masses.

◆ **Universal gravitation constant, G** The constant that occurs in Newton's universal law of gravitation. $G = 6.67 \times 10^{-11}\,\text{N m}^2\,\text{kg}^{-2}$.

To simplify the situation, we will consider the forces acting on only two masses. The masses may be of any magnitude but, to begin with, we will assume that they are *point masses.* That is, all their mass is considered to be at a single point.

The forces acting between two point masses (m_1 and m_2) are proportional to the product of the masses and inversely proportional to their separation (r) squared.

$$F \propto (m_1 \times m_2) \quad \text{and} \quad F \propto \frac{1}{r^2}$$

Putting a constant of proportionality into the relationship, we get Newton's universal law of gravitation:

gravitational force between two (point) masses, $F = G\frac{m_1 m_2}{r^2}$

G is known as the **universal gravitation constant**. It has a value of $\mathbf{6.67 \times 10^{-11}\,N\,m^2\,kg^{-2}}$.

The small value of G reflects the fact that gravitational forces are small unless one (or both) of the masses is very large. G is a fundamental constant which, as far as we know, always has exactly the same value everywhere in the universe and for all time. It should not be confused with g, the acceleration due to gravity, which varies with location. The relationship between g and G is covered later in this topic.

The relationship between force and distance is illustrated in Figure D1.10. Note that exactly the same force always acts on *both* masses (but in opposite directions), even if one mass is larger than the other. This is an example of Newton's third law of motion.

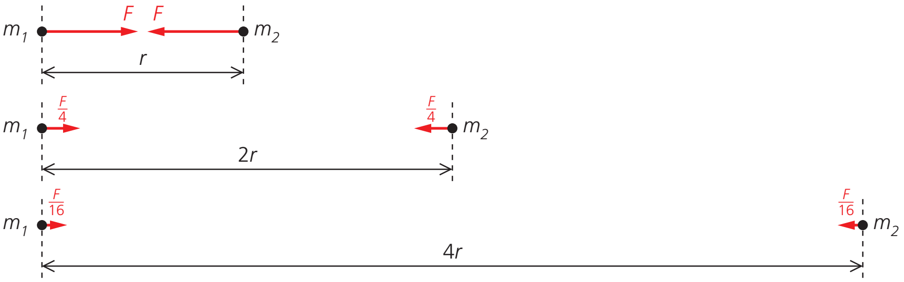


Figure D1.10 The gravitational force between point masses m_1 and m_2 decreases with increasing separation (the vectors are not drawn to scale)

Of course, the mass of an object is not all located at one point, but this does not mean that Newton's equation cannot be used for real masses. The forces between two spherical masses of uniform density located a long way apart are the same as if the spheres had all of their masses concentrated at their centre points. The gravitational effects around a planet (assumed to be spherical) are effectively the same as would be produced by a similar mass concentrated at the centre of the planet.

Newton was also able to confirm his law of gravitation by showing that it was consistent with Kepler's third law, as follows:

For circular gravitational orbits the necessary centripetal force (Topic A.2) is provided by gravity. For a relatively small mass, m, orbiting a much larger mass, M:

$$\frac{\cancel{m} v^2}{r} = \frac{GM\cancel{m}}{r^2} \Rightarrow v^2 = \frac{GM}{r}$$

Then, since $v = \frac{2\pi r}{T}$:

$$\left(\frac{2\pi r}{T}\right)^2 = \frac{GM}{r} \Rightarrow$$

$$\frac{r^3}{T^2} = \frac{GM}{4\pi^2} \text{ (a constant)}$$

This is Kepler's third law (as introduced earlier in this topic), showing a value for the constant on the right-hand side of the equation.

WORKED EXAMPLE D1.3

Calculate the gravitational forces acting between the Earth and a 1.0 kg book on the Earth's surface. (The Earth's mass is 6.0×10^{24} kg and its radius is 6.4×10^{6} m.)

Answer

$$F = G\frac{m_1 m_2}{r^2} = \frac{(6.67 \times 10^{-11}) \times 1.0 \times (6.0 \times 10^{24})}{(6.4 \times 10^{6})^2} = 9.8\,\text{N}$$

This is the *weight* of a 1.0 kg mass on the Earth's surface. The book attracts the Earth up towards it with an equally sized force which has a negligible effect on the Earth. This is another example of Newton's third law.

LINKING QUESTION

- Physics utilizes a number of constants such as G. What is the purpose of these constants and how are they determined? (NOS)

Nature of science: Measurement

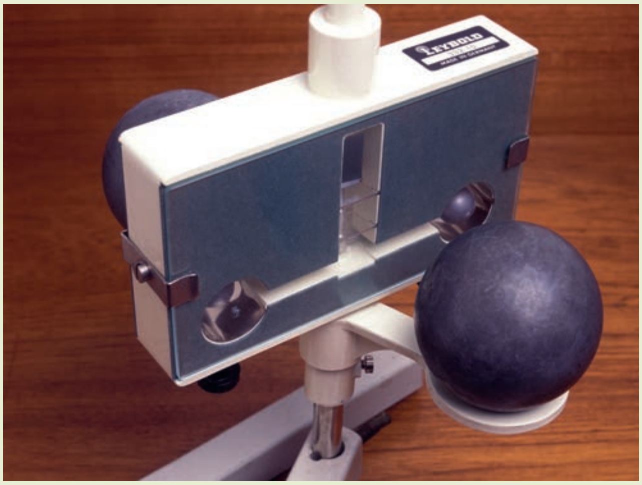
Weighing the Earth

At the time Newton proposed his law of universal gravitation it was not possible to determine an accurate value for the gravitational constant, G. The only gravitational forces that could be measured were those of the weights of given masses on the Earth's surface. The radius of the Earth was known, but that still left two unknowns in the equation $F = Gm_1m_2/r^2$: the gravitational constant and the mass of the Earth. If either of these could be found, then the other could be calculated using Newton's law of gravitation. That is why the determination of an accurate value for G was known as '*weighing the Earth*'.

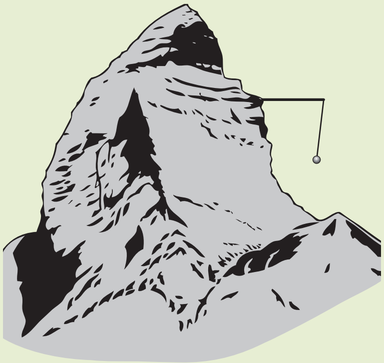
Certainly, it was possible in the seventeenth century to get an approximate value for the mass of the Earth from its volume and estimated average density (using $m = \rho V$). But density estimates would have been little more than educated guesses. We know now that the Earth's crust has a much lower average density (about 3000 kg m^{-3}) than most of the rest of the Earth. However, it was possible to use an estimate of the Earth's mass to calculate an approximate value for the gravitation constant. The first accurate measurement was made more than 100 years later by Henry Cavendish in an experiment that is famous for its precision and accuracy.

To calculate a value for G without needing to know the mass of the Earth (or the Moon, or another planet) required the direct measurement of the force between two known masses. Cavendish used lead spheres (see Figure D1.11) because of their high density (11.3 g cm^{-3}). The forces involved are very difficult to measure because they are so small, but also because similar-sized forces can arise from various environmental factors. (In fact, Cavendish's main aim was to get a value for the density of the Earth rather than to measure G.)

In an early attempt to estimate the gravitational constant and calculate a value for the mass of the Earth, pendulums were suspended near mountains (see an exaggerated representation in Figure D1.12).



■ **Figure D1.11** A modern version of Cavendish's apparatus



■ **Figure D1.12** A pendulum and a mountain attract each other

7 The gravitational force acting on a satellite orbiting 50 km above the Moon's surface was 840 N. Calculate a value for the force if the height above the surface was ten times greater. Radius of Moon = 1.74×10^6 m

8 Estimate the gravitational force between you and your pen when you are 1 m apart.

9 **a** Determine the gravitational force between two steel spheres each of radius 45 cm and separated by 10 cm. Density of steel = $7900\,\mathrm{kg\,m^{-3}}$

b Show that if solid steel spheres with twice the radius were used, with the same separation between their surfaces, the force would increase by a factor of about 20.

10 Calculate the average gravitational force between the Earth and the Sun. (You will need to research the relevant data.)

11 A proton has a mass of 1.7×10^{-27} kg and the mass of an electron is 9.1×10^{-31} kg.

a Estimate the gravitational force between these two particles in a hydrogen atom, assuming that they are 5.3×10^{-11} m apart.

b Compare your answer to the magnitude of the electric force between the same two particles (8.2×10^{-8} N, which is explained in Topic D.2).

c Comment on your answer.

12 Ganymede and Callisto are the two largest moons of Jupiter. Ganymede has an orbital radius of 1.07×10^6 km and orbits every 7.15 Earth days. Callisto has an orbital radius of 1.88×10^6 km and orbits every 16.7 Earth days.

a Determine if this data is consistent with Kepler's third law.

b Calculate the mass of Jupiter.

Gravitational fields

SYLLABUS CONTENT

- Gravitational field strength, g, at a point is the force per unit mass experienced by a small point mass at that point as given by: $g = \frac{F}{m} = \frac{GM}{r^2}$.
- Gravitational field lines.

A region (around a mass) in which another mass would experience a gravitational force is called a gravitational field.

Theoretically, all masses produce gravitational fields around themselves, but in practice we only use the term when discussing the space around very large masses like moons, planets and stars. We all live in the gravitational field of the Earth, while the Earth moves in the gravitational field of the Sun.

TOK

The natural sciences

- What kinds of explanations do natural scientists offer?

Fields

Understanding gravitational, electric and magnetic forces is fundamental knowledge about the universe in which we live. It seems that these forces can act instantaneously across space, even if there is nothing but vacuum (free space) in between. And, in the case of gravity, the forces can act across unbelievably large distances. All this is very difficult to comprehend!

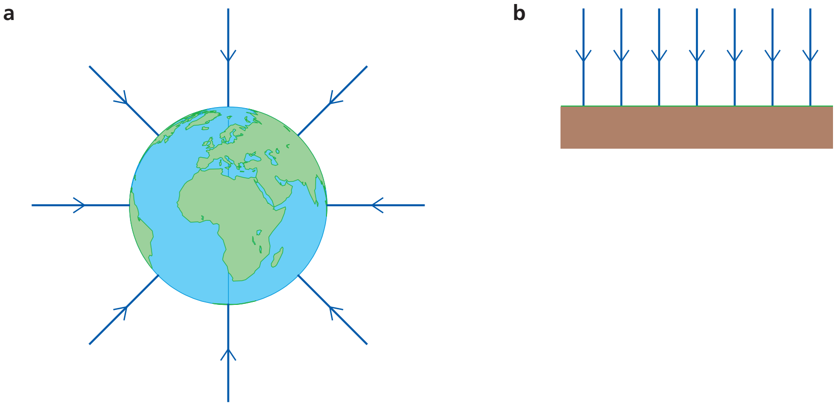
Using the concept of **fields** (gravitational, electric and magnetic) to describe the intermediate spaces may seem to help understanding, but it does not explain the origin of the forces.

Scientists cannot say 'that is just the way it is' and leave it at that. They need a deeper, more fundamental understanding, but that is beyond the requirements of this course.

◆ **Field (gravitational, electric or magnetic)** A region of space in which a mass (or a charge, or a current) experiences a force due to the presence of one or more other masses (charges, or currents – moving charges).

◆ **Field lines and patterns** Representation of fields in drawings by a pattern of lines. Each line shows the direction of force on a mass (in a gravitational field), of force on a positive charge (in an electric field), or on a north pole (in a magnetic field). In any particular drawing, the field is strongest where the lines are closest together.

We often want to represent a gravitational field on paper, or on a screen, and this can be done with gravitational **field lines** as shown in Figure D1.13. The arrows show the direction of the gravitational force that would be experienced by a mass placed at any particular place in the field. Figure D1.13a represents the spreading radial gravitational field lines around the Earth. The lines are closer together nearer to the Earth, which shows that the gravitational field is stronger. Field lines never cross each other; that would mean that gravitational force was acting in two different directions at the same place.



■ **Figure D1.13** Field lines are used to represent gravitational fields on paper or on screen. **a** radial field **b** uniform field

The parallel lines in Figure D1.13b represent a uniform gravitational field, such as in a small region of the Earth's surface where variations in the field are negligible. For example, the room where you are sitting.

Gravitational field strength

We may want to ask the question 'if a mass was put in a particular place, what would be the gravitational force on it?' The answer, of course, depends on the magnitude of the mass, so it is more helpful to generalize and ask 'what would the force be on a unit mass (1 kg)?' If we know this, then we can easily calculate the gravitational force on any other mass.

Gravitational field strength, g, is defined as the force per unit mass that would be experienced by a small test mass placed at that point:

$$g = \frac{F}{m}$$

◆ **Test mass** An object of insignificant mass used in the definition and measurement of gravitational fields.

Reference is made to a 'small **test mass**' because a large mass (compared to the mass, or masses, creating the original field) could have a significant gravitational field of its own. Gravitational field strength is given the symbol g and has the SI unit N kg^{-1}. Gravitational field strength is a vector quantity and its direction is shown by the arrows on field lines.

As explained in Topic A.2, in general, we know from Newton's second law of motion, that $a = F/m$, so that gravitational field strength ($g = F/m$) in N kg^{-1} is numerically equal to the acceleration due to gravity in m s^{-2}.

Imagine you were on an unknown planet and wanted to find experimentally the gravitational field strength. This can be done easily by hanging a small test mass of 1.0 kg on a force-meter, calibrated in newtons. The reading will be the strength of the gravitational field (in N kg^{-1}) and the direction of the field will be the same as the direction of the string – 'downwards' towards the centre of the planet.

WORKED EXAMPLE D1.4

A student measures the time it takes a stone to fall from rest to the ground from a height of 1.18 m to be 0.49 s. Determine the value this gives for the acceleration due to gravity at her location.

Answer

$$s = ut + \frac{1}{2}at^2$$

$$1.18 = 0 + \frac{1}{2} \times a \times 0.49^2$$

$$a = 9.8\,\mathrm{m\,s^{-2}}$$

This is numerically equal to the gravitational field strength, $g = 9.8\,\mathrm{N\,kg^{-1}}$

Common mistake

r in the highlighted equation represents distances from the centre of the planet, or moon. It is not the radius (unless we are calculating g on the surface).

Gravitational field strength around a planet

The gravitational field strength around a large mass (a planet for example) can be determined by combining $g = \frac{F}{m}$ with the equation for gravitational force, $F = G\frac{m_1 m_2}{r^2}$ ($m = m_1$ represents the small mass):

$$g = \frac{F}{m} = \frac{G\cancel{m_1}\,m_2}{\cancel{m_1}\,r^2}$$

Representing the large mass by M, rather than m_2:

$$\boldsymbol{g = G\frac{M}{r^2}}$$

WORKED EXAMPLE D1.5

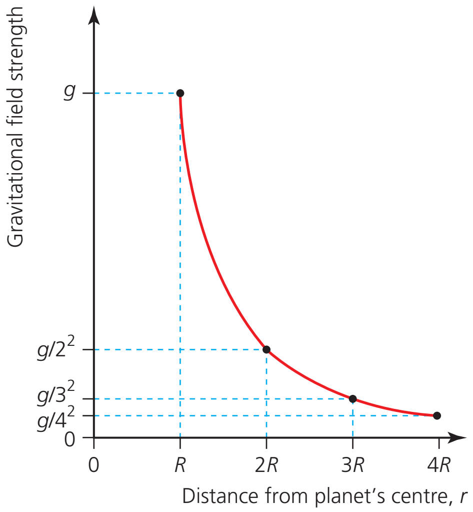
Determine the gravitational field strength on the surface of a planet which has a mass of 4.87×10^{24} kg and a radius of 6.05×10^6 m.

Answer

$$g = G\frac{M}{r^2} = \frac{(6.67 \times 10^{-11}) \times (4.87 \times 10^{24})}{(6.05 \times 10^6)^2} = 8.87\,\mathrm{N\,kg^{-1}}$$

(This planet is Venus.)

Similar to gravitational force, the gravitational field strength around a planet, or a moon, follows an inverse square law. This is represented graphically in Figure D1.14.



The gravitational field strength beneath the surface of a planet cannot be determined from $g = GM/r^2$ because the planet can no longer be considered to be a point mass. At any significant depth the mass above and to the side would also be pulling a 'test mass'. At the centre of a planet the field strength will be assumed to be zero, because the surrounding masses will pull equally in all directions. Moving from the centre to the surface, the gravitational field strength will increase.

🟥 **Figure D1.14** Variation of gravitational field strength with distance from a planet (or moon) of radius R

Tool 3: Mathematics

Determine the effect of changes to variables on other variables in a relationship

For a planet of radius R, the gravitational field strength on its surface can be determined from:

$$g = G\frac{M}{R^2}$$

It is easy to assume, incorrectly, that g decreases for planets of greater radius. In fact, the opposite is true because the mass, M, of a planet also depends on its radius, as shown below.

To determine how the gravitational field strength on the surface of a planet depends on its radius, R, we need to use these facts:

- the volume of a sphere equals $\frac{4}{3}\pi R^3$
- mass, M, is equal to density, ρ, multiplied by volume, V

So, we can write:

$$M = \frac{4}{3}\rho\pi R^3$$

The density of a planet is not uniform, so the value of ρ used here is an average.

Putting this equation for M back into the equation for g we get:

$$g = \frac{4}{3}G\rho\pi\left(\frac{R^3}{R^2}\right)$$

So that:

$$g = \frac{4}{3}G\rho\pi R$$

This equation (which students are not expected to remember) predicts that the gravitational field strength at the surface of a planet is proportional to its radius. From the equation we would expect bigger planets to have stronger fields, but that is only true if they have equal average densities. (The Earth is the densest planet in our Solar system, with an average density of 5510 kg m^{-3}. Venus and Mercury have similar densities to Earth but the density of Mars is significantly lower. The outer planets are gaseous and have lower densities. Saturn has the lowest average density, at 687 kg m^{-3}.)

WORKED EXAMPLE D1.6

Predict a value for the gravitational field strength on the 'surface' of Saturn (radius = 5.8×10^7 m).

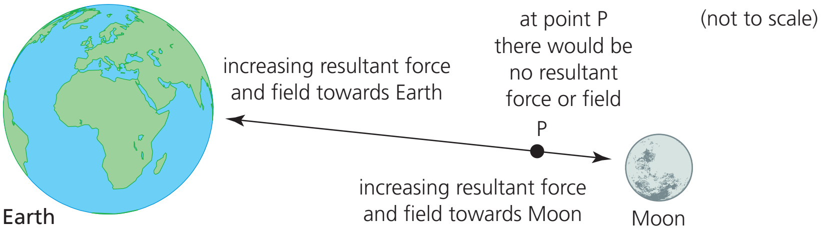
Answer

$$g = \frac{4}{3}G\rho\pi r = \frac{4}{3} \times (6.67 \times 10^{-11}) \times 687 \times \pi \times (5.8 \times 10^7) = 11\,\text{N kg}^{-1}$$

(Accepted value is 10.4 N kg^{-1})

Combining gravitational field strengths

It is possible that a mass may be in two or more separate and significant gravitational fields. For example, we are in the fields of both the Earth and the Moon. For most purposes the Moon's gravitational field on the Earth's surface can be considered to be negligible compared to the Earth's field. But, if a spacecraft is travelling directly from the Earth to the Moon, the gravitational field due to the Earth will get weaker as the Moon's field gets stronger. There will be a point at which the two fields will be equal in strength, but opposite in direction (shown as P in Figure D1.15).



■ **Figure D1.15** Opposing fields cancel at a precise point P between the Earth and the Moon

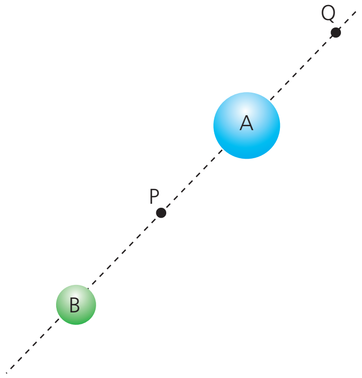
At P, the total gravitational field strength is zero and there will be no resultant force on the spacecraft because the pulls of the Moon and the Earth are equal and opposite. As the spacecraft travels from the Earth to P there is a resultant force pulling it back to Earth but this is reducing in size. After the spacecraft passes P there will be an increasing resultant force pulling the spacecraft towards the Moon.

In general, if two or more masses are creating gravitational fields at a certain point, then the total field is determined by adding the individual fields, remembering that they are vector quantities.

In this chapter as we will only be concerned with locations somewhere on the line passing through the masses, then the vector addition of the two fields is straightforward, as shown in the following Worked example.

WORKED EXAMPLE D1.7

In Figure D1.16 (which is not drawn to scale), P is a point midway between the centres of the planets A and B. At P the gravitational field strength due to A is 4.0 N kg^{-1} and that due to B is 3.0 N kg^{-1}.



■ **Figure D1.16** Point P between the centres of the planets A and B

a Determine the resultant gravitational field strength at P.

b Calculate the combined gravitational field strength at point Q. P and Q are the same distance from A.

Answer

a Taking the field towards the bottom of the diagram to be positive,

$(-4.0) + (+3.0) = -1.0$

The gravitational field strength is 1.0 N kg^{-1} towards A.

b The size of the field due to A is the same at Q as it is at P, although it is in the opposite direction. The strength of the field due to B at Q is 3^2 times less than at P because it is three times further away, but it is in the same direction.

$(+4.0) + (3.0/9) = 4.3$ N kg^{-1} towards A and B.

13 Make a sketch of the gravitational field in the room where you are sitting.

14 The weight of a 12 kg mass on the surface of Mercury would be 44 N.
Calculate the gravitational field strength on the surface of the planet.

15 a Determine the gravitational field strength at a height of 300 km above the Earth's surface. (The radius of the Earth is 6.37×10^6 m. The mass of the Earth is 5.97×10^{24} kg.) Many satellites orbit at about this height.

b Calculate the percentage this value is of the accepted value for the gravitational field strength on the Earth's surface.

16 The gravitational field strength of a planet is $5.8\,\text{N kg}^{-1}$ at a distance of 2.1×10^4 km from its centre. Determine the field strength at a distance 1.4×10^4 km further away.

17 Draw a sketch graph to show how the gravitational field strength varies from the centre of the Earth to a distance of 12.8×10^6 m. (radius of Earth = 6.4×10^6 m)

18 a Calculate the gravitational field strength on the surface of the Moon. The mass of the Moon is 7.35×10^{22} kg and its Radius is 1740 km.

b Calculate the gravitational field strength at a point on the Earth's surface due to the Moon (not the Earth). The distance between the centre of the Moon and the Earth's surface is 3.8×10^8 m.

c State one effect that the Moon's gravitational field has on Earth.

19 Titan is a moon of the planet Saturn. It has an average density of $1900\,\text{kg m}^{-3}$. The gravitational field strength on its surface is approximately 14% of that on Earth. Estimate Titan's radius using the equation given above.

20 Consider Figure D1.16 again, but with different data. If planet A has a gravitational field of $15\,\text{N kg}^{-1}$ at Q, but the combined field at the same point is $16\,\text{N kg}^{-1}$, calculate the combined field at point P.

21 The gravitational fields of the Sun and the Moon cause the tides on the world's oceans. The highest tides occur when the resultant field is greatest (at times of a 'new moon'). Draw a sketch to show the relative positions of the Earth, Sun and Moon when the resultant field on the Earth's surface is:

a greatest

b weakest.

Tool 2: Technology

Use spreadsheets to manipulate data

1. Research the data that will allow you to set up a spreadsheet to calculate the combined gravitational field strengths due to the Earth and the Moon at points along a straight line joining their surfaces.
2. Combine the fields to determine the resultant field and draw a graph of the results.
3. Where does the resultant gravitational field equal zero?

Orbital motion

◆ **Satellite** Object that orbits a much larger mass. Satellites can be **natural** (like the Earth or the Moon), or **artificial** (as used for communication, for example).

◆ **Altitude** Height of an object above the surface of a planet.

The gravitational forces between two masses are equal in size but opposite in direction. However, if one of the masses is very much bigger than the other, we often assume that the force on the larger mass has negligible effect, while the same force acting on the much smaller mass produces a significant acceleration. If the smaller mass is already moving in a suitable direction, then the gravitational force can provide the centripetal force to make it orbit the larger mass. It is then described as a **satellite** of the larger mass. The Earth and the other planets orbiting the Sun, and moons orbiting planets, are all examples of **natural satellites**. In the modern world we are becoming more and more dependent on the **artificial satellites** that orbit around the Earth.

Newton's famous thought experiment was described in Topic A.1: a cannonball fired 'horizontally' from the top of a mountain would move in a parabolic path if there were no air resistance, and hit the ground some distance away, as shown again by path A in Figure D1.17. If the cannonball was travelling fast enough, it could move as shown in path B, and 'escape' from the Earth. Path C shows the path of an object moving with exactly the right speed and direction so that it remains at the same **altitude** (distance above the Earth's surface), that is, it remains in orbit around the Earth: a satellite. Remember that we are assuming that there is no air resistance.

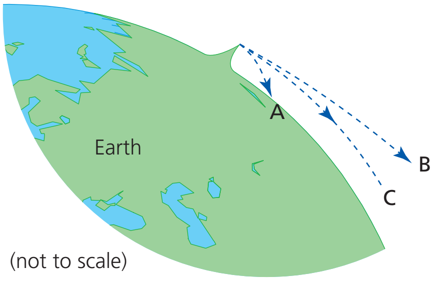


Figure D1.17 The path of objects projected at different speeds from a mountain top

Gravity is the only force acting on the satellite and it is acting continuously and perpendicularly to its instantaneous velocity along path C. As we described in Topic A.2, this is the necessary condition for circular motion. The force of gravity (weight of satellite) is providing the centripetal force.

An actual satellite needs to be at a height which is at least 200 km above the Earth's surface to avoid the effects of air resistance and, at that height, the orbital speed needed is about 8 km s^{-1} (explained below). This was first achieved in 1957 by the Soviet Union with their satellite Sputnik One. Its lowest orbital height was 215 km and it took 96 minutes for each orbit. See Figure D1.18.

Figure D1.19 shows a satellite of mass *m* in orbit around a much larger mass, *M* (a planet, for example). In the mathematical treatment of satellites in this course, we will only consider perfectly circular orbits with constant radius *r*, as shown.

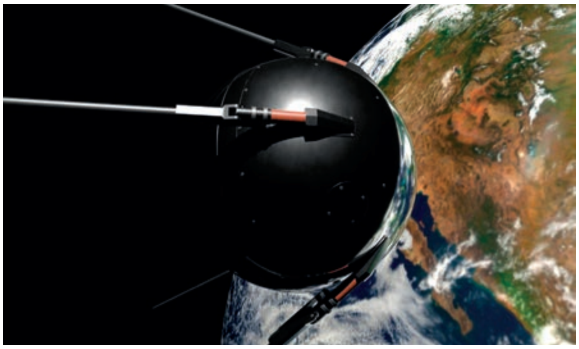


Figure D1.18 Sputnik One

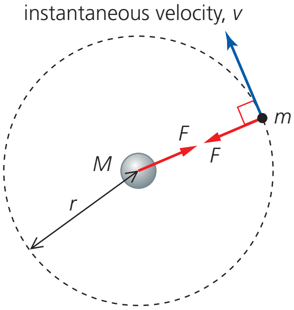


Figure D1.19 A satellite of mass *m* orbiting a planet of mass *M*.

Orbital speed and time period of a satellite

Remembering the equation for centripetal acceleration, we can write:

$$\text{centripetal acceleration, } g = \frac{v^2}{r}$$

Or, considering forces:

$$\text{centripetal force, } mg = \frac{mv^2}{r}$$

In order to maintain a satellite in a circular orbit around the Earth (or other planet), it needs to be given the correct speed, *v*, for its particular radius, *r*, as given by:

$$\frac{v^2}{r} = g$$

This equation enables us to determine the theoretical speed for a satellite which orbits just above the Earth's surface – as in the cannonball thought experiment (Earth's radius = 6.4×10^6 m):

$v^2 = gr = 9.8 \times (6.4 \times 10^6)$

$v = 7.9 \times 10^3\ \text{m s}^{-1}\ (7.9\ \text{km s}^{-1})$

At a more realistic height of 200 km (for example), $g = 9.23\,\text{N kg}^{-1}$ calculated from $g = \frac{GM}{r^2}$, so that the necessary speed is reduced to $7.7\,\text{km s}^{-1}$. Assuming there is no air resistance, a satellite moving with a speed of $7700\,\text{m s}^{-1}$, 200 km above the surface, can orbit the Earth. Knowing the value of g at any particular height enables us to calculate the speed necessary for a circular orbit at that height. The speed does not depend on the mass. All satellites at the same height move with the same speed. If there is no air resistance, a satellite in a circular orbit will continue to orbit the Earth without the need for any engine. The force of gravity acts perpendicularly to motion, so that no work is done by that force.

$$\boldsymbol{v = \frac{2\pi r}{T}}$$

(Topic A.2) can be used to calculate the time period, T, for an orbit.

If the speed of a satellite is greater than the speed necessary for a circular orbit, but less than the *escape speed* (explained later) it will move in an elliptical path. However, for calculations in this course, we will assume that the orbits of planets, moons and satellites are circular.

WORKED EXAMPLE D1.8

Determine the orbital speed and time period of a satellite that orbits the Earth at a distance which is as far above the surface as the centre of the Earth is below.
(Radius of Earth = 6.4×10^6 m)

Answer

At twice the distance from the centre of the Earth the gravitational field strength is reduced to:

$$\frac{9.8}{2^2} = 2.45\,\text{N kg}^{-1}$$

Then:

$$g = \frac{v^2}{r}$$

$$2.45 = \frac{v^2}{(2 \times 6.4 \times 10^6)}$$

$$v = 5.6 \times 10^3\,\text{m s}^{-1}$$

$$T = \frac{2\pi r}{v} = \frac{2 \times \pi \times (6.4 \times 10^6)}{5.6 \times 10^3}$$

$$= 7.2 \times 10^3\,\text{s (two hours)}$$

22 a Calculate values for the gravitational field strengths at heights above the Earth's surface of 1000 km, 10 000 km and 40 000 km.

b Calculate the necessary speeds for circular orbits at these heights.

c Use a compass to draw a scale diagram of the Earth with these orbits around it.

d Determine the times for complete orbits (time periods), T, at these heights and mark them on your diagram.

23 Two satellites of equal mass orbit the same planet as shown in Figure D1.20. Satellite B is twice as far away from the centre of the planet as satellite A. Copy the table and complete it to show the properties of the orbit of satellite B.

	Satellite A	**Satellite B**
Distance from planet's centre	r	$2r$
Gravitational field strength	g	
Gravitational force	F	
Circumference of orbit	c	
Speed	v	
Time period	T	

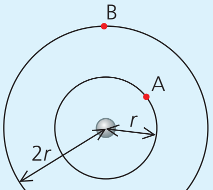


Figure D1.20 Two satellites of equal mass orbiting the same planet

24 The Earth is an average distance of 1.5×10^{11} m from the Sun. Assuming that the orbit is circular,

- **a** calculate the average orbital speed of the Earth around the Sun
- **b** determine the centripetal acceleration of the Earth towards the Sun.
- **c** Use your answers to calculate a value for the mass of the Sun.

25 The gravitational field strength on the surface of Mars is $3.72\,\text{N kg}^{-1}$.

The radius of Mars is 3.4×10^6 m.

- **a** What is the gravitational field strength at a distance of 3.4×10^6 m above the surface?
- **b** Calculate the orbital speed necessary for a satellite orbiting Mars at this height.
- **c** What would be the time period of this satellite?

◆ **Polar satellite orbit** Descriptive of the path of a low-orbit satellite that passes over the poles of the Earth and completes many orbits every day.

◆ **Geosynchronized orbit** Any satellite orbit that has the same period as the Earth spinning on its axis. The orbit must have exactly the correct radius.

◆ **Geostationary orbit** A satellite which appears to remain 'above' the same location on the Earth's surface.

Nature of science: Observations

Uses of satellites

The use of artificial satellites has extended the range and nature of observations that we can make of phenomena both on the Earth and in space. Satellites can be put into orbit at any desired height above the Earth's surface, assuming that they are given the right velocity and are high enough to avoid air resistance. The lower orbits have obvious advantages, especially if the Earth's surface is being monitored for some reason. But, as previously explained, the higher a satellite, the longer its orbital time period. The orientation of the orbit compared with the Earth's axis and whether it is circular or elliptical are also important.

Polar orbits

Many satellites have orbits that pass approximately over both poles of the Earth at heights of up to about 2000 km (Figure D1.21). The **Polar orbits** remain in the same plane as the Earth rotates, so that the satellite passes over different parts of the planet on each orbit. These satellites make many orbits every day.

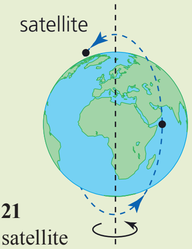


Figure D1.21 Polar-orbiting satellite

Geostationary orbits

A **geostationary** satellite is one that appears to remain in the same position as seen by an observer on the Earth's surface. This is only possible if the period of the satellite is the same as the period of rotation of the Earth (23 h 56 min). As we have seen, this requires that a geostationary satellite is at exactly the right height: 4.2×10^7 m.

Any satellite at this height will have the same period as the rotation of the Earth and they are described as **geosynchronous**. However, to be geostationary the satellites must be made to orbit in the same plane as the equator. Both orbits shown in Figure D1.22 are geosynchronous, but only the orbit in **a** is geostationary.

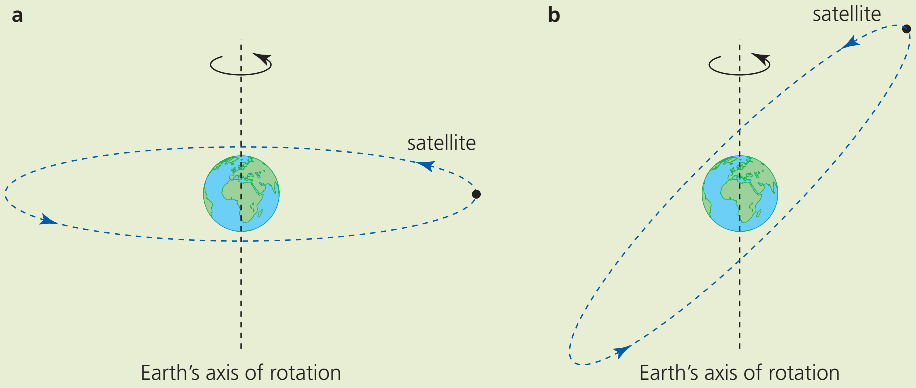


Figure D1.22 Geostationary orbits must be in the plane of the equator

LINKING QUESTION

- How can the motion of electrons in the atom be modelled on planetary motion and in what ways does this model fail? (NOS)

This question links to understandings in Topics E.1 and E.2.

Gravitational potential energy

SYLLABUS CONTENT

- ▶ Gravitational potential energy, E_p, of a system is the work done to assemble the system from infinite separation of the components of the system.
- ▶ Gravitational potential energy for a two-body system as given by: $E_p = -G\frac{m_1m_2}{r}$ where r is the separation between the centres of mass of the two bodies.

◆ **Gravitational potential energy**, E_p, is the work done when bringing all the masses of a system to their present positions from infinity.

We introduced the concept of **gravitational potential energy**, E_p, in Topic A.2. The equation $\Delta E_p = mg\Delta h$ was used to calculate *changes* in gravitational potential energy close to the Earth's surface, where the gravitational field strength, g, can be considered constant ($9.8\,\text{N kg}^{-1}$).

However, a general understanding of gravitational potential energy, which applies *anywhere*, must answer these two questions:

- Is there a true zero of gravitational potential energy and, if so, where is it?
- How can varying values of gravitational field strength be including in gravitational energy calculations?

The zero of gravitational potential energy is chosen to be where the masses are separated by an infinite distance.

TOK

Mathematics and the arts

- Why is mathematics so important in some areas of knowledge, particularly the natural sciences?
- How do mathematicians reconcile the fact that some conclusions seem to conflict with our intuitions?

Infinity

◆ **Infinite** Without limits.

◆ **Finite** Limited.

Infinity is not an actual place, but an abstract concept that appears regularly in physics and mathematics.

The idea of an **infinite** quantity (gravitational field, distance, time, number, and so on) is used in physics to suggest a quantity that is limitless (without end). It is greater than any real, measurable quantity. The opposite of infinite is **finite**, which means within limits. The idea of an infinite series of numbers, an infinite time, or even a field that extends for ever (but becomes vanishingly small) may all seem somehow acceptable to the human mind. However, the concept of an infinite universe gives most of us problems.

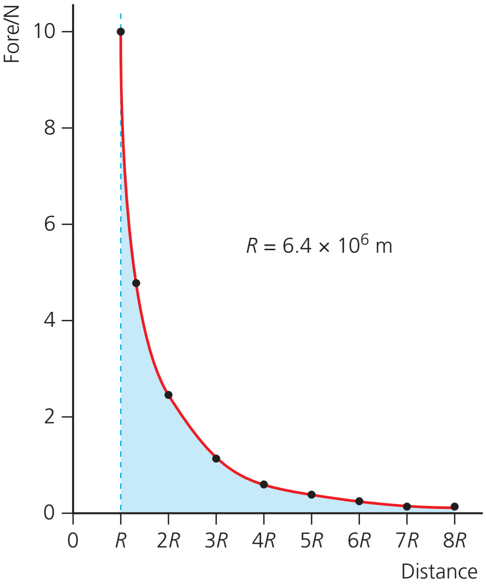
We usually refer to the gravitational potential energy of a single mass, a book on a table for example but, more exactly, the gravitational potential energy is a property of the whole *system* of the book, the table, the rest of the Earth (and the rest of the Universe!). In practice, it is acceptable to talk about the gravitational potential energy of a single object that is very much less massive than the mass creating the gravitational field in which it is situated (a person on the Earth, for example).

Gravitational potential energy is stored between two or more masses because of the gravitational forces between them. In theory, the forces never reduce to zero (consider Newton's law of gravitation), no matter how large the distances.

A 1 kg book placed on a table top which is 0.80 m above the floor has about 8 J ($mg\Delta h$) more gravitational potential energy than if it were placed on the floor. We may consider that the same book on the same table has about 40 J of gravitational potential energy if the room where they are located was on the first floor. If the location was 200 m above sea level, we might say that the book has 2000 J of gravitational potential energy. Defining a zero of gravitational energy which is agreed by everyone (infinity) avoids all these differences and possible misunderstandings.

The total gravitational potential energy of a system, E_p, is defined as the work done when bringing all the masses of the system to their present positions, assuming that they were originally at infinity.

The gravitational potential energy of the book (and the Earth) is the work done in bringing them together from an infinite distance apart. Of course, this is a theoretical exercise, but the determination is straightforward, as follows.



■ **Figure D1.23** Variation of the force on a 1 kg mass with distance from the Earth.

We know (from Topic A.3), work done = force × distance moved in the direction of the force.

The gravitational force between the 1 kg book varies with distance from the Earth, but we know that the work done can be determined from the area under a force–distance graph. See Figure D1.23, in which R represents the radius of the Earth (6.4×10^6 m).

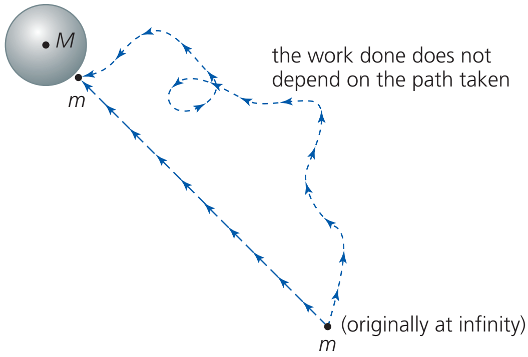
Of course, a separation of 8R is a long way from infinity (!), but we can see from the graph that the force is becoming very small and the gravitational potential energy (= shaded area) is tending to a limit. An accurate determination of the complete area (or the use of calculus) shows that the gravitational potential energy of 1 kg on the Earth's surface is 6.26×10^7 J. However, as explained below, gravitational potential energy is *always* given a negative sign, so that $E_p = -6.26 \times 10^7$ J.

If a mass of 1 kg on the Earth's surface was given $+6.26 \times 10^7$ J of energy, then it would have just the right amount of energy to reach infinity, where it would then have $(-6.26 \times 10^7 \text{ J}) + (6.26 \times 10^7 \text{ J}) = 0$ J of gravitational potential energy.

Top tip!

In this course we show gravitational forces as positive, but it may be considered that gravitational forces between two masses should be shown with negative signs because they are vectors with directions opposite to that of increasing separation. That would then be consistent with gravitational potential energy always being negative.

However, because gravitational forces are always attractive (unlike the electric forces discussed in Topic D.2), we are using only positive signs.



■ **Figure D1.24** Work done when moving in a gravitational field is independent of the path

Gravitational potential energies are *always* given negative values because (positive) energy would have to be supplied to separate the masses to infinity, where a system then has zero gravitational potential energy.

The total work done when moving to, or from, the same locations in a gravitational field does not depend on the path taken. See Figure D1.24.

A ball thrown upwards has been given kinetic energy. As it rises its kinetic energy is transferred to gravitational potential energy. As it falls, the process is reversed. This physics principle is exactly the same for a mass moving large distances, even (in theory) to infinity and back. See Figure D1.25.

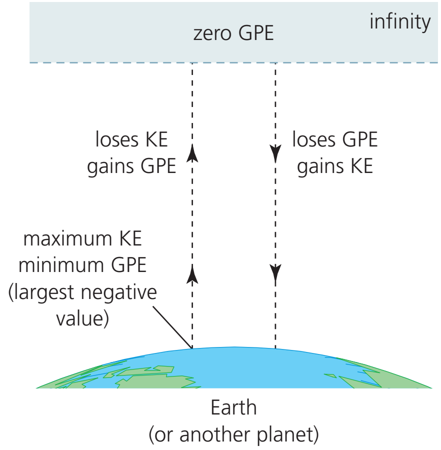


Figure D1.25 Changes of energy when a projectile moves between a planet and infinity

WORKED EXAMPLE D1.9

a Use an area under the graph shown in Figure D1.23 to estimate the change in gravitational potential energy if a mass of 1 kg moved 'up' from the Earth's surface (distance = R) to a height of $2R$ above the surface.

b Determine the change in gravitational potential energy if a 5.0 kg mass moved the same distance 'down' towards the Earth's surface.

c Compare your answer to part **a** to the value obtained using $\Delta E_p = mg\Delta h$, with $g = 9.8\,\text{N kg}^{-1}$.

Answer

a Area ≈ $4.0 \times (2R - R) = 4.0 \times (6.4 \times 10^6) = +2.6 \times 10^7\,\text{J}$

This is a rough estimate of the amount of energy needed to be given to the mass to increase its height above the Earth's surface, and its gravitational potential energy.

b $5.0 \times (-2.6 \times 10^7) = -1.3 \times 10^8\,\text{J}$

The negative sign arises because it represents the fact that the mass has lost gravitational potential energy as it moved closer to the Earth.

c $\Delta E_p = mg\Delta h = 1 \times 9.8 \times (6.4 \times 10^6) = 6.3 \times 10^7\,\text{J}$

It should be clear that this is a very different, and incorrect, answer.

Equation for gravitational potential energy

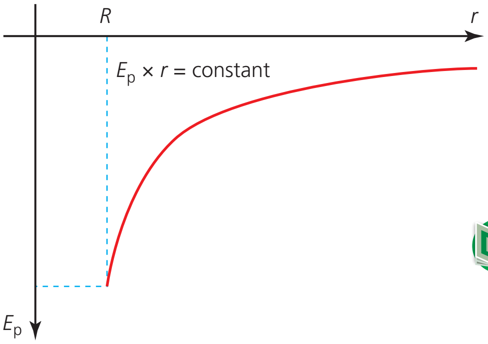


Figure D1.26 Variation of the gravitational potential energy of a mass, E_p, with distance, r, from the surface of a planet or moon of radius R

Clearly it would be inconvenient to have to use graphs to determine every gravitational potential energy. We need a direct equation. However, because the gravitational force is not constant, this is not obtained by a straightforward calculation. It requires calculus, which is not needed for this course. The relationship can be stated as:

gravitational potential energy between two masses, $\boldsymbol{E_p = -G\frac{m_1m_2}{r}}$

See Figure D1.26 for a graphical representation of this relationship.

This equation is consistent with work done = force × distance

$$= \frac{Gm_1m_2}{r^2} \times r$$

WORKED EXAMPLE D1.10

Use the equation above to confirm that the gravitational potential energy of 1.0 kg on the Earth's surface is −6.3 × 10⁷ J (as stated previously).

Mass of Earth = 6.0 × 10^{24} kg.
Radius of Earth is 6.4 × 10^6 m.

Answer

$$E_p = -G\frac{m_1m_2}{r}$$

$$= -\frac{(6.67 \times 10^{-11}) \times (6.0 \times 10^{24}) \times 1.0}{6.4 \times 10^6}$$

$$= 6.3 \times 10^7 \text{ J}$$

26 Explain how it is possible for a body to have negative potential energy.

27 A satellite of mass 200 kg was raised from the Earth's surface to a height of 320 km.

- **a** Determine the change in its gravitational potential energy.
  Mass of Earth = 6.0 × 10^{24} kg
  Radius of Earth = 6.4 × 10^6 m
- **b** What value would the (incorrect) use of using $\Delta E_p = mg\Delta h$ produce?

28 **a** Determine a value for the gravitational potential energy associated with the Earth–Moon system (only).
Mass of Moon = 7.3 × 10^{22} kg
Separation of centres averages at 3.8 × 10^8 m

- **b** Discuss whether it is acceptable to ignore the effect of the Sun in this calculation

29 **a** Use the graph shown in Figure D1.23 to estimate the change in gravitational potential energy when 1 kg moves from a distance 4*R* to a distance 5*R* from the centre of the Earth.

- **b** Compare your answer to a value determined by using the equation:

$$E_p = -G\frac{m_1m_2}{r}$$

30 Calculate the gravitational potential energy of a 200 kg satellite orbiting at a height of 150 km above the surface of the planet Mars.
(The mass of Mars = 6.0 × 10^{23} kg,
radius of Mars = 3.4 × 10^6 m)

Top tip!

The energies of electrons in atoms (and nucleons in nuclei) are also given negative values for the same reason: an attractive force results in a confined system. A negatively charged electron needs to be given energy to free it from the attractive force between it and the positively charged nucleus. After removal, an electron is then considered to have zero electrical potential energy. This is similar to saying that a mass has zero gravitational potential energy when it is a great distance from a planet.

Gravitational potential

SYLLABUS CONTENT

- ▶ Gravitational potential, V_g, at a point, is the work done per unit mass in bringing a mass from infinity to that point as given by: $V_g = -G\frac{M}{r}$.
- ▶ Equipotential surfaces for gravitational fields.
- ▶ The relationship between equipotential surfaces and gravitational field lines.

When calculating gravitational potential energies, we refer to a particular mass in a particular place. But, if we want to answer questions such as: 'how much energy would be needed to put a mass in a specified location?', it is better to use the more generalized concept of **gravitational potential**, V_g. The concept of gravitational potential is used to describe the space around massive objects such as planets, stars, and so on.

◆ **Gravitational potential,** V_g Work done in moving a test mass of 1 kg to a specified point from infinity.

Gravitational potential can be considered as gravitational potential energy per unit mass.

For a relatively small mass, *m*, in the gravitational field of a very much larger mass, *M*, (such as a planet), we can make that clear by rewriting:

$$E_p = -\frac{Gm_1m_2}{r}$$

$$\text{as } E_p = -G\frac{Mm}{r}$$

Then, diving by the small mass, *m*, gives:

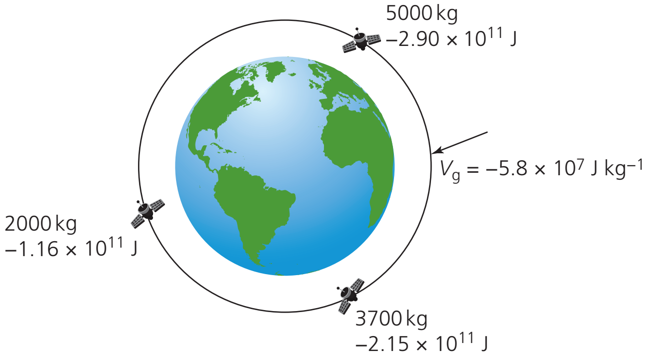


Figure D1.27 Three satellites in orbit

$$\textbf{gravitational potential, } V_g \left(= \frac{E_p}{m}\right) = -G\frac{M}{r}$$

Figure D1.27 may help to explain the usefulness of the concept of potential. It shows three satellites, all at the same height in orbit around the Earth.

Because the satellites have different masses, they have different gravitational potential energies. If we divide their gravitational potential energies by their masses, we get the same result: $-5.8 \times 10^7\,\text{J kg}^{-1}$ (the gravitational potential in that particular orbit).

Calculations with any satellite in the same orbit will produce the same result, and we can label that orbit as having a gravitational potential of $-5.8 \times 10^7\,\text{J kg}^{-1}$.

A more formal definition of gravitational potential:

The gravitational potential at a point is defined as the work done per unit mass (1 kg) in bringing a small test mass from infinity to that point.

Gravitational potential is a scalar quantity; it has no direction. Like gravitational potential energy, the zero of gravitational potential is defined to be at infinity and all values of gravitational potential energy are negative. The SI unit for gravitational potential is J kg^{-1}.

WORKED EXAMPLE D1.11

Consider Figure D1.27.

a If a satellite of mass 4250 kg was placed in the same orbit, calculate its gravitational potential energy.

b Determine the height of the orbit above the Earth's surface.

(Mass of Earth = 6.0×10^{24} kg.
Radius of Earth = 6.4×10^6 m.)

Answer

a gravitational potential energy = potential × mass

$= -(5.8 \times 10^7) \times 4250$

$= -2.5 \times 10^{11}$ J

b $V_g = -G\frac{M}{r}$

$$-5.8 \times 10^7 = -\frac{(6.67 \times 10^{-11}) \times (6.0 \times 10^{24})}{r}$$

$r = 6.9 \times 10^6$ m

height $= (6.9 \times 10^6) - (6.4 \times 10^6)$

$= 5 \times 10^5$ m (500 km)

Common mistake

Students frequently get confused between *potential energy* and *potential*. Perhaps because of the similarity in their names. It may be helpful to use a shopping analogy. A shop may have a wide selection of different sized packets of rice (for example), all at different prices. When faced with such a selection, the most useful information to the shopper is not the prices of each packet, but the prices per unit mass (kg). Similarly, energy per unit mass (gravitational potential) is usually more useful information than individual energies.

31 Determine the gravitational potential on the Earth's surface.

32 If the gravitational potential on the surface of a planet is −4.8 MJ kg^{-1}, determine the gravitational potential energy of an 86 kg mass on the planet.

33 a Calculate the gravitational potential at a height of 1000 km above the surface of Mars. (Mass = 6.4×10^{23} kg, radius = 3.4×10^{6} m)

b Determine the change in gravitational potential energy for a mass of 1200 kg moving from the surface of Mars to a height of 1000 km.

c State whether the change in gravitational potential energy is positive or negative.

34 The gravitational potential at a distance of 1.4×10^{7} m from a planet is $–1.9 \times 10^{7}$ J kg^{-1}. Calculate the gravitational potential at a distance of 3.7×10^{8} m.

◆ **Equipotential line (or surface)** Line (or surface) joining points of equal potential. Equipotential lines are always perpendicular to field lines.

Equipotential surfaces

Drawings of equipotential lines provide useful visualizations of gravitational fields.

An **equipotential surface (or line)** connects places which have the same potential.

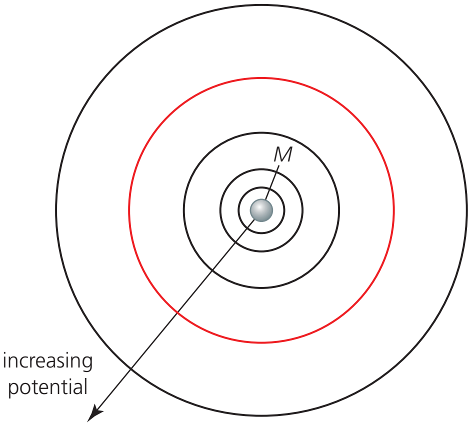


Figure D1.28 Equipotential lines around a spherical mass

Figure D1.28 shows typical equipotential lines around a spherical or point mass, *M*. The circular lines are drawn with equal numerical intervals of potential, which means that they must get further and further apart. In other words, increasing separation of equipotential lines indicates a weakening gravitational field. A three-dimensional representation would have spherical *surfaces*.

As the distance from the mass increases, the potential increases, but we know that the potential on the surface of the mass *M* is negative. This means that the increasing potential is indicated by a negative value decreasing in magnitude.

All equipotential lines form closed loops. A satellite of mass *m,* placed anywhere on any particular equipotential (the red line, for example), will have the same gravitational potential energy. Moving from any one location to any other on the same equipotential line, by any path, requires zero *net* energy input.

◆ **Contour lines** Lines on a map joining places of the same altitude.

Contour lines on a geographical map (see Figure D1.29) are similar to equipotential lines. The lines join places of equal vertical heights (above sea level), which in effect are equipotential lines. Where the lines are closest, the ground is steepest and anything that is free to move, such as water in a river for example, will move perpendicular to the contours.

Equipotential lines and gravitational field lines (as seen in Figure D1.13) can both be used to provide visualizations of the same gravitational field. Figure D1.30 shows their simple relationship:

Field lines are always perpendicular to equipotential lines. They point from higher potential to lower potential.

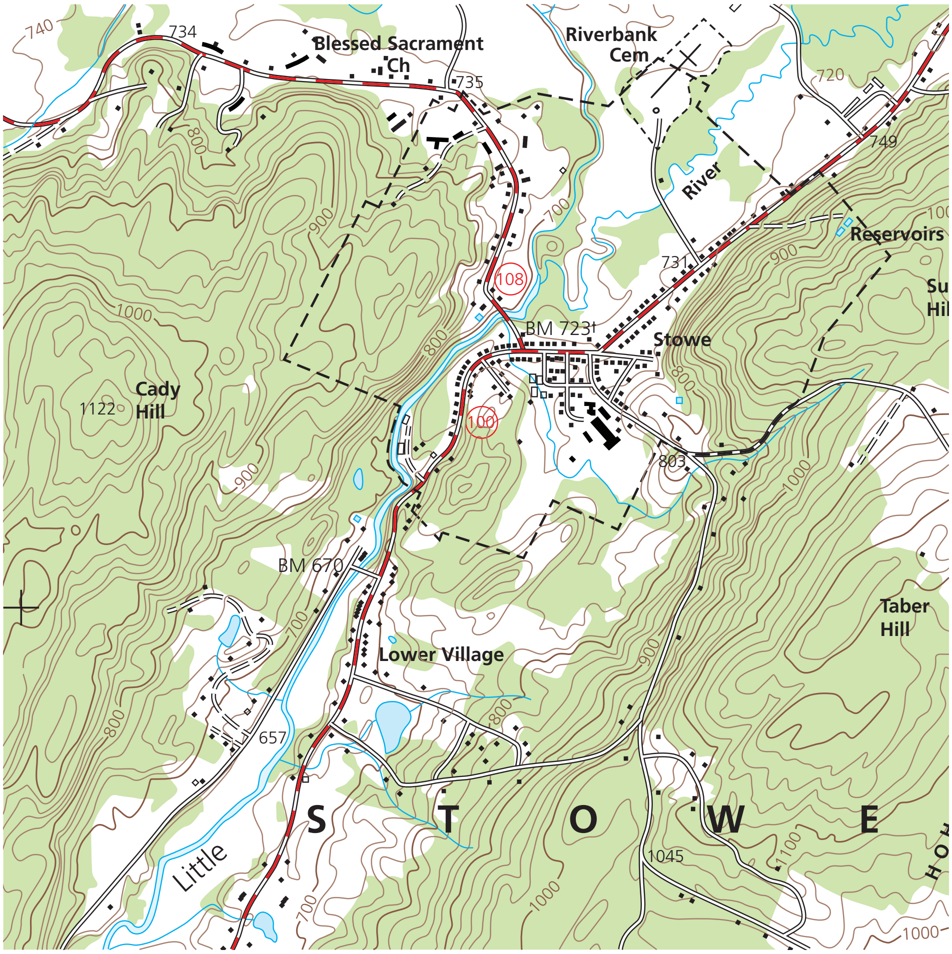


Figure D1.29 Contour map of Stowe, Vermont, USA

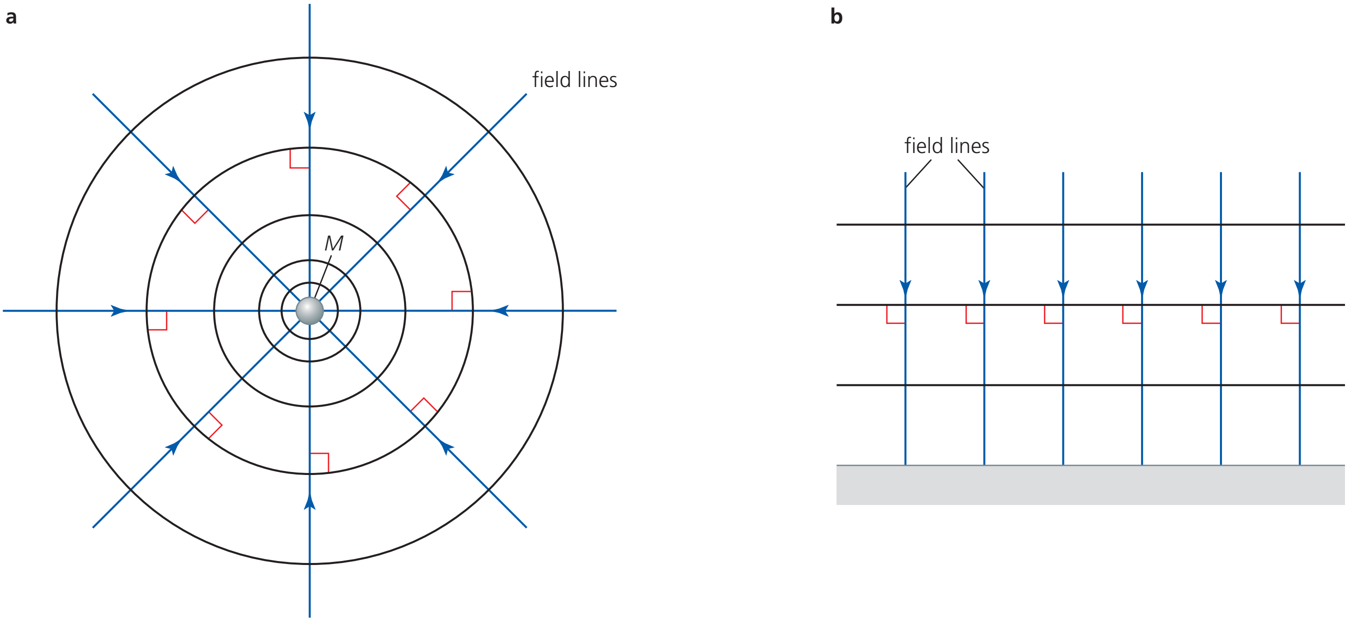
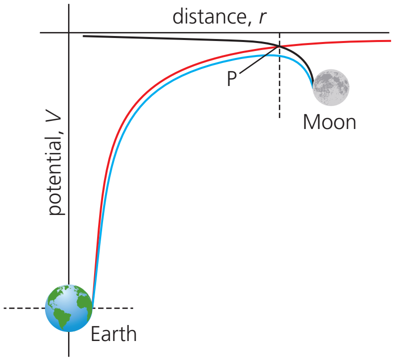


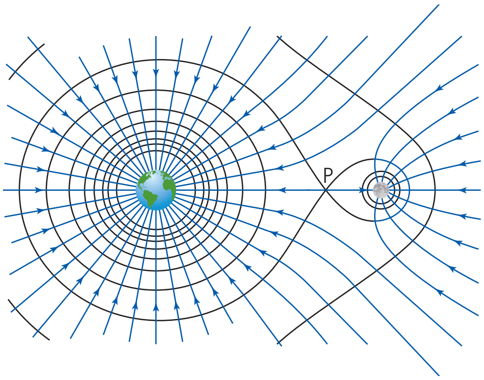
Figure D1.30 Equipotential lines and field lines are perpendicular to each other. **a** radial field; **b** uniform field

Gravitational potential is a scalar quantity, unlike gravitational field strength. This means that, if there are two (or more) large masses creating significant fields, we can determine the potential at any point by simple addition. Figure D1.31 shows an example: the potentials in the Earth–Moon system. The blue line represents the combined potential of the system.

Figure D1.32 shows the equipotential lines (black) and field lines (blue) in the same system.



■ **Figure D1.31** The potentials in the Earth–Moon system



■ **Figure D1.32** Equipotential and field lines around the Earth and Moon.

35 Calculate suitable values of the gravitational potential around the Earth that will enable you to draw a *scale* diagram showing at least three equipotential lines.

36 Discuss the significance of point P seen in Figures D1.31 and D1.32.

37 Explain why contour lines on a geographic map can be considered to be equipotential lines.

Tool 2: Technology

Use spreadsheets to manipulate data

Set up a spreadsheet which will enable you to calculate the combined potentials along a straight line joining the Earth and the Sun. Use the spreadsheet to draw a graph similar to that seen in Figure D1.31.

Gravitational potential difference

SYLLABUS CONTENT

- ▶ Work done in moving a mass, *m*, in a gravitational field as given by: $W = m\Delta V_g$.
- ▶ Gravitational field strength, *g*, as the gravitational potential gradient as given by: $g = -\dfrac{\Delta V_g}{\Delta r}$.

◆ **Gravitational potential difference** Difference in gravitational potential between two points, which equals the work done when 1 kg is moved between the points.

The central theme of this topic is the movement of masses between different places in gravitational fields, for example, the field around the Earth. This means that the difference in potential – the *potential difference* – between two locations is of particular importance.

Gravitational potential difference, ΔV_g, is the work, *W*, done on unit mass (1 kg) when it moves between two points in a gravitational field.

$$\Delta V_g = \frac{W}{m} \text{ or}$$

$$W = m\Delta V_g$$

Work has to be done *on* a mass to increase its gravitational potential energy; that is to move it to a greater potential. (For example, away from a planet to a location where the potential has a lesser negative value.) Then *W* will have a positive value. The mass does the work when it moves to a lesser potential, and *W* will have a negative value. For example, when a mass falls towards the Earth, gravitational potential energy will be transferred from the mass to kinetic energy.

Top tip!

For comparison, you should remember, from Topic B.5, that in electric circuits, the electric potential difference was the defined as the work done *per unit charge* as given by:

$$V = \frac{W}{q}$$

This concept will be developed further in discussion of *electric* fields in Topic D.2.

WORKED EXAMPLE D1.12

A satellite of mass 850 kg is in an orbit 7.9×10^6 m from the centre of the Earth.

a Calculate the gravitational potential in this orbit.

b The satellite is to be moved to an orbit where the gravitational potential is $-5.40 \times 10^7 \text{ J kg}^{-1}$. State whether this is a higher, or lower orbit.

c Calculate the work done in this change of orbit.

Answer

a $V_g = -G\frac{M}{r} = -\frac{(6.67 \times 10^{-11}) \times (6.0 \times 10^{24})}{7.9 \times 10^6}$

$= -5.1 \times 10^7 \text{ J kg}^{-1}$

($5.0658.... \times 10^7$ seen on calculator display)

b The potential is lower in the new orbit (greater negative value), so it must be a lower orbit. See Figure D1.33.

c $W = m\Delta V_g = (850 \times \left((-5.40 \times 10^7) - (-5.07 \times 10^7)\right) = -2.8 \times 10^9 \text{ J}$

The satellite will have to 'lose' this amount of gravitational potential energy to be in the lower orbit.

Note that the answer to **c** is not dependent on the path taken between the two orbits.

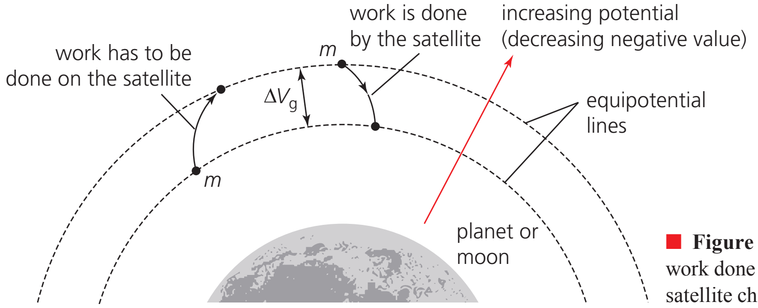


Figure D1.33 work done when a satellite changes orbit

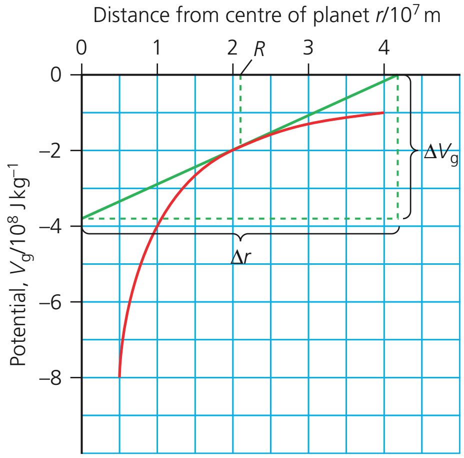


Figure D1.34 Graph showing variation of gravitational potential around a planet

◆ **Gravitational potential gradient** Rate of change of potential with distance, equal in magnitude to field strength.

Gravitational potential–distance graphs

Figure D1.34 shows the variation of gravitational potential around a spherical planet. We can use the gradient of the graph at any point ($\Delta V_g/\Delta r$) to determine the strength of the gravitational field, *g,* which can be explained as follows.

We know that the work done, *W*, when moving a mass, *m,* through a potential difference ΔV_g is given by $W = m\Delta V_g$.

We also know that the work can be calculated from force × distance = $mg \times \Delta r$, where Δr is a small enough distance that the value of *g* does not change significantly. (This assumption would need further explanation or justification at a higher level.)

Hence:

$W = \cancel{m}\,\Delta V_g = \cancel{m}\,g\Delta r$

So that the magnitude of the gravitational field strength $g = \frac{\Delta V_g}{\Delta r}$, called the **gravitational potential gradient**.

Gravitational field strength equals potential gradient: $\boldsymbol{g = -\frac{\Delta V_g}{\Delta r}}$

The negative sign has been added to the equation to show that the direction of the vector quantity *g* is opposite to the direction of increasing potential.

WORKED EXAMPLE D1.13

Determine the magnitude of the gravitational field strength at a distance *r* from the centre of the planet represented in Figure D1.34.

Answer

$$g = \frac{\Delta V_g}{\Delta r} = \frac{(3.8 \times 10^8) - 0}{(4.2 \times 10^7) - 0} = 9.0\,\text{N kg}^{-1}$$

38 Calculate the gravitational potential difference when moving up from a 100 m contour line to a 200 m contour line.

39 Figure D1.35 shows equipotential lines around two equal masses. Draw a sketch to represent the gravitational field lines around the same masses.

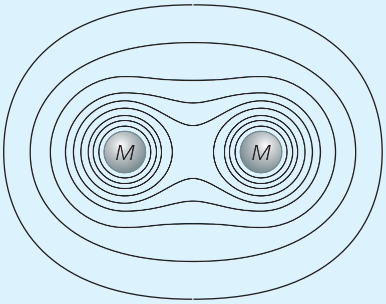


Figure D1.35 Equipotential lines around two equal masses

40 Figure D1.36 shows equipotential lines around a planet. Determine the gravitational potential differences associated with the following movements from:

a 1 to 5 **b** 5 to 1 **c** 3 to 4 **d** 2 to 5.

41 a Draw a graph of the potential around the Earth from its surface to a radius of 2.6×10^7 m.

b Use your graph to determine a value of the gravitational field strength at a radius of 1.7×10^7 m.

Line	Potential J kg^{-1}
A	-4×10^7
B	-6×10^7
C	-8×10^7
D	-10×10^7
E	-12×10^7
F	-14×10^7
G	-16×10^7

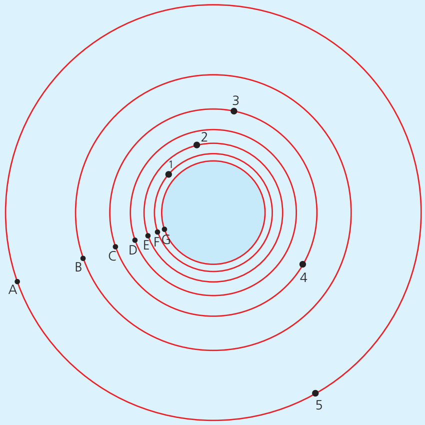


Figure D1.36 Equipotential lines around a planet

ATL D1A: Communication skills

Clearly communicating complex ideas in response to open-ended questions

A useful way to deepen your understanding is to present concepts visually, using charts or other visual organizers to show relationships between concepts. Here are two examples of how the concepts in this topic could be presented in this way:

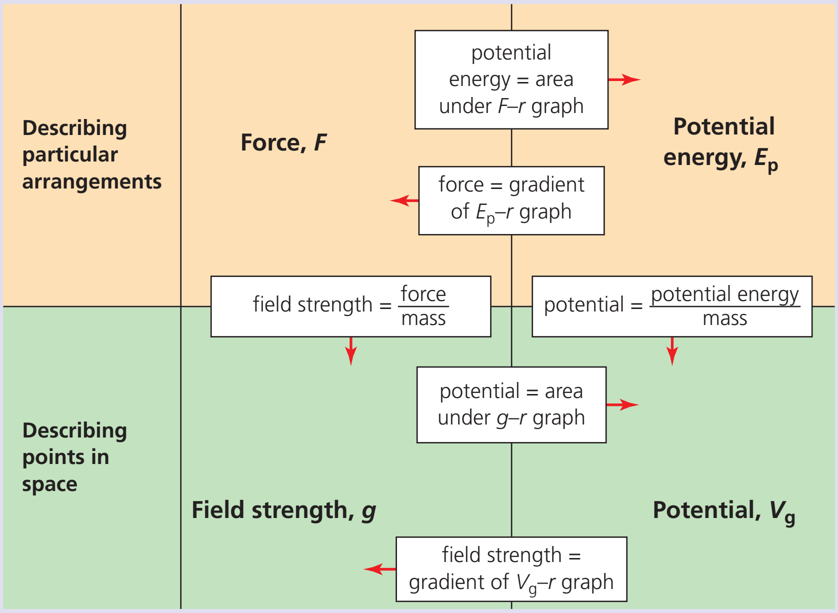
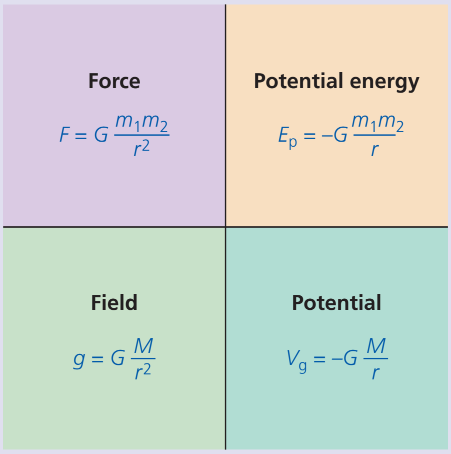


Figure D1.37 Connections between the magnitudes of the four key concepts



Force $F = G\frac{m_1 m_2}{r^2}$	**Potential energy** $E_p = -G\frac{m_1 m_2}{r}$
Field $g = G\frac{M}{r^2}$	**Potential** $V_g = -G\frac{M}{r}$

Figure D1.38 Equations for radial gravitational fields

Can you think of other ways in which these concepts – and the relationships between them – could be represented?

Speeds and energies of satellites

SYLLABUS CONTENT

- The escape speed v_{esc} at any point in a gravitational field as given by: $v_{esc} = \sqrt{\frac{2GM}{r}}$.
- The orbital speed $v_{orbital}$ of a body orbiting a large mass as given by: $v_{orbital} = \sqrt{\frac{GM}{r}}$.
- The qualitative effect of a small viscous drag due to the atmosphere on the height and speed of an orbiting body.

In all of this section, for the sake of simplicity, we will assume that the planet from which a satellite is launched is not rotating.

Escape speed

In theory, an object can be *projected* (not powered) upwards in such a way that it could continue to move away from the Earth forever. For this to be possible the object would need to be given a very high speed. To calculate that speed, we need to consider energies.

◆ **Escape speed** Minimum theoretical speed that an object must be given in order to move to an infinite distance away from a planet (or moon, or star): $v_{esc} = \sqrt{\frac{2GM}{r}}$.

In general, the initial kinetic energy given to an object will be transferred to gravitational energy, but also dissipated due to air resistance in the Earth's atmosphere. But if we assume that the effects of air resistance are negligible, we can calculate the *minimum* theoretical speed that a projectile of mass *m* needs in order to 'escape' from a planet of mass *M*. This is called its **escape speed**, v_{esc}. It can be calculated as follows:

kinetic energy needed = change in gravitational potential energy between location and infinity

$$\frac{1}{2}mv_{esc}^2 = 0 - \left(\frac{-GMm}{r}\right)$$

which leads to:

speed needed to 'escape' from a planet without air resistance (drag), $v_{esc} = \sqrt{\frac{2GM}{r}}$

Note that:

- This speed is the same, regardless of direction of travel (assuming no air resistance).
- We have assumed that there are no other significant gravitational fields, as might be provided by, for example, a moon.

WORKED EXAMPLE D1.14

a Calculate the escape speed from the Earth's surface.

b Outline why all masses have the same escape speed.

Answer

a $v_{\text{esc}} = \sqrt{\frac{2GM}{r}} = \sqrt{\frac{(2 \times (6.67 \times 10^{-11}) \times (6.0 \times 10^{24}))}{6.4 \times 10^{6}}} = 1.1 \times 10^{4}\,\text{m s}^{-1}$

(about 11 km s^{-1})

b Double the mass, for example, will need to gain double the gravitational potential energy. So, it needs double the kinetic energy, which it will have with double the mass at the same speed.

LIINKING QUESTION

- How is the amount of fuel required to launch rockets into space determined by considering energy?

Although the equation above can be used for any radius, it assumes that the mass starts with zero kinetic energy. If, for example, we wanted to know the escape speed needed for a satellite already in a steady orbit, we need to take into consideration the initial kinetic energy of the satellite. See below.

Orbital speed

We have already seen that a satellite needs to have the correct **orbital speed** to maintain a circular orbit. From Kepler's third law, we showed that:

$$\frac{r^3}{T^2} = \frac{GM}{4\pi^2}$$

Replacing using T with $\frac{2\pi r}{v}$ gives us:

$$\frac{r^3 v^2}{4\pi^2 r^2} = \frac{GM}{4\pi^2}$$

which simplifies to (using v_{orbital} instead of v):

speed required to maintain a circular orbit, $\boldsymbol{v}_{\textbf{orbital}} = \sqrt{\frac{\boldsymbol{GM}}{\boldsymbol{r}}}$

◆ **Orbital speed** For a satellite in a circular orbit, its speed must have the correct value for the chosen radius: $v_{\text{orbit}} = \sqrt{\frac{GM}{r}}$.

WORKED EXAMPLE D1.15

What speed is required for a satellite to maintain a circular orbit 100 km above the Moon's surface? (mass of Moon = 7.35 × 10^{22} kg, radius of Moon = 1737 km)

Answer

$$v_{\text{orbital}} = \sqrt{\frac{GM}{r}}$$

$$= \sqrt{\frac{(6.67 \times 10^{-11}) \times (7.35 \times 10^{22})}{1.837 \times 10^{6}}}$$

$$= 1.63 \times 10^{3}\,\text{m s}^{-1}$$

WORKED EXAMPLE D1.16

The average distance between the centre of the Moon and the centre of the Earth is 3.84×10^8 m. The Earth has a mass of 6.0×10^{24} kg. Determine a value for the time for each orbit of the Moon around the Earth (assuming that its path is circular).

Answer

$$\frac{r^3}{T^2} = \frac{GM}{4\pi^2} \Rightarrow \frac{(3.84 \times 10^8)^3}{T^2} = \frac{(6.67 \times 10^{-11}) \times (6.0 \times 10^{24})}{4\pi^2}$$

$T = 2.4 \times 10^6$ s (27 days)

Artificial satellites

To put a satellite into orbit we need to provide enough energy to:

- increase the gravitational potential energy of the satellite
- increase the gravitational potential energy of the launch vehicle fuel, etc.
- give the satellite the required kinetic energy for the required orbit
- overcome fictional forces
- allow for thermal energy dissipation.

The gravitational and kinetic energy given to the orbiting satellite will only be a small percentage of the total energy transferred. This is not a very efficient process!

Once a satellite is in orbit, the energy relationships are less complicated. As we have seen:

gravitational potential energy of a satellite of mass *m* orbiting a planet, or moon, of mass *M* at a distance *r* from the planet's (or moon's) centre is given by:

$$E_p = -G\frac{Mm}{r}$$

For a satellite already in orbit:

$$\text{Total energy, } E_T = E_k + E_p = \frac{1}{2}mv_{\text{orbital}}{}^2 + \left(-G\frac{Mm}{r}\right)$$

But we know that:

$$v_{\text{orbital}}{}^2 = \frac{GM}{r}$$

so that:

$$\text{kinetic energy of a satellite in a circular orbit, } E_k = \frac{1}{2}\frac{GMm}{r}$$

$$\text{Then, total energy, } E_T = \frac{1}{2}\frac{GMm}{r} + \left(-\frac{-GMm}{r}\right)$$

so that:

$$\text{total energy of a satellite in a circular orbit, } E_T = -\frac{1}{2}\frac{GMm}{r}$$

Note that:

$$E_T = \frac{1}{2}E_p = -E_k$$

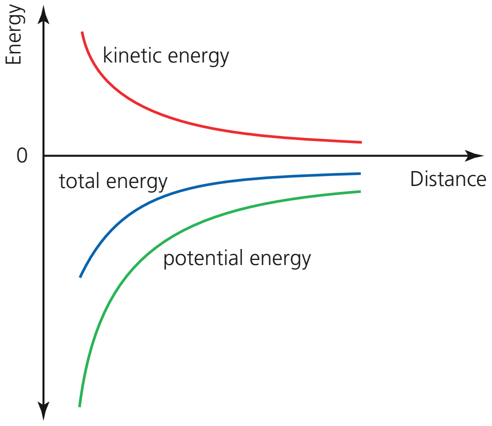


Figure D1.39 Energies of a satellite

A satellite in a circular orbit does not have enough energy to escape the gravitational field, so that both its potential energy and its total energy are negative. Figure D1.39 shows these relationships graphically.

WORKED EXAMPLE D1.17

A 500 kg satellite is orbiting at a height of 300 km above the surface of the planet Mars.

Mass of Mars = 6.4×10^{23} kg. Radius of Mars = 3.4×10^{6} m.

Determine the satellite's:

a gravitational potential energy

b kinetic energy

c total energy.

Answer

a $E_p = -G\dfrac{Mm}{r} = -\dfrac{(6.67 \times 10^{-11}) \times (6.4 \times 10^{23}) \times 500}{3.4 \times 10^{6}} = -6.3 \times 10^{9}\,\text{J}$

b $E_k = \dfrac{1}{2}G\dfrac{Mm}{r} = +3.1 \times 10^{9}\,\text{J}$

c $E_T = -\dfrac{1}{2}\dfrac{GMm}{r} = -3.1 \times 10^{9}\,\text{J}$

Changing orbits

- Continuing Worked example D1.17, suppose that it was required that a satellite was to be re-directed to an orbit 200 km higher. Then the new values are:
  $E_p = -5.93 \times 10^{9}$ J. This is an increase $\approx 0.35 \times 10^{9}$ J.
  $E_k = +2.96 \times 10^{9}$ J. This is a decrease $\approx 0.17 \times 10^{9}$ J.
  $E_T = -2.96 \times 10^{9}$ J. This is an increase $\approx 0.17 \times 10^{9}$ J.
  The satellite will have less kinetic energy because its necessary orbital speed is less, but there is a greater gain of gravitational potential energy due to the increased height. Overall, energy must be *supplied* for the change.
- If it required that a satellite *already in orbit* (radius *r*) is to 'escape' the Earth's gravity, we can calculate the extra energy needed from: required gain in gravitational potential energy = existing orbital kinetic energy + extra kinetic energy needed
  $$\text{extra kinetic energy needed} = \frac{GMm}{r} - \frac{1}{2}\frac{GMm}{r} = \frac{1}{2}\frac{GMm}{r}$$
  This is the same as its existing orbital kinetic energy.
- If a satellite is to be re-directed to a lower orbit, energy must be *removed*. It has to travel faster in its new orbit and gain kinetic energy, but there is an even greater reduction in gravitational potential energy.
- Satellites in low orbits may experience some very slight air resistance (viscous drag). This results in a dissipation of kinetic energy to thermal energy. It would travel more slowly, but as it moves to a lower height it gains an even greater amount of kinetic energy. Overall, its speed increases and the effects of increasing air resistance result in even greater dissipation of thermal energy. An uncontrolled satellite will spiral towards Earth, burn up and disintegrate.

LINKING QUESTION

- How can air resistance be used to alter the motion of a satellite orbiting Earth?

This question links to understandings in Topic A.2.

42 a Explain what you think ‘burn up’ means in the paragraph above.

b Research into an occasion when a satellite actually crashed on the Earth’s surface and find out what happened.

43 A typical air molecule travels at 450 m s^{-1} at room temperature. Explain why the Earth’s atmosphere does not spread out into space away from the planet.

44 A satellite of mass 820 kg is orbiting at a height of 320 km above the Earth’s surface. Calculate:

a its gravitational potential energy

b its kinetic energy

c its total energy (Earth’s radius = 6.4×10^6 m, Earth’s mass = 6.0×10^{24} kg)

d the speed it would need to have in order to escape from the Earth.

45 a A 300 kg satellite is in orbit around the Moon at an altitude of 60 km. Calculate how much extra energy it needs to escape from the Moon. (Mass of Moon = 7.3×10^{22} kg, radius of Moon = 1.7×10^6 m)

b State any assumptions you made in answering **a**.

46 a Determine a value for the escape velocity from the surface of the planet Mars. (Mass of Mars = 6.4×10^{23} kg, radius of Mars = 3.4×10^6 m)

b Outline why this escape speed is less than the escape speed from Earth.

47 A satellite in a geosynchronized orbit has a time period of 24 hours.

a Determine the radius of this orbit.

b Calculate the orbital speed of the satellite.

48 What happens to the total energy of a satellite in a circular orbit if it encounters some air resistance, moves to a lower orbit, but gains speed? Explain your answer.

49 a Suggest what effect the spin of the Earth will have on the escape speed.

b Suggest why satellite launch sites are often close to the equator.

50 Draw a Sankey diagram to represent the energy flows as a satellite is launched from the surface of the Earth and then enters an orbit. Assume that the whole process is very inefficient.

51 Titania is a moon of the planet Uranus. It orbits at an average distance of 4×10^8 m from the centre of Uranus. The planet has a mass of 8.7×10^{25} kg.

a Determine the time period (Earth days) of Titania’s orbit.

b Calculate its average orbital speed.

c Calculate the strength of the gravitational field of Uranus at the height of Titania.

ATL D1B : Research skills

Using search engines and libraries effectively

Use the internet to find out the latest progress on space launch systems that intend to propel spinning rockets upwards from the Earth’s surface by giving them a large amount of kinetic energy. After the rocket has significantly slowed down and reached an altitude of about 60 km, then the engines are ignited for the rest of the trip into orbit.

D.2 Electric and magnetic fields

Guiding questions

- Which experiments provided evidence to determine the nature of the electron?
- How can the properties of fields be understood using both an algebraic approach and a visual representation?
- What are the consequences of interactions between electric and magnetic fields?

Electric charge

SYLLABUS CONTENT

- ▶ The direction of forces between the two types of electric charge.
- ▶ The conservation of electric charge.

ATL D2A: Communication skills

Clearly communicating complex ideas in response to open-ended questions

The concept of electric charge was introduced in Topic B.5, in which the electric charges of electrons, protons and ions were briefly described. Before beginning this topic, review the beginning of Topic B.5: Electric charge and its conservation.

Make notes on the important concepts you find there. Use a visual organizer and/or diagrams to connect the key concepts.

Charge is measured in *coulombs*, C. One coulomb is a relatively large amount of charge and we often use microcoulombs (1 μC = 10^{-6} C) and nanocoulombs (1 nC = 10^{-9} C).

All protons have a positive charge of +1.60 × 10^{-19} C and all electrons have a negative charge of −1.60 × 10^{-19} C.

Any quantity of charge consists of a whole number of these charged particles, each ± 1.6 × 10^{-19} C. Charge is *quantized*.

1.6 × 10^{-1}9 C is called the elementary charge and it is given the symbol *e*.

(Note: nuclear particles – protons and neutrons – are themselves composed of smaller particles called *quarks*. Quarks also have quantized charge, but the charge is quantized into multiples of ±*e*/3. However, quarks cannot be observed as isolated particles. Knowledge of quarks is not required in this course.)

Later in this topic, we will describe Millikan's famous experiment to determine the charge of an electron.

LINKING QUESTION

- Charge is quantized. Which other physical quantities are quantized? (NOS)

This question links to understandings in Topic E.2.

Electrostatic charging and discharging

SYLLABUS CONTENT

- Electric charge can be transferred between bodies using friction, electrostatic induction and by contact, including the role of grounding (earthing).

◆ **Charge (to)** Add or remove electrons, so that an object acquires an overall net charge, for example, by friction. (To 'charge' a battery has a different meaning.)

◆ **Electrostatics** The study of the effects of charges at rest (that is, not electric currents).

Everyday objects contain an enormous number of charged particles and they usually have equal numbers of positive charges (protons) and negative (electrons) charges, so that they have no overall charge and are therefore described as being *neutral*.

Negatively charged electrons are the outermost particles in atoms and some of them are not tightly bound to atoms. If electrons can be added to, or removed from, a neutral object, it will then have an overall charge and we describe the object as being **charged**, which can then result in **electrostatic** effects. Protons, unlike electrons, are located in the nuclei of atoms and cannot be separated or moved from their positions, so they are not involved in producing electrostatic effects.

If a neutral object is given excess electrons, it becomes negatively charged. If the number of electrons is reduced, the object becomes positively charged.

Charging by friction

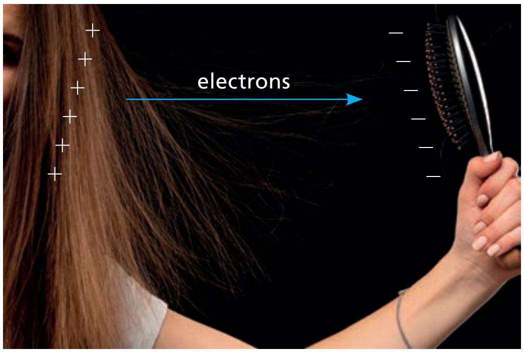


Figure D2.1 When you brush your hair, individual hairs may move apart because of the repulsion between similar charges

One common example is seen in Figure D2.1: when brushing dry hair with a plastic brush, electrons can be transferred in the process, one object (the hair brush) gains electrons, becoming negatively charged, while removing electrons from the other object (the hair), leaving it with a positive charge. The two objects will then attract each other. In this example, individual hairs with similar charge can also be seen to be repelled from each other.

In school experiments, in order to produce electrostatics effects, insulating rods are often rubbed with cloths. Depending on the materials, one will become positively charged and the other negatively charged, as electrons are transferred between the rod and the cloth. These insulating materials have much fewer mobile charge carriers (free electrons) than metals, but if metals were used, the charges would not stay in the same place.

When friction occurs between substances, electrons are often transferred between their surfaces.

Charging by friction is not limited to solids. For example, electrostatic effects can be produced when liquids flow through pipes, or when air flows past fans.

Very high electrostatic voltages can be produced using friction with specially designed (and safe) apparatus. Figure D2.2 shows an example.

Figure D2.2 Classroom electrostatic generator

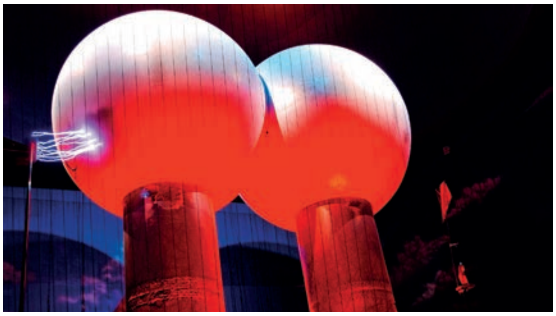


Figure D2.3 The world's largest Van de Graaff frictional electrostatic generator is in the Boston Museum of Science (USA). It has two 4.5 m diameter spheres mounted on 7.6 m insulating poles. It can generate a potential difference of up to seven million volts.

Charging and discharging by contact

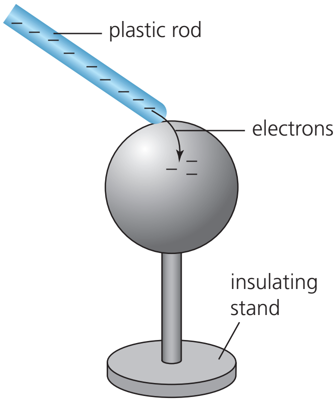


Figure D2.4 Charging a conducting sphere negatively by contact

If a charged object comes into actual physical contact with another object, it is possible for charges (electrons) to flow between them. Figure D2.4 shows a laboratory example: an isolated metal-coated sphere is touched by a charged plastic rod and some of the excess electrons flow off the rod onto the sphere. When the rod is removed, the excess negative charge will remain on the sphere. If the rod was positively charged, electrons would flow in the opposite direction.

In effect, the charge has been *shared*, but the amount of charge that flows off the original object (a plastic rod in this example) depends on many factors (sizes, shapes, ability to conduct and so on).

If a charged object is touched by a non-insulator which is connected to the ground, electrons will be easily attracted onto, or off, the object, so that it quickly loses its overall charge, although the effect on the ground is insignificant. This is called **discharging**. Charged objects usually tend to become discharged easily because charges can flow through the air, especially if the air is humid (high water content), as in wet weather conditions.

When objects come in contact with each other, excess electrons, or a deficit of electrons may be shared between them.

Figure D2.5 Symbol for a ground connection

◆ **Discharge** Flow of electrons to or from an object that reduces the overall charge on it.

◆ **Earth (ground) connection** A good conductor connected between a point on a piece of apparatus and the ground. This may be part of a safety measure, or to ensure that the point is kept at 0 V.

If we wish to make sure that an object is not charged, a good conducting path (of low resistance) is made with the earth / ground. This is called **earthing** (or **grounding**). In domestic wiring this is done by connecting a thick copper wire to a metal plate in the ground, or by connecting to a metal water pipe. In this way, the frames of metal devices can be kept safe at 0 V – the same as the Earth. In some experiments it may be necessary to keep one point in a circuit at 0 V and this is also done with a connection to earth (ground). The symbol for ground connection is shown in Figure D2.5.

Charged objects will become discharged if they are connected to the ground. If this is done deliberately, it will happen very quickly and is called grounding. The object will then be at 0 V.

Charging by electrostatic induction

If a charged object is brought close to an uncharged object, but *without touching*, forces will be exerted on electrons in the uncharged object. Some electrons in the surface near the charged object will be either repelled or attracted, and this results in some charge separation. See Figure D2.6, in which a positively charged rod is attracting some electrons towards it, leaving the other side of the sphere positively charged.

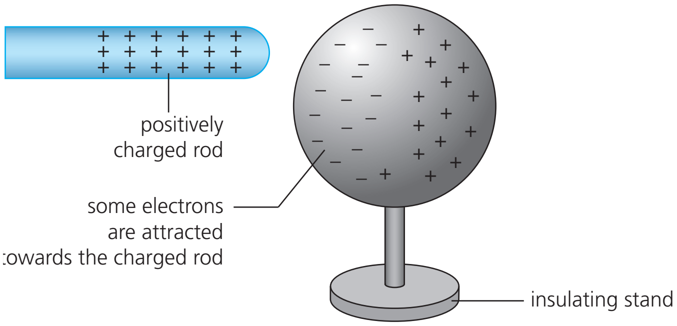


Figure D2.6 Charge separation induced without contact

◆ **Electrostatic induction** Movement of charged particles (electrons) caused by the influence of a nearby charged object, but without physical contact.

If the rod was negatively charged, some electrons would be repelled to the far side of the sphere, leaving the side closest to the rod with an excess of positive charge.

Charge separation caused, without contact, by a nearby charge is known as **electrostatic induction**.

(In this sense of the word, *induction* is being used to describe something being made to happen without physical contact)

When the charged rod is removed the electrons will redistribute evenly.

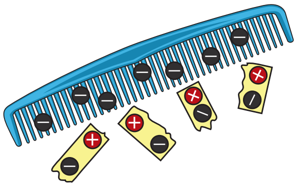


Figure D2.7 An example of electrostatic induction

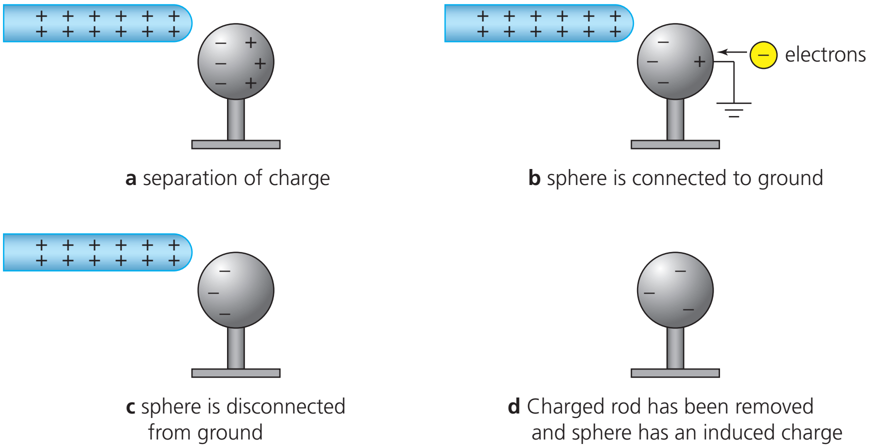
Electrostatic induction is needed to explain most of the electrostatic effects we may see in everyday life. For example, Figure D2.7 shows how a comb (which has been previously charged by friction) can attract uncharged small pieces of paper (without contact).

The excess electrons on the charged comb repel some electrons on the pieces of paper. The comb then attracts the paper because the paper nearest the comb now has a positive charge.

Electrostatic induction can be the best way to charge an object for an experiment, because it does not involve sharing charge. Figure D2.8 shows how. When the sphere is grounded, electrons flow onto the sphere and they will remain there when the connection is removed. Using a negatively charged rod can result in a positively charged sphere.

Common mistake

Do not confuse *electrostatic* induction (described here) with *electromagnetic* induction (Topic D.4).



a separation of charge

b sphere is connected to ground

c sphere is disconnected from ground

d Charged rod has been removed and sphere has an induced charge

Figure D2.8 Charging by induction

Dangers of static electricity

Large-scale electrostatic effects can be unwanted and even dangerous. Lightning is an obvious example (see the activity later in this topic). Cars and planes can become charged as they move through the air or along the ground, and this could be a problem when they stop for refuelling – any sparks from a charged vehicle might cause an explosion of the fuel and air. This risk can be prevented by making sure that the vehicle and the fuel supply are well grounded (see question 5).

1 Describe the movement of electrons that causes a plastic rod to become positively charged when it is rubbed with a dry cloth.

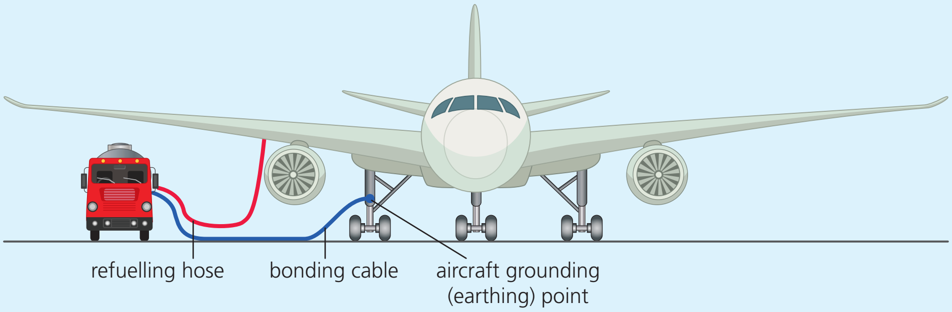
2 Describe and explain an electrostatic effect that you have seen in your home.

3 Two conducting spheres have charges of −20.0 μC and +6.0 μC. If the spheres are identical and come briefly into contact, determine the charge on each sphere when they are separated.

4 If you were given two conducting spheres on insulating stands (similar to those seen in the previous figures), explain how you could make one positively charged and the other negatively charged by using a negatively charged plastic rod.

5 Explain:

- **a** how an electrostatic effect could be dangerous when refuelling an aircraft (Figure D2.9)
- **b** how grounding can prevent the problem arising
- **c** how sparks could occur at a petrol (gas) station. Use the internet to learn about how dangerous this may be.



■ **Figure D2.9** Refuelling an aircraft

Electric forces: Coulomb's law

SYLLABUS CONTENT

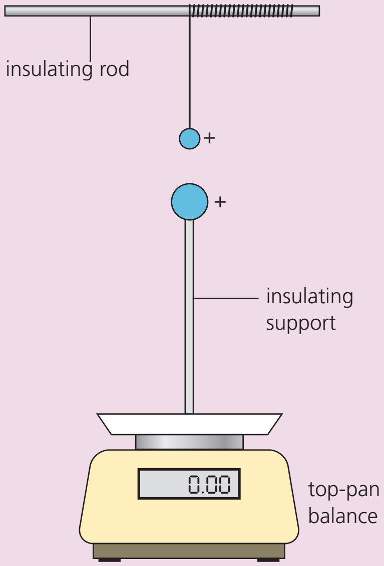
- ▶ Coulomb's law as given by: $F = k\frac{q_1q_2}{r^2}$ for charged bodies treated as point charges where $k = \frac{1}{4\pi\varepsilon_0}$

Inquiry 2: Collecting and processing data

Collecting data

Identify issues that might arise when attempting to collect accurate data

Figure D2.10 shows the apparatus that a student plans to use to determine how the force between charges depends on the magnitudes of the charges and their separation. Identify the issues that might arise when the student attempts to collect accurate data.



■ **Figure D2.10** Possible investigation of electric forces

TOK

The natural sciences

- Why are many of the laws in the natural sciences stated using the language of mathematics?

Another useful analogy

Electric forces and fields can be unfamiliar and difficult to visualize. Gravitational fields (Topic D.1) are generally considered easier to understand because they are more familiar. Fortunately, there is a close mathematical *analogy* between these two types of field. A thorough understanding of gravitational fields will be of great help in studying this topic.

The forces between point charges can be represented by an inverse square law: $F \propto 1/r^2$, where r is the distance between the charges (q_1 and q_2). See Figure D2.11, which shows the forces between similar charges (both positive, or both negative). If the charges were opposite (one positive, the other negative) the forces would be attractive.

◆ **Coulomb's law** There is an electric force between two point charges, q_1 and q_2 given by $F = k\frac{q_1q_2}{r^2}$, where r is the distance between them and k is the Coulomb constant.

◆ **Coulomb constant, k** The constant that occurs in the Coulomb's equation. $k = 8.99 \times 10^9\,\text{N m}^2\,\text{C}^{-2}$. $k = \frac{1}{4\pi\varepsilon_0}$, where ε_0 is the electrical permittivity of free space.

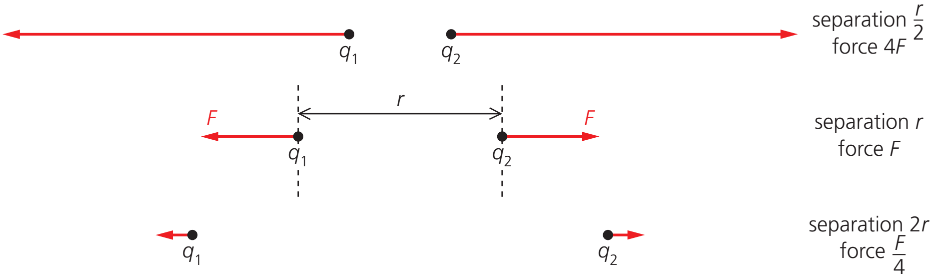


Figure D2.11 The repulsive force varies with distance between similar charges

Coulomb's law represents the relationship between the forces, F, between two point charges and their separation, r. It has a similar form to Newton's law of gravitation:

The forces between two point charges (q_1 and q_2) separated by a distance r:

$$F = k\frac{q_1q_2}{r^2}$$

Figure D2.12 French physicist Charles Augustin de Coulomb (1736-1806)

Charged bodies which are spherical can be considered to act as point charges (at the centres of the spheres), so that they also obey Coulomb's law.

The law was first published by Charles Augustin de Coulomb (Figure D2.12) in 1783.

If the two charges are of the same type (positive and positive, or negative and negative), the forces will have positive signs, representing repulsive forces. If the two charges are opposite (positive and negative), the forces will have negative signs, meaning that the forces are attractive.

The constant k is known as the **Coulomb constant**. It has the value $\mathbf{8.99 \times 10^9\,N\,m^2\,C^{-2}}$.

k is not a fundamental constant (unlike G), because it can be further simplified:

$$\text{Coulomb constant, } k = \frac{1}{4\pi\varepsilon_0}$$

Permittivity

The $1/4\pi$ in the expression for k represents the radial nature of the force and ε_0 represents the electric properties of free space (vacuum).

ε_0 is called the electrical **permittivity of free space** and it has a value of **$8.85 \times 10^{-12}\,C^2\,N^{-1}\,m^{-2}$**.

The electrical permittivity of free space, ε_0, is a fundamental constant which represents the ability of free space to transfer an electric force and field.

◆ **Permittivity of free space, ε_0** Fundamental constant that represents the ability of a vacuum to transfer an electric force and field, $\varepsilon_0 = 8.85 \times 10^{-12}\,C^2\,N^{-1}\,m^{-2}$.

◆ **Permittivity (electric) of a medium, ε** Constant that represents the ability of a particular medium to transfer an electric force and field. Often expressed as **relative permittivity**: $\varepsilon_r = \frac{\varepsilon}{\varepsilon_0}$ (no units), which is also sometimes called **dielectric constant**.

The permittivities of other substances are all greater than ε_0, although dry air has similar electrical properties to free space. This means that the force between two charges in air would be reduced if the air was replaced by another medium.

The **permittivity of a particular medium, ε,** is divided by the permittivity of free space to give the **relative permittivity, ε_r,** of the medium. Some examples are shown in Table D2.1. (Relative permittivity is sometimes known as the **dielectric constant** of the medium.)

$$\text{relative permittivity} = \frac{\text{permittivity of medium}}{\text{permittivity of free space}}$$

$$\varepsilon_r = \frac{\varepsilon}{\varepsilon_0}$$

Table D2.1 The approximate relative permittivities of some common insulators.

Free space (a vacuum)	1 (by definition)
dry air	1.0005
polythene	2
paper	4
concrete	4
rubber	6
water	80

Because it is a ratio, relative permittivity does not have a unit. For example, if the permittivity of a certain kind of rubber was $4.83 \times 10^{-11}\,C^2\,N^{-1}\,m^{-2}$, its relative permittivity would be:

$$\frac{4.83 \times 10^{-11}}{8.85 \times 10^{-12}} = 5.46$$

WORKED EXAMPLE D2.1

A point charge of $4.5 \times 10^{-8}\,C$ is situated in air 3.2 cm from another charge of $-1.3 \times 10^{-7}\,C$.

a Determine the electrical force between them.

b If they were separated by polythene, calculate the approximate force between the charges.

Answer

a $F = k\frac{q_1 q_2}{r^2} = \frac{(8.99 \times 10^9) \times (4.5 \times 10^{-8}) \times (-1.3 \times 10^{-7})}{(3.2 \times 10^{-2})^2} = -5.1 \times 10^{-2}\,N$

The negative sign represents an attractive force.

b Polythene has a relative permittivity of about 2, so the force would be divided by 2 (approximately), $F \approx -3 \times 10^{-2}\,N$.

6 Given that the electrical permittivity of free space is $8.85 \times 10^{-12}\,C^2\,N^{-1}\,m^{-2}$, show that the Coulomb constant has a value of $8.99 \times 10^9\,N\,m^2\,C^{-2}$.

7 The force between two identical point charges was 5.0 N when they were separated by 7.6 cm.
What were the magnitudes of the charges?

8 The surfaces of two insulated conducting spheres were separated by 14.0 cm. One sphere had a radius of 2.7 cm and had a charge of $3.6 \times 10^{-7}\,C$. The other had a radius of 3.9 cm and had a charge of $-4.8 \times 10^{-7}\,C$.

a Determine the force between the spheres in magnitude and direction.

b State any assumption you made when answering **a**.

9 Calculate the force between two point charges of 7.4 C and 2.2 C which are separated by 1.2 m in a non-conducting liquid of relative permittivity 3.1.

10 The force between two point charges was 2.7×10^{-6} N when they were separated by 29 cm.
Predict the force between the same two charges if the separation was increased to 40 cm.

11 Calculate the value of the force between a proton and an electron in a hydrogen atom (separation = 5.3×10^{-11} m).

Nature of science: Patterns and trends

Electric and gravitational forces compared

Two or more charged particles experience *both* electric and gravitational forces between them and it informative to compare the magnitudes of these forces. The forces between an electron and a proton gives perhaps the obvious example.

The electric force between a proton and an electron is about 10^{39} × greater than the gravitational force. This is truly an unimaginably large number! Immediately, we can see that gravitational forces are totally insignificant when discussing atomic particles.

Both types of force follow a similar inverse square law, so why are electrical forces apparently insignificant on the very large scale? For example, gravitation seems to dominate an understanding of the formation and motions of planets and stars. This is because gravitational forces are only attractive and increase with the size of the masses involved, but electric forces are both attractive and repulsive. On the small scale, separate charges result in significant electric forces, but on the very large scale the enormous numbers of positive and negative charges are usually approximately balanced, so that electrostatic effects are insignificant.

LINKING QUESTION

- What are the relative strengths of the four **fundamental forces**?

This question links to understandings in Topics D.1, D2, E.1 and E.3.

◆ **Fundamental forces (interactions)** Strong nuclear, electromagnetic and gravitational forces (and the weak nuclear force) are the four fundamental forces.

◆ **Radial field** Field (electric or gravitational) that spreads out from a point equally in all directions.

Electric fields

SYLLABUS CONTENT

- Electric field lines.
- The relationship between field line density and field strength.
- The electric field strength as given by: $E = \frac{F}{q}$.
- The electric field strength between parallel plates as given by: $E = \frac{V}{d}$.

A region in which a charge would experience an electric force is called an *electric field*.

Electric fields are represented on paper, or on a screen, with electric field lines. The direction of electric forces depends on the nature of the charges but, by convention we choose that:

electric field lines always point in the directions of the forces on positive charges.

Figure D2.13 shows the two most basic *electric fields*: **radial fields** around an isolated point positive charge and around an isolated point negative charge

The field lines can never cross each other and the field is strongest (in a particular diagram) where the field lines are closest together (densest).

Figure D2.14 shows the combined electric field of two equal point charges. If the charges were both positive the field lines would point in the opposite directions. Figure D2.15 shows the field around opposite charges of equal magnitude.

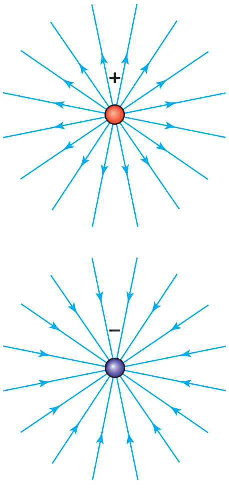


Figure D2.13 Radial electric fields around point charges

Figure D2.16 shows the electric field around a charged solid conducting sphere (in this example the excess charge is negative). The mobile charges (electrons) repel each other, so that they are evenly distributed on the outer surface of the sphere. The resulting electric field is perpendicular to the surface, but there is no field inside the sphere. The field is identical to that around a point charge at the centre of the sphere which had the same charge as all the excess electrons combined. The same is true for a hollow conducting sphere.

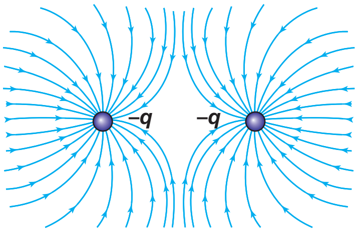


Figure D2.14 Field around two similar point charges

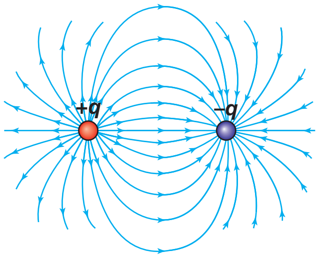


Figure D2.15 Field around two opposite point charges

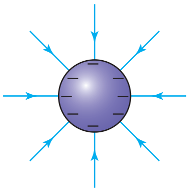


Figure D2.16 Electric field around a charged solid sphere

Electric field lines must be perpendicular to any conducting surface.

If this was not true, there would be a component of the electric force acting on the electrons, so that they would rearrange until the field was perpendicular.

If a surface is flat, the forces between adjacent mobile charges will be parallel to the surface and this results in an even distribution of charges, as seen in Figure D2.17a. But if the surface has a variable curvature, the forces between adjacent charges will not be parallel to the surface and will vary with the amount of curvature. (Figure D2.17b). This results in charge becoming concentrated where the curvature changes most suddenly, that is, near points. See Figure D2.18.

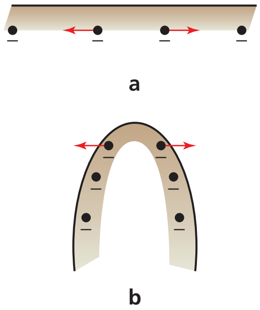


Figure D2.17 Forces between charges near the surface of a conductor

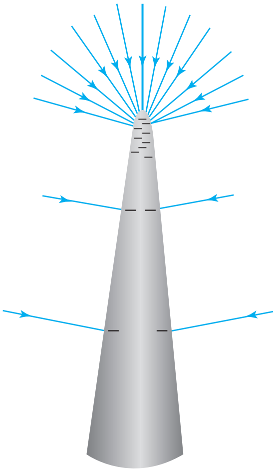


Figure D2.18 Concentration of charge and electric field near points

Figure D2.19 shows an important arrangement: the *uniform electric field* that can be created between parallel metal plates. The positive terminal on the battery attracts electrons, so that the top plate becomes positively charged. The lower plate becomes negatively charged because electrons have been repelled away from the negative terminal of the battery. The field is weaker and non-uniform beyond the edges of the plates. At the points midway between the edges of the two plates, the strength of the field is half of its maximum value.

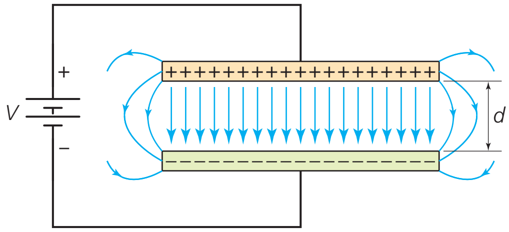


Figure D2.19 Creating a uniform electric field

The arrangement seen in Figure D2.19 can be used with high voltages if strong, uniform fields are needed in experiments.

12 By considering components of forces, discuss why field lines must always be perpendicular to conducting surfaces.

13 Suggest why pointed conductors (see Figure D2.20) are a good way of discharging static electricity.

Figure D2.20 Static dischargers on the wing of an aircraft

14 Consider Figure D2.19. Draw electric field lines to represent the electric field if the battery was reversed and the p.d. halved.

15 An uncharged metal sphere is between two charged metal plates as shown in Figure D2.21. Copy the figure, show where charges are induced and add lines to represent the electric field.

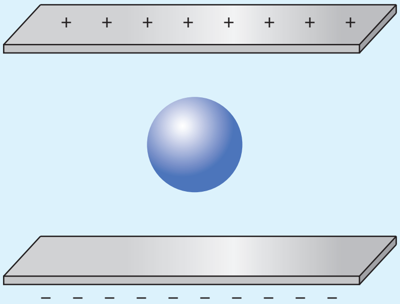


Figure D2.21 An uncharged metal sphere between two charged metal plates

Electric field strength

Electric field strength is defined in a similar way to gravitational field strength:

◆ **Electric field strength, *E*** The electric force per unit charge.

Electric field strength, *E*, is defined as the force per unit charge that would be experienced by a small positive test charge placed at that point:

$$E = \frac{F}{q}$$

SI unit: $\mathrm{N\,C^{-1}}$

TOK

The natural sciences

- What is the role of imagination and intuition in the creation of hypotheses in the natural sciences?
- Are gravitational and electric fields real?

The effects of gravitational and electric forces are easily observable, but we have defined these fields in term of such forces, so do the fields really exist if there are no forces acting? For example, is there really an electric field around a stationary charge if there is no other charge present? Or, is the concept of ‘field’ just an imaginary device constructed to help us understand that forces can act without contact?

WORKED EXAMPLE D2.2

If there is a force of 3.4×10^{-6} N acting on a point charge of 6.7 nC, calculate the magnitude of the electric field strength at that location. ($1\text{ nC} = 1 \times 10^{-9}$ C)

Answer

$$E = \frac{F}{q} = \frac{3.4 \times 10^{-6}}{6.7 \times 10^{-9}} = 5.1 \times 10^{2}\,\mathrm{N\,C^{-1}}$$

There is not enough information in this question to know the direction of the field.

An expression for the strength of the constant electric field between parallel plates (Figure D2.19) can be obtained by considering the work done as a charge, q, moves from one plate to the other:

work done = force × distance = potential difference × charge (from Topics A.3 and B.5)

$W = Fd = Vq$

Rearranging, and remembering that $E = \frac{F}{q}$, gives:

DB

electric field strength between parallel metal plates, $E = \frac{V}{d}$

Expressed in this way we can see that $\mathrm{V\,m^{-1}}$ is an alternative to $\mathrm{N\,C^{-1}}$ as the units for electric field strength.

(As we shall see later, this is an example of electric field strength equalling the potential gradient.)

Tool 3: Mathematics

Using units, symbols and numerical values

Consider again the situation shown in Figure D2.19. The space between the plates has a vacuum, and a charge, q, is next to the negative plate. The charge will accelerate towards the positive plate and gain kinetic energy. In this process, the work done on the charge is $Fd = Vq$, as explained above.

Consider a numerical example:

An electron (mass 9.110×10^{-31} kg and charge -1.60×10^{-19} C) is accelerated across a distance of 5.0 cm by a potential difference of 3000 V.

Assuming that the electron starts with zero kinetic energy, its final kinetic energy equals the work done on it by the uniform electric field, $W = 3000 \times (1.60 \times 10^{-19}) = 4.80 \times 10^{-16}$ J. (If we want to determine the speed of the electron, we can equate this to $\frac{1}{2}mv^2$, which gives an answer of $3.25 \times 10^{7}\,\mathrm{m\,s^{-1}}$.)

In situations similar to this, rather than using joules as the unit of energy, it is more common and easier to use the **electronvolt**.

One electronvolt is an amount of energy equal to 1.60×10^{-19} J. This the amount of energy gained by a charge of 1.60×10^{-19} C when accelerated by a potential difference of 1.00 V. (work done, $W = qV$).

In the previous example, the work done on the electron by the electric field (= kinetic energy it gains) can be stated as 3000 eV, without the immediate need for any further calculations.

Common mistake

Although it is called an *electron*volt, this unit is widely used for the atomic-scale energies of any particles, or radiation. Some examples: a proton accelerated by 5 kV will gain 5 keV of energy; A doubly charged ion accelerated by 2000 V will gain 4 keV of energy; a gamma ray (Topic E.1) may transfer 5 MeV of energy.

◆ Electronvolt, eV
An amount of energy equivalent to that which is transferred when an electron is accelerated by a potential difference of 1 V. $1\,\mathrm{eV} = 1.60 \times 10^{-19}$ J.

WORKED EXAMPLE D2.3

Parallel metal plates are separated by a distance of 0.50 cm. Determine the potential difference needed to create an electric field of one million $\mathrm{V\,m^{-1}}$.

Answer

$$E = \frac{V}{d}$$

$$1.0 \times 10^{6} = \frac{V}{0.0050}$$

$$V = 5.0 \times 10^{3}\ \mathrm{V}$$

16 Calculate the value of electric field strength that would exert a force of 6.3×10^{-14} N on a proton.

17 a Calculate the electric field strength along a straight wire of length 38 cm if there is a potential difference of 0.0010 V between its ends.

b What force would this field exert on a free electron in the wire?

c Determine the acceleration of the electron.

d Explain why the electron is not accelerated to an extreme speed.

18 a Calculate the electric field strength between parallel metal plates separated by 8.0 cm when a p.d. of 15 kV is connected across them.

b How much energy is gained by an electron accelerated between the plates in:

i eV **ii** joules?

19 If the electric field strength exceeds about 3 kV mm^{-1}, air can begin to conduct electricity. Predict the potential difference needed across 20 cm for this to happen.

Electric field strength around a point charge

The strength of an electric field around a point charge decreases with the square of the distance (an inverse square law relationship):

$$E = \frac{F}{q} = \frac{kq}{r^2}$$

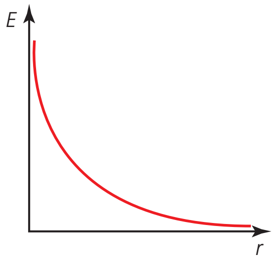


Figure D2.22 Electric field strength around a positive charge

WORKED EXAMPLE D2.4

Determine the electric field strength at a distance of 1.0 m from a charge of $+2.9 \times 10^{-8}$ C.

Answer

$$E = \frac{kq}{r^2} = \frac{(8.99 \times 10^9) \times (+2.9 \times 10^{-8})}{1.0^2}$$

$$= 260\,\text{N}\,\text{C}^{-1}$$

Figure D2.22 shows the variation of electric field strength around a point positive charge. The field strength would be negative if the charge was negative.

Combining electric fields

Electric fields can be combined to determine a resultant by using vector addition.

This is straightforward for places on the line that joins charges, for example: if a field of 420 N C^{-1} to the left and a field of 550 N C^{-1} to the right act at a point; the combined field is 130 N C^{-1} to the right.

More generally, we can use a parallelogram and scale drawing (or trigonometry) as seen in Figure D2.23.

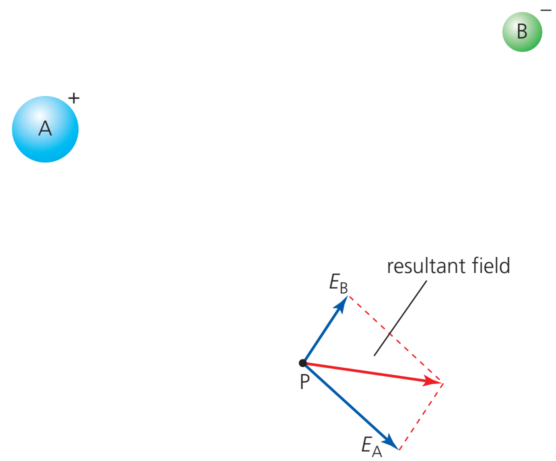


Figure D2.23 Determining a resultant electric field

WORKED EXAMPLE D2.5

Figure D2.24 shows two separate electric fields acting at a point, P. Determine the resultant field.

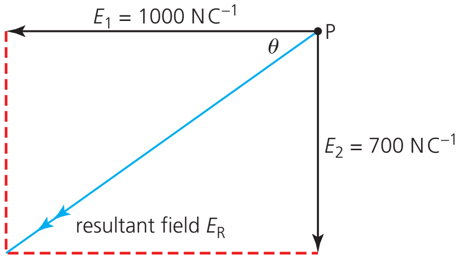


Figure D2.24 Two separate electric fields acting at a point P

Answer

Using Pythagoras, $E_R^{\ 2} = E_1^{\ 2} + E_2^{\ 2} = 1000^2 + 700^2$

$E_R = 1220\,\text{N}\,\text{C}^{-1}$

Electric field is a vector quantity, so the answer must include a direction:

$$\tan\theta \text{ (as shown)} = \frac{700}{1000} = 0.700$$

$\theta = 35°$

20 Sketch a graph to represent the variation of electric field strength with distance around a negatively charged conducting sphere.

21 Calculate the electric field strength 15 cm from a charge of −8.4 μC when it is in:

a air

b a material of relative permittivity 1.6.

22 The electric field strength 150 cm from a point charge is $2.0 \times 10^5\,\text{N}\,\text{C}^{-1}$.
Determine at what distance the field strength would be one million N C^{-1}.

23 A nucleus of a carbon atom has a charge of +6*e*.
Determine the distance from its centre where the electric field strength has a value of $3.0 \times 10^{10}\,\text{N}\,\text{C}^{-1}$.

24 a Calculate the electric field strength midway between point charges of 26 nC and −10 nC when they are separated by a distance of 50 cm.

b Determine the field strength at a point which is 35 cm from the negative charge and 15 cm from the positive charge.

25 A point charge *Q* is 8.3 cm from a charge of +56 μC. The electric field strength is zero at a point which is 4.7 cm from *Q* on a line joining the two charges. Determine the charge of *Q*.

26 The average electric field strength just above the surface of the Earth is about 150 N C^{-1}, directed downwards.
Estimate the total resultant charge carried by the Earth. (The radius of Earth is 6.4×10^6 m.)

27 Sketch the approximate shape of the electric field lines around two charged spheres of equal radius, *R*, and separated by 4*R*: one with a charge of +*Q*, the other with a charge of −4*Q*.

Millikan's experiment

SYLLABUS CONTENT

- Millikan's experiment as evidence for quantization of electric charge.

A strong, uniform electric field was an essential component of the famous 1909 experiment to determine the charge of an electron. See Figure D2.25.

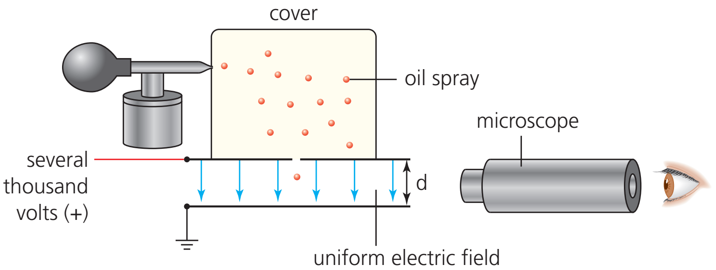


Figure D2.25 Millikan's oil drop experiment

Robert Millikan and Harvey Fletcher used this apparatus to determine the small quantities of charge on oil drops. Details are provided below. The drops can be electrostatically charged because of friction when they are sprayed into the upper compartment. (Additionally, the charge on the drops can be changed using X-rays or a radioactive source, but the details are not needed here.)

The importance of this famous experiment lies in the fact that

the charges of *all* drops were multiples of the same number (-1.60×10^{-19} C).

So, for example, the following oil drop charges ($\times 10^{-19}$ C) could have been determined: −9.62, −11.20, −4.84, −17.58, −6.43, −8.00. Allowing for experimental uncertainty, statistical analysis informs us that it is highly probable that all these numbers are multiples of −1.60.

A simple analogy may help. Suppose you were given six sealed bags each containing an unknown number of the same coins. The masses of the coins (excluding the bags) were 84 g, 63 g, 98 g, 35 g, 56 g and 42 g. What was the probable mass of one coin? (Answer: 7 g)

When enough measurements reach the same conclusion, it effectively becomes a certainty.

Electric charge is not a continuous quantity. It can only have certain discrete values (multiples of e). We say that it is quantized.

Millikan's experiment confirms that electric charge is a quantized quantity.

Experimental details – understanding Millikan's experiment

If the potential difference, V, between the plates is varied, then the electric force, F, on a charged drop changes. With the correct potential difference, it is possible for the electrical force and gravitational force to become equal and opposite. The resultant force will then be zero and the drop will be stationary. See Figure D2.26.

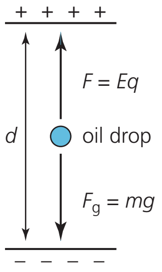


Figure D2.26 Balanced forces on a stationary oil drop

$$F = mg = Eq = \frac{Vq}{d}$$

If V, d and m are known, q can be calculated.

WORKED EXAMPLE D2.6

What potential difference across parallel metal plates separated by 2.1 cm is necessary to keep an oil drop of mass 2.7×10^{-14} kg stationary if it has five excess electrons?

Answer

$$mg = \frac{Vq}{d}$$

$$2.7 \times 10^{-14} \times 9.8 = \frac{(V \times 5 \times (1.60 \times 10^{-19}))}{2.1 \times 10^{-2}}$$

$$V = 6.9 \times 10^{3}\,\mathrm{V}$$

28 An oil droplet has a weight of 7.68×10^{-15} N.

- **a** If it is stationary between two horizontal metal plates which are 1.0 cm apart with a voltage of 120 V across them, determine the charge on the oil droplet.
- **b** How many excess electrons are on the droplet?

29 Show, with an approximate calculation, why it may be reasonable to ignore the buoyancy force in the previous calculation.

30 Explain why it is reasonable to assume that the masses of the coin bags described above (84 g, 63 g, 98 g, 35 g, 56 g and 42 g) lead to the conclusion that the mass of each coin is 7 g.

Note that there is also a buoyancy force (see Topic A.2) acting upward on the oil drop. This force is much less than the weight of the oil drop and has been ignored for the sake of simplicity. For very accurate work it would need to be included in the calculation.

The mass of a spherical oil drop can be determined from its dimensions and the density of the oil. More accurately (the drop will not be perfectly spherical), the terminal speed of a drop can be used to determine its mass using Stokes's law (Topic A.2).

Nature of science: Models

Lightning

Figure D2.27 Lightning over Kuala Lumpur

Water and ice droplets of different sizes in clouds are variously affected by convection currents. As they move past each other, friction causes the transfer of electrons (as described earlier in this topic). This typically results in the upper and lower surfaces of the cloud becoming charged differently and charge separation induced on the ground. The result can be a strong electric field between the cloud and the ground.

This can be a complex situation. Scientific explanations often require that reasonable assumptions are made so that a simplified model can be suggested that allows us to understand complicated natural phenomena. In the case of lightning, we could use a simplified model such as that shown in Figure D2.28.

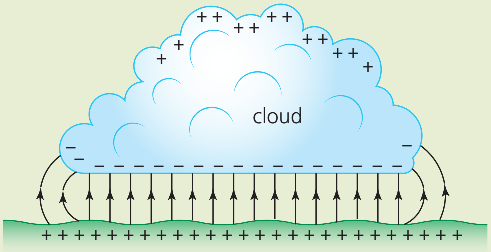


Figure D2.28 Electric field under a storm cloud

When the potential difference between the ground become large enough the (wet) air can dramatically conduct electricity: a lightning strike.

Some approximate numerical values:

p.d., $V = 2.0 \times 10^{6}\,\mathrm{V}$

separation of ground and cloud, $d = 0.5\,\mathrm{km}$

$$E = \frac{V}{d} = 4000\,\mathrm{V\,m^{-1}}$$

Charge transferred in lightning strike of $\Delta t = 0.2\,\mathrm{s}$

$\Delta q = 3000\,\mathrm{C}$

so that current:

$$I = \frac{\Delta q}{\Delta t} = 15\,000 A$$

Power $= IV = 2.0 \times 10^{6} \times 15\,000 = 3.0 \times 10^{10}\,\mathrm{W}$

Total energy transferred $= P\Delta t = 3.0 \times 10^{10} \times 0.2 = 6.0 \times 10^{9}\,\mathrm{J}$

Magnetic fields around permanent magnets

SYLLABUS CONTENT

- Magnetic field lines.

◆ **Magnetic forces** Fundamental forces that act across space between all moving charges, currents and/or permanent magnets.

◆ **Permanent magnet** Magnetized material that creates a significant and persistent magnetic field around itself.

◆ **Ferromagnetic materials** Materials from which permanent magnets are made.

◆ **Magnetic poles (north and south)** Regions in a magnetic material where the field is strongest

Magnetic forces can act across space in a similar way to gravitational and electrical forces, but the equations for magnetic forces and fields are different in form from the other two. Magnetic effects are very closely connected to electrical effects.

A magnetic field exists anywhere that a magnetic force occurs.

Magnetic fields are produced around all moving charges (currents).

Any 'stationary' charge has an electric field around it but, if a charge moves, there is also a magnetic field around it, which is perpendicular to the electric field. We frequently refer to the combinations as *electromagnetic* fields.

Before considering the important subject of the production of temporary magnetic fields when currents flow in circuits, we will consider the **permanent magnets** with which we are all familiar, similar to that seen in Figure D2.29.

Figure D2.29 Magnet on a refrigerator door

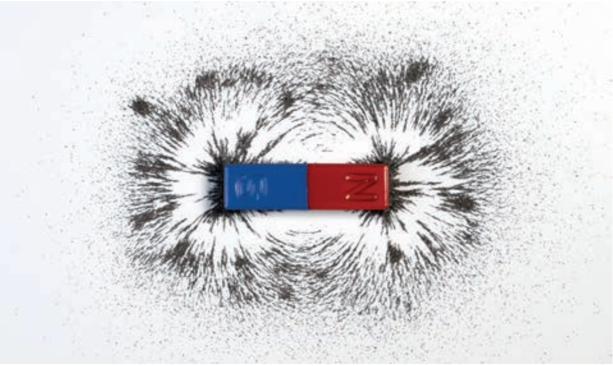


Figure D2.30 Iron filings show the shape of the magnetic field around a bar magnet

The motion of electrons within atoms creates tiny magnetic fields, but in most elements these effects cancel each other, so that they have no significant magnetic properties. Iron is a notable exception because it can be magnetized. Iron alloys, nickel, cobalt and some rare Earth metals can also be magnetized. They are known as **ferromagnetic materials**. After a ferromagnetic material has been magnetized, it may lose its magnetism quickly, it is then described as being 'soft' (magnetically). Pure iron is an important example. However, many other ferromagnetic materials, steel for example, are magnetically 'hard' and can retain their magnetism for a long time.

The simplest permanent magnets are made in the shape of a straight bar. See Figure D2.30. The shape of the magnetic field around the bar magnet has been shown by sprinkling tiny pieces of iron (iron filings) around the magnet. This is explained later.

The magnetism is effectively concentrated at the ends of the bar and these are called **magnetic poles**. We have seen that there is only one type of mass, but two types of charge (positive and negative) which can be isolated from each other. Magnetism is different: there are only two types of magnetic pole, but they *always* occur in pairs. Magnetic poles are called magnetic north poles and magnetic south poles. Confusingly, these names have no direct link with geography. An explanation is included later.

One end of the bar magnet seen in Figure D2.30 is a magnetic north pole (N), the other is a magnetic south pole (S). This simple arrangement is called a magnetic **dipole**.

◆ **Dipole** Two close electric charges (or magnetic poles) of equal magnitude but of opposite sign (or polarity).

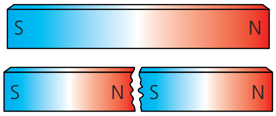


Figure D2.31 Cutting a bar magnet in half

If the magnet was cut in half, we would *not* have separate poles. The result would be two smaller, weaker magnets, but they would still have a magnetic north pole at one end and a magnetic south pole at the other end. See Figure D2.31.

Magnetic field lines

Magnetic field lines are used to represent magnetic fields on paper or screen. As with gravitational and electric fields, we give a direction to magnetic field lines.

Magnetic field lines always form closed loops. The direction of field lines (around a magnet) is from a magnetic north pole to a magnetic south pole.

Usually we do not show the field lines *inside* magnets but, when we do, their direction is from the south magnetic pole to the north magnetic pole. See Figure D2.32.

As with other field lines, magnetic field lines can never cross each other and the field is strongest where the lines are closest together (in a specific diagram). Figure D2.32 shows clearly that the field outside the magnet is strongest close to the poles.

Magnets are often made into U-shapes, as seen in Figure D2.33. This strengthens the field near to the poles.

A strong *uniform* magnetic field is often needed for experiments. Figure D2.34 show how this can be achieved with parallel permanent magnets.

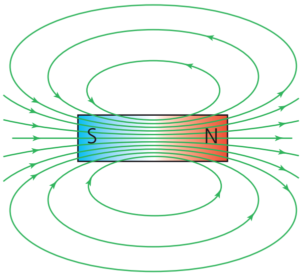


Figure D2.32 Field lines in and around a bar magnet

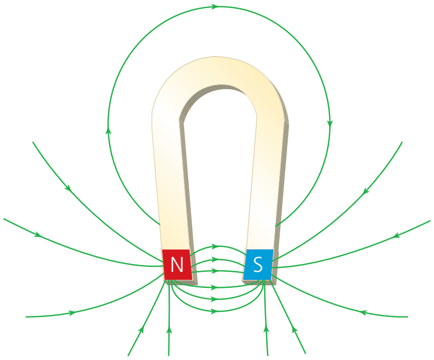


Figure D2.33 The field lines near a U-shaped magnet

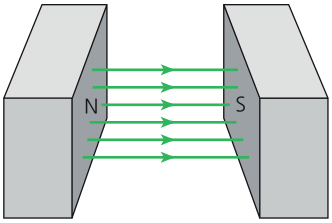


Figure D2.34 Creating a strong, uniform magnetic field

Forces between permanent magnets

Similar magnetic poles repel each other, opposite poles attract.

If two bar magnets are brought close to each other, the forces between them will cause them to align (if at least one of them is free to move), as shown in Figure D2.35.

Many **compasses** use this effect.

◆ **Compass** A device for determining direction. Small plotting compasses are used to investigate the shapes of magnetic fields in the laboratory.

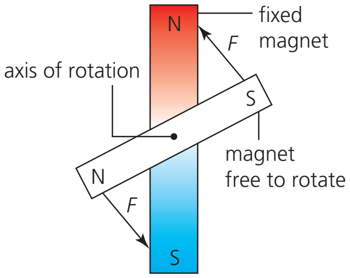


Figure D2.35 Magnetic forces forming a couple (see Topic A.4)

◆ **Induced magnetism** When a ferromagnetic material becomes magnetized because it is in an external magnetic field.

Induced magnetism

When ferromagnetic materials are located in magnetic fields, they tend to become magnetized to some extent.

This is called **induced magnetism**. These effects may reduce, or disappear, if the material is removed from the field.

This effect explains why, for example, a permanent magnet can attract unmagnetized nails, as shown in Figure D2.36. Each nail becomes magnetized and can then induce magnetism in other nearby nails.

In a similar way, all the tiny iron filings seen in Figure D2.30 each get magnetically induced and then line up with the field of the magnet.

Figure D2.36 Example of induced magnetism

The Earth's magnetic field

The Earth behaves as a very large, weak bar magnet with a magnetic south pole near the geographic North Pole. See Figure D2.37.

Many compasses detect the Earth's magnetic field in order to indicate direction. The compass needle is a small permanent magnet which is able to rotate freely. It aligns with the Earth's magnetic field, so that the north pole of the compass magnet points towards the south pole of the Earth's magnetic field, which is close to the geographic North Pole.

The magnetic north pole of a magnet is so called because that end of a freely suspended magnet points towards geographic North (where there is a south magnetic pole).

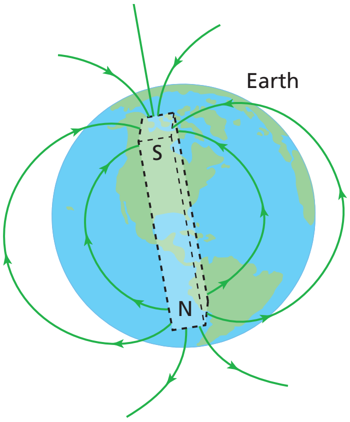


Figure D2.37 The Earth's magnetic field

Detecting magnetic fields

As explained above, a compass effectively detects the Earth's magnetic field, and small 'plotting compasses' are widely used to detect magnetic fields around magnets in school laboratories. They point along the magnetic field lines, effectively from magnetic north to magnetic south. See Figure D2.38.

The plotting compasses are also in the Earth's magnetic field, but the Earth's field is weak in comparison to the field close to the bar magnet. Iron filings are also widely used to show the shape of a magnetic field, as seen in Figure D2.30. In recent years tiny *magnetometers* for measuring magnetic fields have become common. They are to be found in most mobile phones.

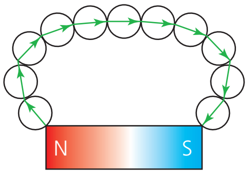


Figure D2.38 Plotting compasses indicating a magnetic field line

Magnetic fields around currents in wires

As was stated at the beginning of this topic, magnetic fields are produced by currents, so the best place to start when understanding magnetism is with the simplest situation.

Magnetic field around a steady current in a long straight wire

The magnetic field lines around a current in a long straight wire are circular.

The increasing separation of the field lines indicates that the field strength decreases with increasing distance from the current (see Figure D2.39). This basic magnetic field can be demonstrated using iron filings or compasses, but the field is not strong unless a large current is used. The direction of the field can be predicted by using the **right-hand grip rule**, as seen in Figure D2.40.

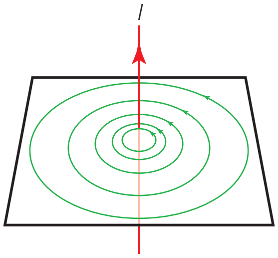


Figure D2.39 Field around a constant current in a long straight wire

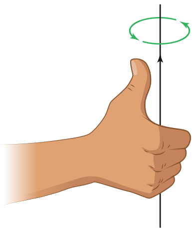


Figure D2.40 Right-hand grip rule

Magnetic fields around steady currents in coils and solenoids

As we would expect, to produce a stronger field requires a greater current in the wire, but there is an easier way: wind insulated wire into a coil or **solenoid**, then each extra turn of wire increases the field strength (with the same current). A solenoid is a coil of insulated wire wrapped regularly so that the turns do not overlap and it is significantly longer than it is wide.

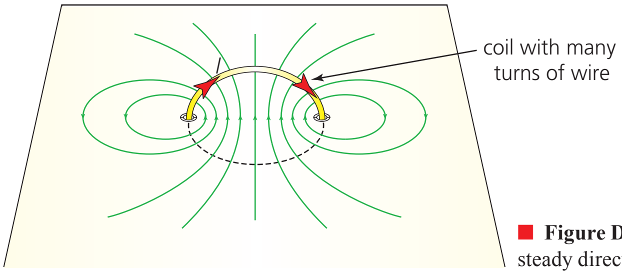


Figure D2.41 Magnetic field around a steady direct current in a circular coil

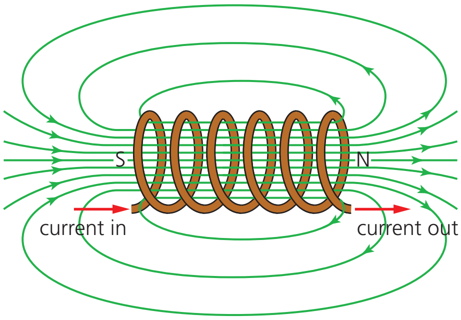


Figure D2.42 Magnetic field due to the current in a solenoid

Figure D2.42 shows the magnetic field in and around a solenoid. (The number of turns shown has been reduced for clarity.) Of especial importance is the strong, uniform field *inside* the solenoid. Comparing Figure D2.42 with Figure D2.32, it is clear that the magnetic field produced by a current in a solenoid is similar in shape to the magnetic field of a bar magnet.

The **polarity** of the magnetic field (which end is a north pole, and which end is a south pole), depends on the direction of the current. It can be predicted using the right-hand grip rule. (Alternatively, when viewed from an end, that end is acting as a south pole if the current is clockwise.)

◆ **Right-hand grip rule** Rule for determining the direction of the magnetic field around a current.

◆ **Solenoid** Long coil of wire with turns that do not overlap (helical).

◆ **Polarity** Separation of opposite electric charges or opposite magnetic poles, which produces uneven effects in a system.

◆ **Electromagnet** Magnet which needs the flow of an electric current in a coil to produce magnetic effects.

◆ **Soft iron** Form of iron (pure or nearly pure) that is easily magnetized and demagnetized.

Electromagnets

Strong **electromagnets** can be made by winding a coil or solenoid on soft iron. A simple example is shown in Figure D2.43.

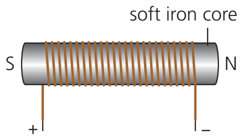


Figure D2.43 Electromagnet

The strength of the magnetic field can be controlled by changing the magnitude of the current and using a soft iron core greatly increases the strength of the field. Importantly, the electromagnet loses its magnetism as soon as the current is turned off. If the core was made of steel, it would retain much of its magnetism when the current was disconnected.

Electromagnets have a very wide range of uses.

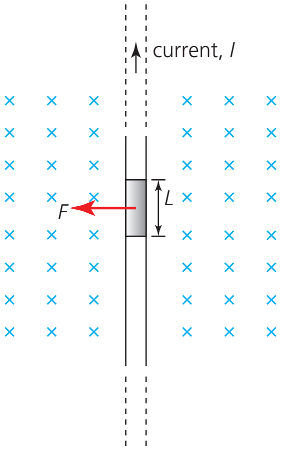


Figure D2.44 Force on current in a magnetic field. (The crosses represent a magnetic field directed perpendicularly into the page.)

Magnetic field strength

We have seen that gravitational field strength $g = \frac{F}{m}$ and electric field strength $E = \frac{F}{q}$ but magnetic field strength is less easily defined. We will describe the strength of a magnetic field in terms of the magnetic force experienced by a current flowing across the field.

We will see in Topic D.3 that moving charges experience magnetic forces when they move across a magnetic field. The simplest example would be the electrons in a current in a straight wire *which is perpendicular* to a magnetic field, as seen in Figure D2.44.

The magnitude of the magnetic force, F, is proportional to three things: the magnitude of the current, I, the length of conductor in the field, L, and **magnetic field strength**, which is given the symbol B.

$F = BIL$ or, rearranging:

magnetic field strength, $B = \frac{F}{IL}$

The SI units of B are newtons per amp metre. 1 N A^{-1} m^{-1} is known as 1 **tesla, T**. 1 T corresponds to a very strong magnetic field, so, milliteslas (mT) and microteslas (μT) are in common use.

◆ **Magnetic field strength, B** Defined in terms of the force on a current: The field which produces a force of 1 N on each 1 m length of a conductor carrying a current of 1 A perpendicularly across the field.

◆ **Tesla, T** Unit of magnetic field strength. 1 T = 1 N A^{-1} m^{-1}

◆ **Permeability (magnetic)** Constant that represents the ability of a particular medium to transfer a magnetic force and field.

◆ **Permeability of free space, μ_0** Fundamental constant that represents the ability of a vacuum to transfer a magnetic force and field, $\mu_0 = 4\pi \times 10^{-7}$ T m A^{-1}.

WORKED EXAMPLE D2.7

Calculate the magnetic field strength that produces a force of 5.0×10^{-4} N on each metre length of a long straight conductor carrying a constant current of 2.0 A. (The field and the current are perpendicular to each other.)

Answer

$$B = \frac{F}{IL} = \frac{5.0 \times 10^{-4}}{1.0 \times 2.0}$$

$$= 2.5 \times 10^{-4}\ \text{T}$$

Magnetic field strength around a current in a long straight wire

We have previously described the shape of the magnetic field around a current in a long, straight wire, now we will consider how we can calculate values for the field strength.

The magnetic field is created by the charges moving in the current and it spreads around the wire. The **magnetic permeability** of a medium, or free space, represents its ability to transfer a magnetic field and force. It may be considered analogous to *electric permittivity*, which describes electric properties. The magnetic properties of air are almost identical to the magnetic properties of free space.

The magnetic **permeability of free space** is given the symbol μ_0 and has the value **$4\pi \times 10^{-7}$ T m A^{-1}**.

The magnetic field strength around a straight current depends on the:

- magnitude of the current, I
- perpendicular distance from the wire, r
- magnetic permeability of the air surrounding the wire (≈ permeability of free space).

Magnetic field strength at a distance r from a current I in a long straight wire in air, $B = \frac{\mu_0 I}{2\pi r}$

WORKED EXAMPLE D2.8

Determine at what distance from a long straight wire carrying a current of 5.0 A, the resulting magnetic field has a strength of 100 μT.

Answer

$$B = \frac{\mu_0 I}{2\pi r}$$

$$100 \times 10^{-6} = \frac{((4\pi \times 10^{-7}) \times 5.0)}{(2\pi r)}$$

$r = 0.01$ m (1 cm)

For comparison, the Earth's magnetic field strength averages about 50 μT, which is comparable to the field within one or two centimetres of the current in this Worked example.

31 a Sketch the magnetic field pattern around two bar magnets as seen in Figure D2.45.

b Sketch the magnetic field pattern around the two bar magnets if the polarity of one of the magnets was reversed.

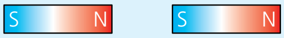


Figure D2.45 Two bar magnets

32 Explain in detail how it is possible for a bar magnet to attract an unmagnetized steel pin.

33 Describe where you would expect the magnetic field of the Earth to be strongest, and in what direction does it act?

34 Consider Figure D2.43.

a Describe how the polarity of the electromagnet can be determined.

b Sketch the magnetic field that this electromagnet would produce.

35 State three applications of electromagnets.

36 The electromagnet seen in Figure D2.43 has been wound on a 'soft' iron core. What difference would it make if the core was made of steel?

37 Draw a graph to represent how the magnetic field strength due to a 2.0 A current in a long straight wire varies with perpendicular distances up to 10 cm from the wire.

38 Show that the SI units of permeability are T m A^{-1}.

HL ONLY

LINKING QUESTION

- How can moving charges in magnetic fields help probe the fundamental nature of matter?

This question links to understandings in Topics D.3, E.1 and E.2.

◆ **Electric potential energy** E_p, is the work done when bringing all the charges of a system to their present positions from infinity.

Electric potential energy

SYLLABUS CONTENT

- ▶ The electric potential energy, E_p, in terms of work done to assemble the system from infinite separation.
- ▶ The electric potential energy for a system of two charged bodies as given by: $E_p = k\frac{q_1 q_2}{r}$.

Electric potential energy is stored in any system of charges because of the forces between them. As with gravitational potential energy, for the same reasons, we chose infinity to be the zero of **electric potential energy**:

The zero of electric potential energy is chosen to be when the charges are separated by an infinite distance.

The total electric potential energy of a system, E_p, is defined as the work done when bringing all the charges of the system to their present positions, assuming that they were originally at infinity.

We saw in Topic D.1 that gravitational potential energy can only ever be negative because the forces are always attractive, meaning that energy has to be supplied to increase their separation.

However, electric potential energies can be negative (if the forces are attractive between opposite charges), or positive (if the forces are repulsive between similar charges). In other words, we need to *supply* energy to separate charges which are attracted to each other, but energy is released (to kinetic energy) as opposite charges are repelled apart from each other.

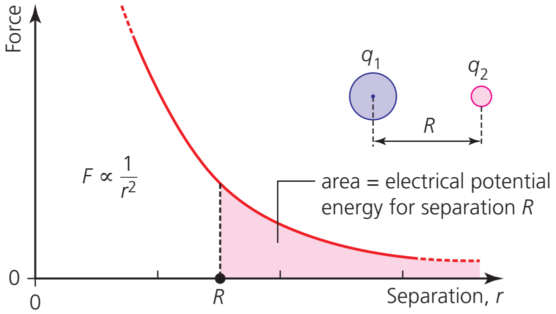


Figure D2.46 Area under graph represents electric potential energy

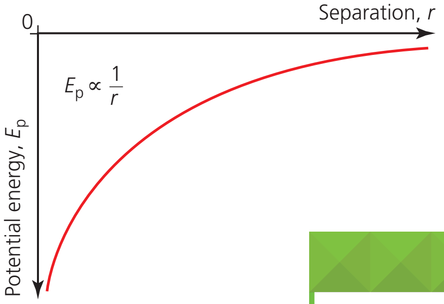


Figure D2.47 Electric potential energy variation with separation between oppositely charged point charges.

Calculating electrical potential energies

Electrical potential energy can be determined from the area under a force–distance graph, as shown in Figure D2.46 for similar charges.

However, for point charges, using the following equation is easier (analogous to the gravitational potential energy equation seen in Topic D.1).

electric potential energy of two point charges, $E_p = k\frac{q_1 q_2}{r}$

This equation can be used for two point charges, q_1 and q_2, separated by a distance r. It can also be used with isolated spherical conductors, when r is then the distance between their centres. The sign of the electric potential energy will be dependent on the signs of q_1 and q_2.

Figure D2.47 represents this inverse relationship graphically, for oppositely charged point charges. If the charges both had the same sign, the potential energy would be positive.

Top tip!

Do not confuse the symbols for electric potential energy, E_p, and electric field strength, E.

WORKED EXAMPLE D2.9

Calculate the electric potential energy that was stored between two isolated spherical conductors: one had a radius of 2.5 cm and charge −4.7 × 10^{-8} C, the other had a radius of 1.5 cm and charge −6.3 × 10^{-8} C. Their surfaces were separated by 1.7 cm.

Answer

Separation of centres, $r = 2.5 + 1.5 + 1.7 = 5.7$ cm

$$E_p = k\frac{q_1 q_2}{r} = \frac{(8.99 \times 10^9) \times (-4.7 \times 10^{-8}) \times (-6.3 \times 10^{-8})}{5.7 \times 10^{-2}}$$

$$= +4.7 \times 10^{-4}\ \text{J}$$

The energy is positive because the charges are repelled from each other and they would gain kinetic energy if they were free to move.

39 There was a force of −4.7 × 10^{-7} N between two point charges when they were separated by 40 cm.

- **a** Draw a force–separation graph to show how the force varies over distances of 10 to 100 cm.
- **b** Use your graph to determine the change in electrical potential energy in the system if the separation was increased from 40 cm to 90 cm.

40 The surfaces of two identically charged spheres are separated by 12 cm. If the radius of each sphere is 2.8 cm and the electrical potential energy in the arrangement is 3.6 × 10^{-4} J, determine the charge on each sphere.

41 Determine how much electrical potential energy is associated with the proton–electron arrangement in a hydrogen atom. (separation = 5.3 × 10^{-11} m)

Electric potential

SYLLABUS CONTENT

- The electric potential is a scalar quantity with zero defined at infinity.
- The electric potential V_e at a point is the work done per unit charge to bring a test charge from infinity to that point as given by: $V_e = \frac{kQ}{r}$.
- The electric field strength E as the electric potential gradient as given by: $E = -\frac{\Delta V_e}{\Delta r}$.
- The work done in moving a charge q in an electric field as given by: $W = q\Delta V_e$.
- Equipotential surfaces for electric fields.
- The relationship between equipotential surfaces and electric field lines.

◆ **Electric potential** Work done in moving a test charge of +1 C to a specified point from infinity.

The concept of **electric potential**, V_e, is used to describe points in the space around charges (Compare with gravitational potential.)

Electric potential can be considered as electric potential energy per unit charge.

More precisely, it is defined as follows:

The electric potential at a point is defined as the work done per unit charge (1 C) in bringing a small positive test charge from infinity to that point.

The SI unit for electric potential is J C^{-1}. This should be familiar from discussing potential difference in Topic B.5: 1 J C^{-1} is called 1 volt (V).

For a relatively small charge in the electric field of a larger charge, we can make that clear by rewriting

$$E_p = k\frac{q_1 q_2}{r} \text{ as } E_p = k\frac{Qq}{r}$$

Then, dividing by the small charge, q, gives

Electric potential around a point charge Q:

$$\boldsymbol{V_e = \frac{kQ}{r}}$$

The potential around a negative charge will be negative (as shown by the equation). Increasing the distance, r, from the charge, $-Q$, reduces the magnitude of the negative potential, which is equivalent to an increase in potential. This is similar to gravitational fields.

The potential around a positive charge will be positive (as shown by the equation). Increasing the distance, r, from the charge, $+Q$, reduces the magnitude of the positive potential, which is equivalent to a decrease in potential.

WORKED EXAMPLE D2.10

Calculate the electric potential due to a point charge of -1.00×10^{-8} C at a distance of:

a 1.00 m **b** 2.00 m.

Answer

a $V_e = \frac{kQ}{r} = \frac{(8.99 \times 10^9) \times (-1.00 \times 10^{-8})}{1.00} = -89.9\text{ V}$

b $V_e = \frac{kQ}{r} = \frac{(8.99 \times 10^9) \times (-1.00 \times 10^{-8})}{2.00} = -45.0\text{ V}$

The potential increases by 45.0 V when moving from 1.00 m to 2.00 m from the charge.

Electric equipotential surfaces and lines

Drawings of equipotential lines provide useful visualizations of electric fields.

An equipotential surface (or line) connects places which have the same electric potential.

Equipotential lines are always perpendicular to electric field lines.

No overall work is done if a charge moves between different positions on the same equipotential line (surface).

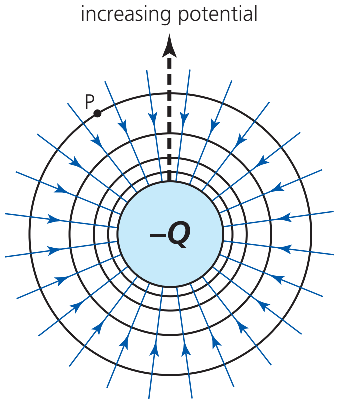


Figure D2.48 Equipotential and field lines around a negative charge

Figure D2.48 shows electric equipotential lines and electric field lines around a spherical negative charge, −Q. The circular lines are drawn with equal numerical intervals of potential, which means that they must get further and further apart because the field is weakening. A three-dimensional representation would have spherical *surfaces*.

A test positive charge placed at P would be attracted to −Q, as shown by the arrows on the field lines. Moving a test charge further away from −Q requires work to be done, so that the electric potential energy and potential must increase.

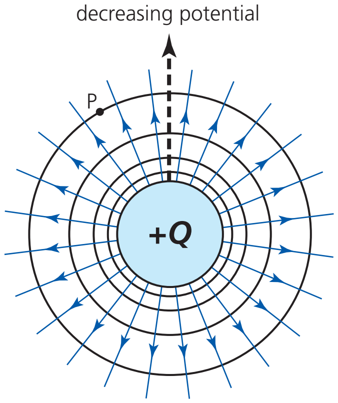


Figure D2.49 Equipotential and field lines around a positive charge

In Figure D2.49 the central charge is positive, +Q. A test positive charge placed at P would be repelled from +Q. so that the electric potential energy and potential must decrease if it is able to move.

Figure D2.50 represents the field and potential *inside* and outside of a positively charged conducting sphere (solid or hollow).

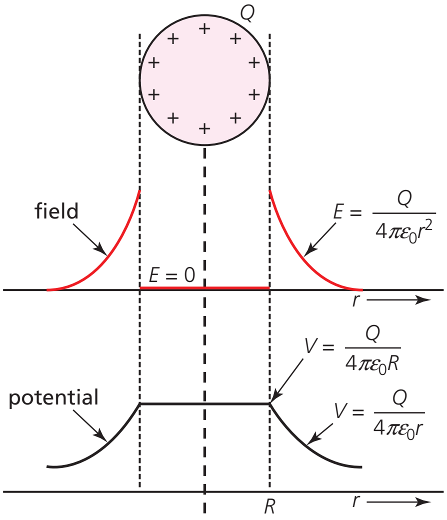


Figure D2.50 Electric field and potential of a conducting sphere

◆ **Electrode** Conductor used to make an electrical connection to a non-metallic part of a circuit.

ATL D2B: Thinking skills

Applying key ideas and facts in new contexts

Mapping electric fields

Figure D2.51 shows a way in which potential and potential difference can be mapped experimentally using conducting paper.

Two or more metal **electrodes** are placed in good electrical contact with a special type of paper which has been coated with carbon so that it conducts electricity, but still has significant resistance. The shape and location of the electrodes can be varied. A p.d. is connected across the electrodes and, typically, one electrode is kept at 0 V. A movable probe is connected between 0 V and a point of interest in the electric field between the electrodes. Lines of equipotential are easily identified. The voltmeter will display the potential at that point. Alternatively, the voltmeter can be used to determine the p.d. between any two points in the field.

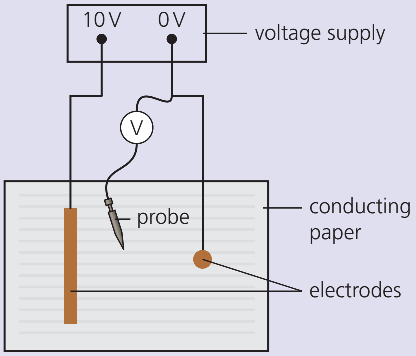


Figure D2.51 mapping potential

Faraday's cage

If a constant electric field is applied to a metal conductor surrounding a space, free electrons will very quickly redistribute themselves on the outer surface depending on the strength and direction of the field, and the shape of the conductor. For a spherical conductor the charge distribution would be the same everywhere on the surface.

The safest place to be in a lightning storm is inside a conductor, like a car or a building (that has a lightning conductor if it is in an exposed location).

Figure D2.52 shows a dramatic example of a 'Faraday' cage.

Electronic equipment can be protected by putting it inside a Faraday cage.

Effective electromagnetic shielding is used widely to protect important equipment from external electromagnetic waves, or to stop the emission of electromagnetic signals.

■ **Figure D2.52** A Faraday cage, showing sparks on the outside, but with someone safe inside

In pairs, research Faraday cages and other forms of electromagnetic shielding. Using what you know about electric fields and conductors, explain how they shield objects placed inside from electric fields.

Combining electric potentials

Electric potential is a scalar quantity and potentials can be combined by simple addition.

WORKED EXAMPLE D2.11

Determine the combined potential at a point which is 34 cm from a point charge A of -1.9×10^{-7} C and 45 cm from a point charge B of $+2.3 \times 10^{-7}$ C.

Answer

$$V_e = \left(\frac{kQ}{r}\right)_A + \left(\frac{kQ}{r}\right)_B = \frac{(8.99 \times 10^9) \times (-1.9 \times 10^{-7})}{(34 \times 10^{-2})} + \frac{(8.99 \times 10^9) \times (+2.3 \times 10^{-7})}{(45 \times 10^{-2})} = -4.3 \times 10^2 \text{ V}$$

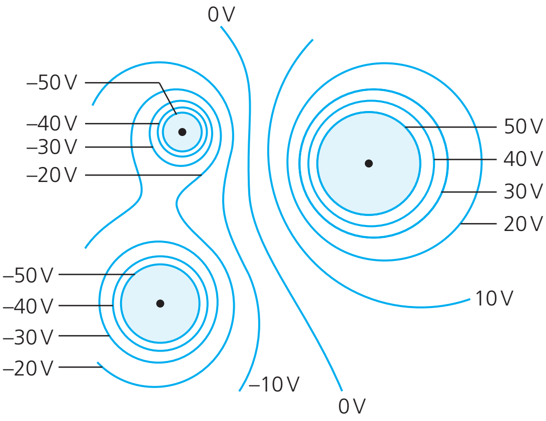


Figure D2.53 shows an example of potential mapping around three charged spheres. Remember that, if required, field lines can be drawn perpendicular to the equipotential lines.

■ **Figure D2.53** Equipotential lines around three charged spheres

42 If 4.95×10^{-5} J of energy were transferred when a charge of $5.1\,\mu$C was moved from a certain point to earth (ground), determine the magnitude of the potential at the point.

43 At what distance from an isolated point charge of 4.6×10^{-8} C would the electric potential have a value of −3000 V?

44 Sketch the equipotential and field lines around two point charges of different magnitudes if

a they have similar signs

b they have opposite signs.

45 Figure D2.54 shows the electric field lines around an isolated charged conductor. Make a copy of Figure D2.54 and add lines of equipotential.

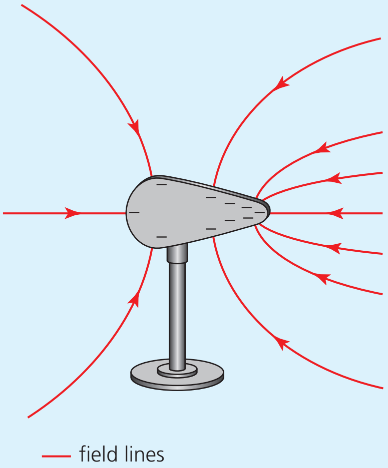


Figure D2.54 Electric field lines around an isolated charged conductor

46 Figure D2.55 shows a kind of *coaxial cable* that is widely used for transferring data, such as signals to televisions. The outer copper mesh is maintained at 0 V and the signal is sent as an electromagnetic wave in the insulator between the central copper wire and the surrounding mesh. This design, with its earthed outer mesh, helps reduce interference to and from other electromagnetic waves. Make a sketch of a cross-section of the cable and add electric field lines and equipotential lines.

Figure D2.55 Coaxial cable

47 Figure D2.56 represents the variation of potential between two conductors. Sketch the electric field pattern of this arrangement.

48 A hollow conducting sphere has a radius of 6.3 cm. If it is charged with −4.3 nC, determine values for the electric field strength and the electric potential at a distance of:

a 20.0 cm from the centre of the sphere

b 3.0 cm from the centre of the sphere.

49 A point charge of -1.9×10^{-7} C is placed 56 cm from another point charge of 5.6×10^{-8} C. Identify the location of one place where the electric potential is zero.

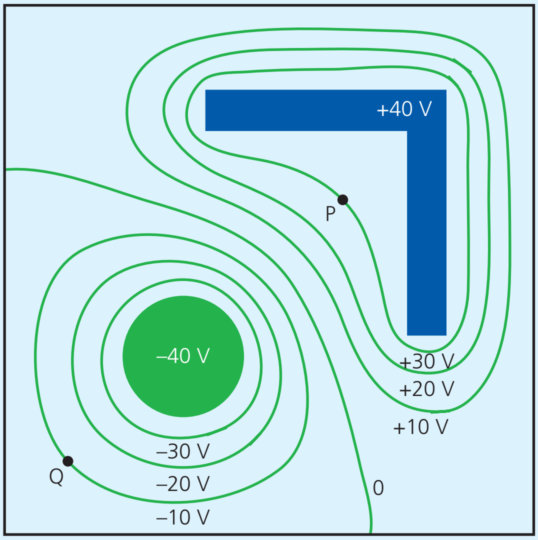


Figure D2.56 The variation of potential between two conductors

Electric potential difference

The central theme of this topic is the movement of charges between different places in electric fields. This means that the difference in potential – the *potential difference* – between two locations is of particular importance.

◆ **Potential difference, p.d. ΔV_e** Difference in electric potential between two points, which equals the work done when a charge of 1 C is moved between the points.

Electric potential difference, ΔV_e, is the work, W, done on unit charge (1 C) when it moves between two points in an electric field.

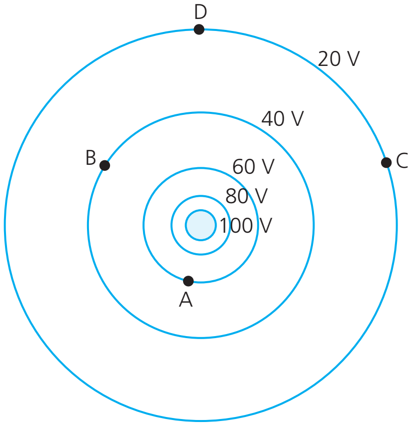
$$\Delta V_e = \frac{W}{q} \quad \text{or} \quad W = q\Delta V_e$$

The unit for electric potential difference is the same as for potential: the volt (joule per coulomb).

The same concept (electric potential difference) was used frequently in Topic B.5, where it was abbreviated to p.d. or referred to as voltage. In this topic we are applying the term more generally in two and three dimensions, whereas in Topic B.5 our focus was just on electric potential differences across components in electrical circuits.

WORKED EXAMPLE D2.12

Consider Figure D2.57 which shows equipotential lines around a conducting sphere.



■ **Figure D2.57** Equipotential lines around a conducting sphere

a State whether the sphere is positively or negatively charged.

b Calculate the potential difference when moving from:

  i C to A

  ii C to D.

c Determine how much work is done when a charge of +2.0 C moves from B to C.

Answer

a Positively charged (potentials are positive)

b **i** (+60) – (+20) = + 40 V

  ii (+20) – (+20) = + 0 V

c $W = q\Delta V_e = 2.0 \times [(+20) - (+40)] = -40$ J

The negative sign shows that electric potential energy falls. (A freely moving positive charge will be repelled from the positive sphere and gain kinetic energy.)

Electric potential–distance graphs

We know that the work done, W, when moving a charge, q, through a potential difference ΔV_e is given by $W = q\Delta V_e$.

We also know that the work can be calculated from force × distance = $Eq \times \Delta r$, where Δr is a small enough distance that the value of E does not change significantly.

Hence $W = \cancel{q}\Delta V_e = E\cancel{q}\Delta r$, so that:

DB

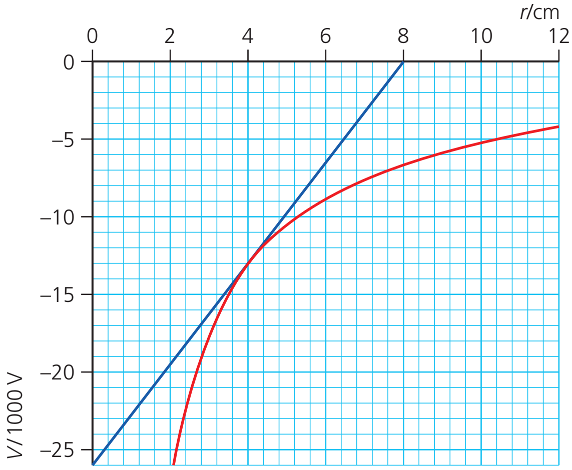
$$\text{electric field strength, } \boldsymbol{E} = -\frac{\Delta \boldsymbol{V}_e}{\Delta \boldsymbol{r}}$$

where $\frac{\Delta V_e}{\Delta r}$ is called the electric *potential gradient.*

The negative sign has been added to the equation to show that the direction of the vector quantity E is opposite to the direction of increasing potential.

In other words, electric fields exist where electric potential is changing. If potential is constant, then the electric field strength is zero.

Figure D2.58 shows how the value of potential varies around a negative point charge. The tangent to the curve can be used to determine the gradient at any required distance (the example is for $r = 4.0$ cm).



■ **Figure D2.58** Variation of potential around a point charge

WORKED EXAMPLE D2.13

Determine:

a the electric field strength at a distance of 4.0 cm from the point charge represented in Figure D2.58

b the value of the point charge involved.

Answer

a $E = -\frac{\Delta V_e}{\Delta r} = -\frac{0 - (-26 \times 10^3)}{(8.0 \times 10^{-2})}$

$= -3.3 \times 10^5\,\mathrm{N\,C^{-1}}$

b $V_e = \frac{kQ}{r}$

$-15 \times 10^3 = \frac{(8.99 \times 10^9) \times Q}{3.5 \times 10^{-2}}$

$Q = -5.8 \times 10^{-8}\,\mathrm{C}$

50 Two points in an electric field have potentials of 12.7 V and 15.3 V. Determine how much energy will be transferred when an electron moves between these points:

a in eV **b** in J.

51 Calculate how much energy is gained by a charge of 2.0 C when it passes through a battery that has a terminal p.d. of 12 V.

52 A charge of +4.5 C is moved from point P to point Q in Figure D2.56.

a Determine how much energy is transferred.

b State whether work is done on the charge, or by the charge.

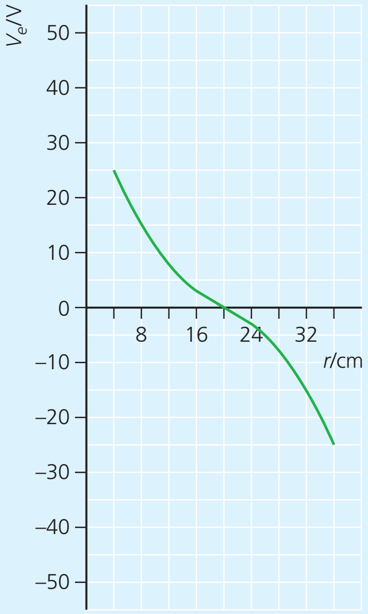
53 Figure D2.59 shows how the electrical potential varies with distance, *r*, from a certain point, P.

a State where the electric field strength is a minimum.

b Determine the magnitude of the electric field strength:

i 12 cm from P **ii** 32 cm from P.

c Suggest what arrangement of charges might produce this variation in potential.

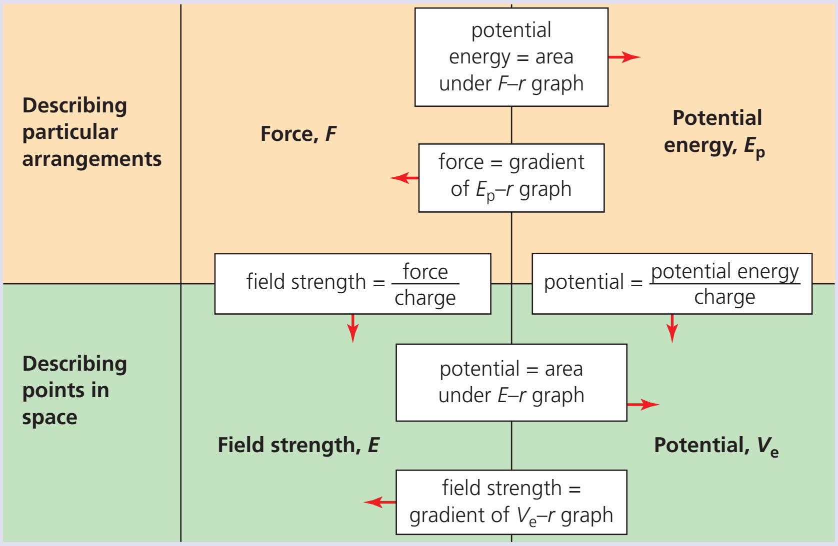


■ **Figure D2.59**

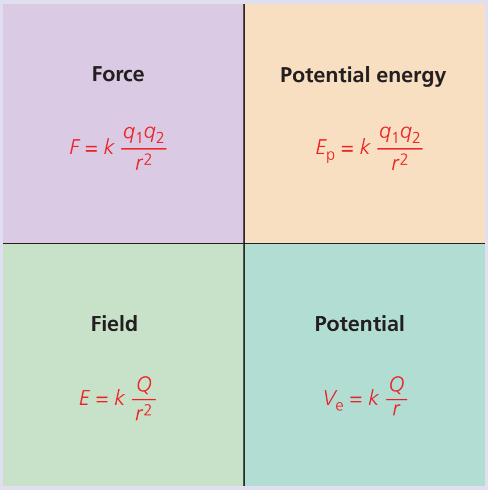
ATL D2C: Communication skills

Clearly communicating complex ideas in response to open-ended questions

As you may have seen in Topic D.1, we can summarize key concepts and their connections using a visual organizer. Here are two visual organizers for the key concepts in this topic.



■ **Figure D2.60** Connections between the four key concepts



Force $F = k\frac{q_1q_2}{r^2}$	**Potential energy** $E_p = k\frac{q_1q_2}{r^2}$
Field $E = k\frac{Q}{r^2}$	**Potential** $V_e = k\frac{Q}{r}$

■ **Figure D2.61** Equations for radial electric fields

Can you think of other ways in which you might represent the concepts from this topic and their connections?

LINKING QUESTION

- How are electric and magnetic fields like gravitational fields?

This question links to understandings in Topic D.1.

D.3 Motion in electromagnetic fields

Guiding questions

- How do charged particles move in magnetic fields?
- What can be deduced about the nature of a charged particle from observations of it moving in electric and magnetic fields?

The motion of charged particles in uniform electric fields

SYLLABUS CONTENT

- ▶ The motion of a charged particle in a uniform electric field.
- ▶ The motion of a charged particle in a perpendicularly orientated uniform electric field.

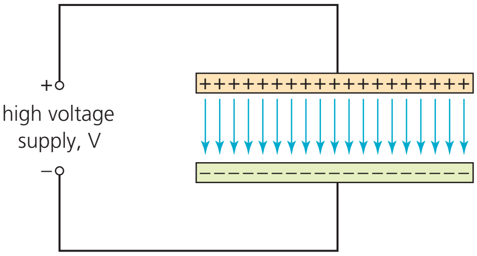


Figure D3.1 Producing a uniform electric field

In this section we will discuss the motion of charged particles (electrons, protons and ions) that are free to move in uniform electric fields. We will assume that the particles are in a vacuum (unless otherwise stated), so that their movements do not involve collisions with other particles (air molecules).

In Topic D.2 we explained that the easiest way of producing a uniform electric field was by connecting a potential difference across parallel metal plates, as shown again in Figure D3.1.

Stationary charged particles

From an understanding of the kinetic theory of matter, it should be appreciated that particles are never truly ‘stationary’. However, even a small p.d. can accelerate charged particles to speeds very much greater than their random velocities without the p.d. So, assuming that a particle is stationary to begin with will not result in any significant error when determining its final speed and kinetic energy.

A particle with charge q situated in an electric field will experience a force $F = Eq$ towards the oppositely charged plate (Topic D.2). Figure D3.2 shows two opposite charges in an electric field.

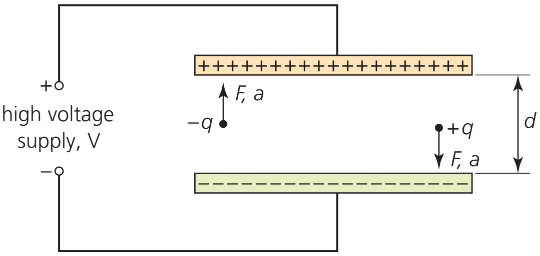


Figure D3.2 Forces and accelerations of opposite charges in a uniform electric field

Since $E = \frac{V}{d}$, $F = \frac{Vq}{d}$.

The forces will cause the charges to accelerate perpendicularly towards the plates. The equations of motion (Topic A.1) can then be used to determine how the charge moves.

‘Stationary’ mobile charges will accelerate along a field line in a uniform electric field.

WORKED EXAMPLE D3.1

An electron is very close to the negatively charged plate seen in Figure D3.2.

a Determine its speed when it reached the positive plate if there was a p.d. of 2000 V across the metal plates which were separated by 8.0 cm.

b State any assumptions you made when answering part **a**.

Answer

a $F = \frac{Vq}{d} = \frac{2000 \times (1.60 \times 10^{-19})}{0.080} = 4.0 \times 10^{-15}\,\text{N}$

Then, acceleration can be determined using Newton's second law ($F = ma$):

$a = \frac{F}{m} = \frac{4.0 \times 10^{-15}}{9.110 \times 10^{-31}} = 4.4 \times 10^{15}\,\text{m}\,\text{s}^{-2}$

Finally, the final speed can be determined from $v^2 = u^2 + 2as$:

$v^2 = 0^2 + (2 \times 4.4 \times 10^{15} \times 0.080)$

$v = 2.7 \times 10^7\,\text{m}\,\text{s}^{-1}$

b The calculation has assumed that:

- the electron started with speed $u = 0$
- there were no gas molecules between the plates (that would collide with the electrons).

Charged particles moving across a uniform electric field

This situation is analogous to the motion of masses projected in uniform gravitational fields. Both situations involve motion which can be analysed as a constant speed in one direction and an acceleration in a perpendicular direction. See Figure D3.3.

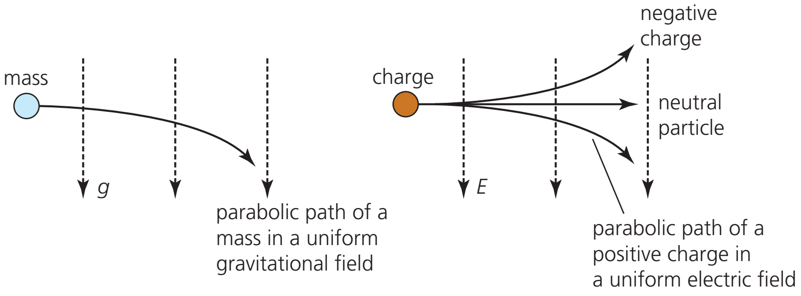


Figure D3.3 Comparing motion in gravitational and electric fields

Freely moving charged particles travelling across electric fields will move in parabolic paths.

WORKED EXAMPLE D3.2

An electron travelling horizontally at a constant speed of $5.5 \times 10^7\,\text{m}\,\text{s}^{-1}$ is directed into a uniform electric field of $2.8 \times 10^5\,\text{N}\,\text{C}^{-1}$ acting vertically downwards.

a Determine the force and acceleration of the electron in the field (magnitude and direction).

b If the field extends for a horizontal distance of 10 cm, determine the time that the electron spends in the field.

c Determine the vertical displacement of the electron from its original path as it leaves the field. (Charge on electron is -1.6×10^{-19} C, mass of electron is 9.110×10^{-31} kg.)

Answer

a $F = Eq = (2.8 \times 10^5)(1.60 \times 10^{-19}) = 4.5 \times 10^{-14}\,\text{N}$ upwards

$a = \frac{F}{m} = \frac{4.48 \times 10^{-14}}{9.110 \times 10^{-31}} = 4.9 \times 10^{16}\,\text{m}\,\text{s}^{-2}$ upwards

(4.48×10^{-14} seen on calculator display)

($4.9177... \times 10^{16}$ seen on calculator display)

b Horizontal speed is constant:

$v = \frac{s}{t}$

$5.5 \times 10^7 = \frac{0.10}{t}$

$t = 1.8 \times 10^{-9}\,\text{s}$

($1.81818... \times 10^{-9}$ s seen on calculator display)

c Using the equation of motion (Topic A.1):

$s = ut + \frac{1}{2}at^2 = 0 + \left(\frac{1}{2} \times (4.918 \times 10^{16}) \times (1.818 \times 10^{-9})^2\right)$

$= 0.081\,\text{m}$ (8.1 cm)

Tool 3: Mathematics

Express measurements and processed uncertainties to an appropriate number of significant figures or level of precision

Worked example D3.2 shows the effect of 'rounding off' too early in a multi-step calculation. All three answers are correctly given to 2 significant figures (the same as the data given in the question). However, if those answers were used in part **c**, a different answer (0.079 m) would be obtained.

◆ **Particle beams** Streams (flows) of very fast-moving particles, most commonly charged particles (electrons, protons or ions), moving across a vacuum. Properties of the individual particles can be investigated by observing the behaviour of the beams in electric and/or magnetic fields.

◆ **CERN** European organization for nuclear research.

◆ **Cathode** An electrode out of which (conventional) current flows.

◆ **Thermionic emission** Release of electrons from a very hot metal surface.

◆ **Anode** An electrode into which (conventional) current flows.

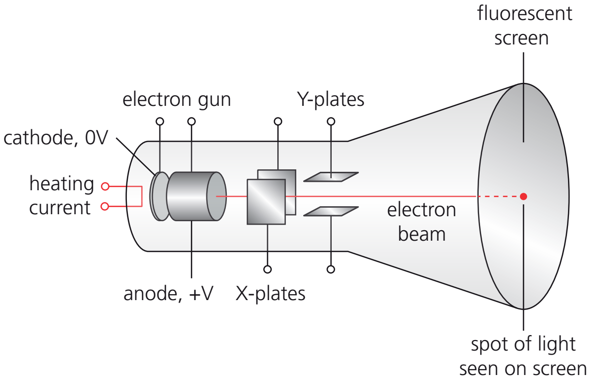
Particle beams

So far, we have been discussing the motion of *individual* charged particles but, in practice,

experiments with charged particles will involve large numbers moving together with the same velocity as a **particle beam**.

Experiments with particle beams have had great historic importance (for example, the discovery of the electron), and they continue to be an essential part of the latest research into nuclear physics (at **CERN**, for example, see Figure D3.12).

We will use the production of an electron beam as an example. Figure D3.4 shows a type of electron beam deflection tube that is commonly used to demonstrate to students the production and properties of electron beams.



■ **Figure D3.4** Electron deflection tube

The heating current is used to heat the metal **cathode** (the terminal connected to 0 V) and the thermal energy supplied increases the kinetic energy of the free electrons in the metal. Some of the electrons have energy to be released (emitted) from the metal's surface. This process is called **thermionic emission**. A large positive voltage is applied to the other terminal (called the **anode**) and this accelerates the electrons into a beam travelling with very high speeds to the right (as shown). The tube contains a vacuum. This arrangement is commonly called an 'electron gun'. When the beam of electrons strikes the fluorescent screen at the end of the tube, some of their kinetic energy is transferred to visible light in the form of a spot that can be easily observed.

If a p.d. is connected across the 'X-plates' the beam (and the spot) is deflected to the left or to the right. If a p.d. is connected across the 'Y-plates' the beam is deflected up or down.

1 **a** Calculate the force exerted on a singly charged positive ion in an electric field of 1200 $\mathrm{N\,C^{-1}}$ acting vertically downwards.

b Determine the mass of the ion if it started to accelerate at $7.4 \times 10^9\,\mathrm{m\,s^{-2}}$.

c State in which direction the ion will accelerate.

2 A proton is accelerated by a p.d. of 3.7×10^4 V connected between two parallel metal plates which are 2.8 cm apart.

a Calculate the strength of the electric field.

b Assuming that the electric field is uniform, determine what force the proton will experience.

c Determine the maximum amount of energy that the proton can gain when it has been accelerated in

i eV **ii** J.

3 Show that the final speed of an electron accelerated from rest over a distance of 5.0 cm in a vacuum by a uniform electric field of $9.2 \times 10^4\,\mathrm{N\,C^{-1}}$ is about forty million metres per second.

4 An electron beam consisting of electrons travelling at speeds of $1.3 \times 10^7\,\mathrm{m\,s^{-1}}$ is directed horizontally into a uniform electric field of $7.4 \times 10^4\,\mathrm{N\,C^{-1}}$ acting vertically upwards.

- **a** Determine the force and acceleration of the electrons in the field (magnitude and direction).
- **b** If the field extends for a horizontal distance of 8.5 cm, calculate how much time the electrons spend in the field.
- **c** Determine the vertical displacement of the electron from its original path as it leaves the field.
- **d** State the name given to the shape of the electron beam's path.

5 An alpha particle is emitted from the nucleus of a radium atom (Topic E.1) with kinetic energy of 4.8 MeV.

- **a** Show that the initial speed of the alpha particle is between ten million and twenty million metres per second. (Mass of alpha particle = $6.64 \times 10^{-27}\,\mathrm{kg}$)
- **b** An alpha particle has a charge of $3.2 \times 10^{-19}\,\mathrm{C}$. Determine the force acting on the particle when it is moving perpendicularly across a uniform electric field of strength $4.9 \times 10^4\,\mathrm{N\,C^{-1}}$.
- **c** Explain, by considering your answers to Question 4, but without a detailed calculation, why the particle will not be significantly deflected in the field.

The motion of charged particles in uniform magnetic fields

SYLLABUS CONTENT

- The motion of a charged particle in a uniform magnetic field.
- The motion of a charged particle in a perpendicularly orientated uniform magnetic field.
- The magnitude and direction of the force on a charge moving in a magnetic field as given by: $F = qvB\sin\theta$

A 'stationary' charge in a magnetic field will not experience any magnetic force.

Any charge *moving* in a magnetic field will experience a magnetic force unless its motion is parallel to the magnetic field (that is, if it is moving along a magnetic field line). The magnitude of the force increases with the speed of the particle.

Consider Figure D3.5 which shows a charge q entering a magnetic field of strength B. The constant velocity, v, of the charge makes an angle θ with the direction of the magnetic field.

A moving charge will experience a force which is perpendicular to both the directions of its velocity and the magnetic field.

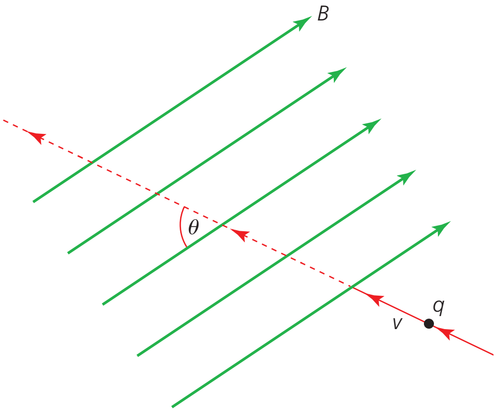


Figure D3.5 An individual charge entering a magnetic field

Since in Figure D3.5 the velocity and the field are both in the plane of the paper, we know that the force on the charge will act perpendicularly into, or out of, the paper. The direction can be predicted using **Fleming's left-hand rule**, as shown in Figure D3.6.

If the moving charge shown in Figure D3.5 was positive, using Fleming's left-hand rule predicts that the force on the particle acts perpendicularly downwards, into the paper. If the particle was negatively charged, the same rule predicts that the force on the particle acts perpendicularly upwards, out of the paper.

Nature of science: Observations

Once we have understood that a charged particle needs to be *moving* across a magnetic field in order to experience a force, we probably should not be surprised that the magnitude of that force increases with the speed of the particle. However, this is an example in which fundamental physics seems to contradict our 'common sense'. An explanation requires a relativistic treatment.

◆ **Left-hand rule (Fleming's)** Rule for predicting the direction of the magnetic force on moving charges, or a current in a wire.

Top tip!

Any moving charge can be considered to be an electric current. Remember that, by convention, the direction of current is *always* shown to be the direction in which positive charges are moving (Topic B.5). If the charges are negative, the conventional current will be shown to be in the opposite direction to their velocity.

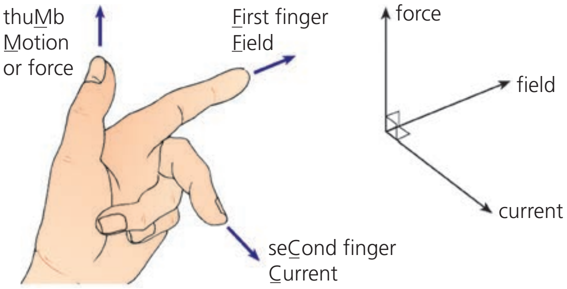


Figure D3.6 Fleming's left-hand rule predicts the direction of the force

Because of the magnetic force, the charged particle will not continue to move in a straight line. Because the force is perpendicular to the velocity, this is the necessary condition for circular motion (Topic B.2) *if* the particle is moving perpendicularly across the field.

A charged particle moving perpendicularly across a uniform magnetic field will follow a circular path.

The charged particle will move along an arc of a circular path as long as it remains in the magnetic field. Figure D3.7 shows the four different possibilities, each for particles travelling with different speeds. Crosses represent fields into the paper/screen and dots represent fields out of the paper/screen.

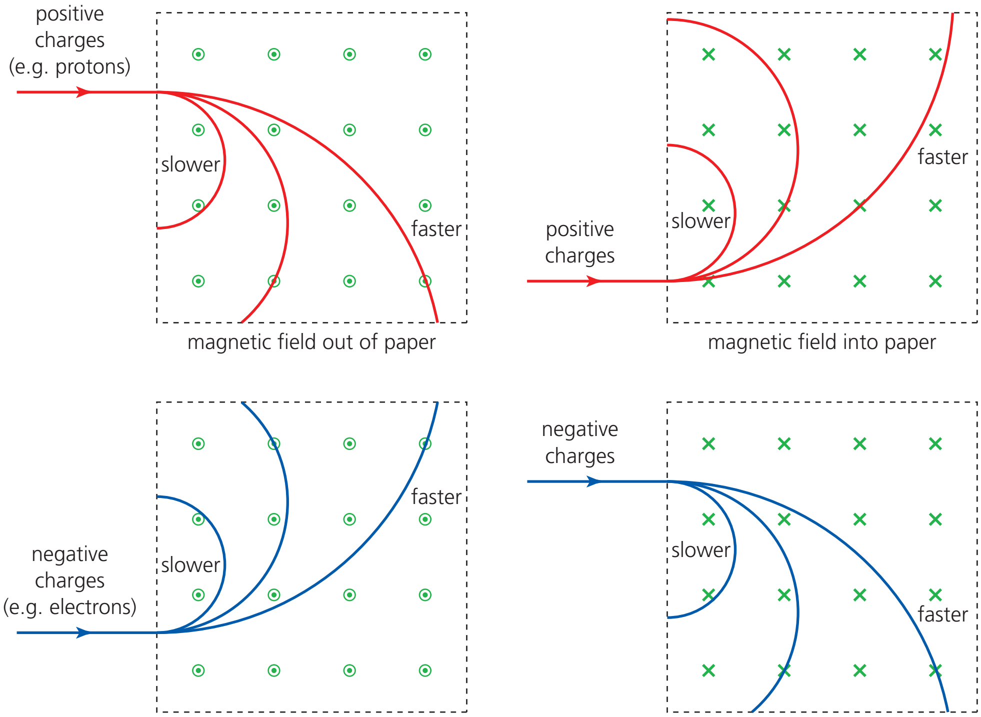


Figure D3.7 Circular paths of charges moving perpendicular to magnetic fields

During its circular motion in a perpendicular magnetic field, the speed and the kinetic energy of the charged particle will remain constant.

TOK

Knowledge and the knower

- *How do we acquire knowledge?*

Mapping

'**Mapping**' is the process of representing information in the form of a diagram, map, or picture. For example, you may choose to *map* the basic ideas in the physics course in order to show the interconnections between them.

The purpose of mapping is to simplify and make something easier to understand.

The use of lines and patterns to represent fields is accepted by the scientific community as possibly the only way of presenting these difficult ideas simply to the human mind. No one thinks that the lines are 'real' and it might be argued that such a simplification in some ways restricts our understanding, or imagination, about the subject because it channels our thoughts in certain prescribed directions. The mapping of any knowledge is a simplification to aid understanding and one which has obvious appeal but, like all simplifications, has its limitations.

Make a list of the key concepts introduced in Theme D (so far), then display them on a full page annotating the connections between them.

◆ **Mapping** Representing the interrelationships between ideas, knowledge or data by drawing.

LINKING QUESTION

- How are the properties of electric and magnetic fields represented? (NOS)

Equation for the force on a charged particle moving across a magnetic field

The magnitude of the force, F, on a charged particle moving across a uniform magnetic field depends on the:

- charge of the particle, q
- velocity of the particle, v
- strength of the magnetic field, B
- angle between the field and the velocity, θ.

The force is proportional to q, v, B and $\sin\theta$, so that:

the magnitude of the magnetic force on a charge moving in a magnetic field is given by:

$$\boldsymbol{F = qvB\sin\theta}$$

WORKED EXAMPLE D3.3

Determine the force experienced by an electron moving with a speed of $4.7 \times 10^6\,\text{m s}^{-1}$ at an angle of 50° across a magnetic field of strength 0.56 T.

Answer

$F = qvB\sin\theta = (1.60 \times 10^{-19}) \times (4.7 \times 10^6) \times 0.56 \times \sin 50° = 3.2 \times 10^{-13}\,\text{N}$

If a particle is moving *perpendicularly* to a uniform magnetic field, $\sin\theta = 1$ so that the equation for the force reduces to $F = qvB$. This magnetic force can be considered as the *centripetal force* causing circular motion:

we know from Topic A.2, centripetal force $F = \dfrac{mv^2}{r}$

So that:

$$qvB = \frac{mv^2}{r}$$

which can be rearranged to show that:

the radius of the circular path of a charged particle moving in a perpendicular magnetic field can be determined from:

$$r = \frac{mv}{qB}$$

WORKED EXAMPLE D3.4

Determine the radius of the path followed by a proton moving with a speed of $1.9 \times 10^7\,\text{m s}^{-1}$ perpendicularly across a magnetic field of strength 0.30 T. (Charge on proton is $+1.60 \times 10^{-19}\,\text{C}$, mass of proton is $1.673 \times 10^{-27}\,\text{kg}$.)

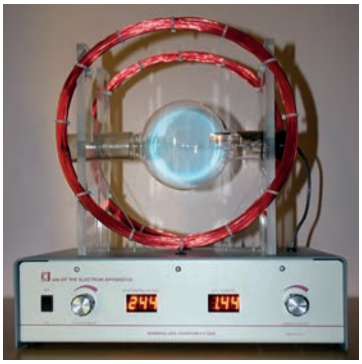
Answer

$$r = \frac{mv}{qB} = \frac{(1.673 \times 10^{-27}) \times (1.9 \times 10^7)}{(1.60 \times 10^{-19}) \times 0.30} = 0.66\,\text{m}$$

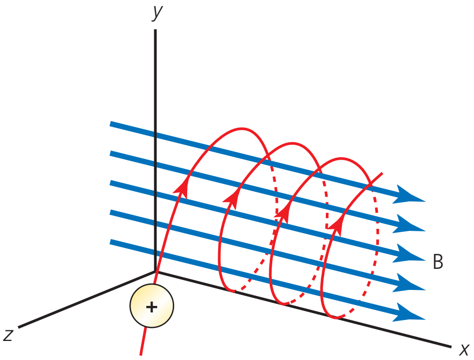
Figure D3.8 shows a school laboratory experiment to demonstrate the circular path of electrons moving perpendicular to a uniform magnetic field.

An electron beam is produced by an **electron gun** arrangement, as described previously (in Figure D3.4 but not visible in D3.8). The electrons are fired into a perpendicular uniform magnetic field, which is produced by steady currents in the coils which can be seen in Figure D3.8, which also shows the resultant circular path of the electron beam. The path of the electrons can be seen because the tube contains a very small amount of an inert gas at very low pressure. The gas molecules gain energy from collisions with the electrons and then emit light.

◆ **Electron gun**
Component that fires a beam of electrons across a vacuum.



■ **Figure D3.8** Electrons moving in circular paths



■ **Figure D3.9** Helical path

WORKED EXAMPLE D3.5

The electrons moving in the circular path seen in Figure D3.8 had been accelerated from rest by a voltage of 5000 V.

a Determine their maximum energy in:

  i electronvolts   **ii** joules.

b Calculate their maximum speed.

c If the strength of the magnetic field was 0.0033 T, what was the radius of the electrons' path?

Answer

a **i** 5000 eV

  ii $qV = (1.60 \times 10^{-19}) \times 5000 = 8.00 \times 10^{-16}\,\text{J}$

b $qV = \frac{1}{2}mv^2$

$$8.00 \times 10^{-16} = \frac{1}{2} \times (9.110 \times 10^{-31}) \times v^2$$

$$v = 4.19 \times 10^7\,\text{m s}^{-1}$$

c $r = \frac{mv}{qB} = \frac{((9.110 \times 10^{-31}) \times (4.19 \times 10^7))}{((1.60 \times 10^{-19}) \times 0.0033)} = 0.072\,\text{m}\ (7.2\,\text{cm})$

◆ **Helical** In the shape of a spiral.

If a charged particle is moving across a magnetic field but not moving perpendicularly or parallel to the field, its path will be **helical** (like a spiral), as shown in Figure D3.9 for positively charged particles.

6 Determine the magnetic force acting on a proton moving at an angle of 32° across a magnetic field of 5.3×10^{-3} T at a speed of $3.4 \times 10^5 \text{ m s}^{-1}$.

7 An electron is moving at a speed of $1.6 \times 10^7 \text{ m s}^{-1}$ perpendicularly to a magnetic field of 1.4×10^{-4} T. Calculate the radius of its path.

8 **a** Outline how it is possible for a charged particle to move through a magnetic field without experiencing a force.

b Discuss whether it is possible for the same particle to move through electric and gravitational fields without experiencing forces.

9 An alpha particle (Topic E.1) has a charge of $+3.2 \times 10^{-19}$ C and a mass of 6.7×10^{-27} kg. . It moves perpendicularly across a magnetic field in a vacuum with a speed of $1.4 \times 10^7 \text{ m s}^{-1}$

a If it experiences a magnetic force of 4.1×10^{-14} N, determine the strength of the field.

b Describe the path of the alpha particle.

c When a similar alpha particle moved at an angle across the same magnetic field, the force it experienced was 3.3×10^{-14} N. Determine the angle between the field and the particle's velocity.

d Describe the shape of the particle's trajectory.

10 **a** A beam of singly charged ions ($q = 1.6 \times 10^{-19}$ C) is projected perpendicularly across a magnetic field of strength 0.87 T. Determine the speed of each ion if they each experience a magnetic force of 1.1×10^{-12} N.

b The ions move in the arcs of circular paths of radius 2.78 m. Determine their mass.

c If the beam was replaced with doubly charged ions of the same mass, but moving with half the speed, predict the radius of their path.

11 **a** Electrons are accelerated into a beam by a p.d. of 7450 V in a vacuum. Determine the kinetic energy of the electrons in:

i electronvolts **ii** joules.

b Calculate the final speed of the electrons.

c Determine the strength of magnetic field needed to make these electrons move in a circle of radius 14.8 cm.

d If the accelerating voltage is halved, predict by what factor the radius of the electrons' path will change (in the same field).

12 A charge of $+4.8 \times 10^{-19}$ C moving perpendicularly across a magnetic field of 1.9×10^{-2} T experiences a force of 9.5×10^{-14} N.

a Determine the speed of the particle.

b What electric field would be needed to produce the same force on this charge?

c Draw the path of the particle moving across the fields in a direction such that these two forces could be equal and opposite to each other (so that the resultant force on the particle was zero).

ATL D3A: Communication skills

Being curious about the natural world

Find out how the Aurora Borealis is formed (see Figure D3.10)

■ **Figure D3.10** The Aurora Borealis

Motion of charged particles in both an electric field and a magnetic field

SYLLABUS CONTENT

- The motion of a charged particle in perpendicularly orientated uniform electric and magnetic fields.

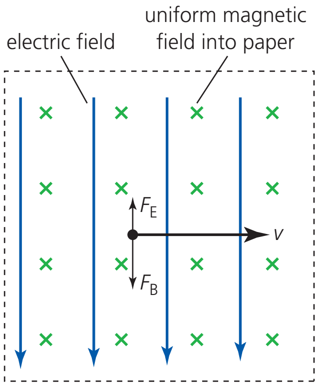


Figure D3.11 Three vectors perpendicular to each other

Any charged particle which is moving across both an electric and a magnetic field will experience two forces, one force parallel to the electric field and one force perpendicular to the magnetic field.

Consider the specific situation in which a charged particle is moving with a velocity v *perpendicularly* across an electric field, E, and both vectors are *perpendicular* to a magnetic field, B. That is, the three vectors are perpendicular to each other as shown in Figure D3.11. The directions of the two forces depend on the nature of the charge. In the example shown (for a negative charge), the forces are in opposite directions because of the relative directions of the fields. This is also true for a positively charged particle moving in this arrangement of fields.

This perpendicular arrangement of fields is of particular importance because, by adjusting the strengths of the two fields, it is possible for the forces on the charged particles to be made equal and opposite. That is, the forces cancel each other out, so that the particle continues its original motion in a straight line with a constant speed.

If the electric force, F_E = magnetic force, F_B: $Eq = Bqv$, so that:

if the motion of a charged particle is unaffected by perpendicular electric and magnetic fields, then, its velocity:

$$v = \frac{E}{B}$$

Tool 3: Mathematics

Check an expression using dimensional analysis of units

The units of both sides of the equation $v = \frac{E}{B}$ can be checked to confirm that they are equivalent to each other.

Units of E divided by units of B are:

$$\frac{\mathrm{N\,C^{-1}}}{\mathrm{N\,A^{-1}m^{-1}}} = \frac{\cancel{\mathrm{N\,A^{-1}}}\,\mathrm{s^{-1}}}{\cancel{\mathrm{N\,A^{-1}}}\,\mathrm{m^{-1}}} = \mathrm{m\,s^{-1}}$$

(the same units as v.)

WORKED EXAMPLE D3.6

An electron with a velocity of $5.9 \times 10^6\,\text{m s}^{-1}$ is passing between parallel metal plates which are separated by 10 cm. A uniform magnetic field of 42 mT is acting perpendicular to both the plates and the velocity of the electrons.

a Determine what p.d. across the plates will keep the electrons travelling in a straight line.

b If the direction of the electric field is downwards and the electrons are moving to the right, calculate the necessary direction of the magnetic field.

Answer

a $v = \dfrac{E}{B}$

$$5.9 \times 10^6 = \frac{E}{42 \times 10^{-3}}$$

$E = 2.5 \times 10^5\,\text{N C}^{-1}$ (or V m^{-1})

Then, since for a parallel plate arrangement $E = \dfrac{V}{d}$

$V = Ed = 2.5 \times 10^5 \times 0.10 = 2.5 \times 10^4\,\text{V}$

b The conventional current is to the left and electric force on a negatively charged electron is in the opposite direction to the electric field: upwards. The magnetic force must be downwards. Using the left-hand rule, the magnetic field must be directed towards the observer.

Charge to mass ratio of particles, *q/m*

◆ **Charge to mass ratio (of a particle)** The ratio *q/m* affects the motion of charged particles in electric and magnetic fields (important when charge and mass are not known separately).

If the exact nature of a particle is unknown, that is, neither the mass nor the charge of a particle is known, then the **charge / mass ratio**, q/m, becomes all-important. All particles with the same velocity, v and charge / mass ratio will follow paths of the same radius, r, when they pass into the same magnetic field, B, as shown by $r = \dfrac{mv}{Bq}$, as seen before:

In other words:

We cannot determine the charge on an unknown particle if we do not know its mass, or we cannot determine the mass of an unknown particle if we do not know its charge.

The charge to mass ratio of electrons was determined by J.J. Thomson (in 1897) before their mass and charge were confirmed separately.

Rearranging the previous equation, we get:

$$\frac{q}{m} = \frac{v}{Br}$$

Experiments such as that shown in Figure D3.8 can determine the radius of the particle beam's path in a known magnetic field. But to determine a mass to charge ratio, the speed of the particles must also be determined. This can be done as explained previously: perpendicular electric and magnetic fields are adjusted until the particles' motions are unaffected. Then $v = \dfrac{E}{B}$:

For example, if the charges moved with constant velocity when they were moving perpendicular to an electric field of $4.7 \times 10^6\,\text{N C}^{-1}$ and a magnetic field of 190 mT, then :

$$\text{speed, } v = \frac{E}{B} = \frac{4.7 \times 10^6}{190 \times 10^{-3}} = 2.5 \times 10^7\,\text{m s}^{-1}$$

If the same particle beam was directed perpendicularly across a separate magnetic field of strength 1.8 mT and the result was movement in the arc of a circle of radius 7.8 cm (similar to that seen in Figure D3.8), then:

$$\frac{q}{m} = \frac{v}{Br} = \frac{2.5 \times 10^{7}}{(0.0018 \times 0.078)} = 1.8 \times 10^{11}\,\mathrm{C\,kg^{-1}}$$

These values are consistent with a beam of electrons:

$$\text{charge} = 1.60 \times 10^{-19}\,\mathrm{C},\ \text{mass} = 9.110 \times 10^{-31}\,\mathrm{kg},\ \frac{\text{charge}}{\text{mass}} = 1.8 \times 10^{11}\,\mathrm{C\,kg^{-1}}$$

◆ **Particle accelerator** Apparatus designed to produce particle beams.

◆ **Superconducting** Without significant electrical resistance; only occurring at very low temperatures.

LINKING QUESTION

- How can conservation of energy be applied to motion in electromagnetic fields?

This question links to understandings in Topic A.3.

TOK

Knowledge and technology, and The natural sciences

- Why might some people regard science as the supreme form of all knowledge?
- To what extent are technologies merely extensions to the human senses, or do they introduce radically new ways of seeing the world?

CERN

The letters of CERN represent the Conseil Européen pour la Recherche Nucléaire. It has an informative website.

The main activity at CERN is the use of **particle accelerators** to produce the extremely high particle energies needed to investigate the fundamental forces and particles of nature. This is achieved by the use of extremely large and strong magnetic fields to force charged particles to keep moving faster and faster in circular paths. Then, subatomic particles are made to collide together at speeds close to the speed of light.

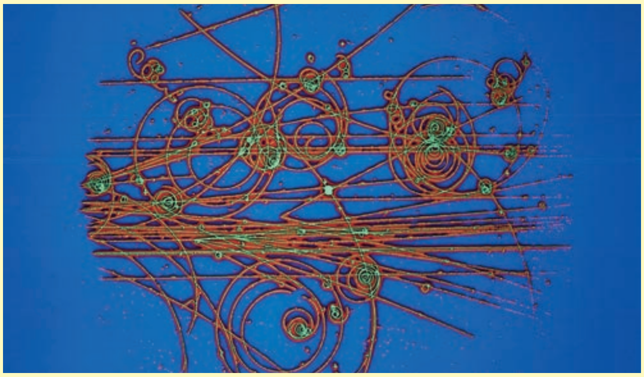
The Large Hadron Collider, shown in Figure D3.12, is the world's largest particle accelerator. *Hadrons* are a class of subatomic particles which includes protons and neutrons. The accelerator has a radius of 4.3 km and a circumference of 27 km with numerous **superconducting** magnets.

Figure D3.13 shows an example of the types of paths that can be produced by subatomic particles (produced following collisions) as they then move through a strong perpendicular magnetic field (in a *bubble chamber*). Measurements made from such images can lead to a determination of particle properties.

The following is a quote from the CERN website: *'The process (of colliding subatomic particles) gives us clues about how the particles interact, and provides insights into the fundamental laws of nature. We want to advance the boundaries of human knowledge by delving into the smallest building blocks of our universe'.*

Would you agree that CERN are aiming to improve the most fundamental form of knowledge?

■ **Figure D3.12** The Large Hadron Collider at CERN is underground and has a radius of about 4 km



■ **Figure D3.13** Curved paths of individual particles in a nuclear physics bubble chamber

13 We have stated that a particle will move in a circular path in a vacuum in a magnetic field which is perpendicular to its velocity. Figure D3.14 shows the path of a negatively charged particle in a container which contains some gas at low pressure.

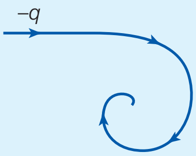


Figure D3.14 The path of a negatively charged particle

- **a** State the direction of the magnetic field.
- **b** Describe how the following quantities are changing:
  - **i** radius **ii** velocity **iii** kinetic energy.
- **c** Explain why these changes are occurring.

14 A charged particle is travelling parallel to, and mid-way between, two parallel metal plates which are separated by 15 cm and have a p.d. of 12.5 kV across them.

- **a** If the particle has a speed of $5.7 \times 10^6\,\text{m s}^{-1}$, what strength of magnetic field can be used to keep the particle moving with the same velocity?
- **b** State the direction in which the magnetic field must act.
- **c** Explain why you do not need to know the particle's charge to answer part **a**.

15 Use the internet to learn what a mass spectrometer is used for, and how they use electric and magnetic fields.

LINKING QUESTIONS

- What causes circular motion of charged particles in a field?
- How can the orbital radius of a charged particle moving in a field be used to determine the nature of the particle?
- How are the concepts of energy, forces and fields used to determine the size of an atom?

These questions link to understandings in Topics E.1 and E.2.

Forces on current-carrying conductors

SYLLABUS CONTENT

- The magnitude and direction of the force on a current-carrying conductor in a magnetic field as given by: $F = BIL\sin\theta$
- The force per unit length between parallel wires as given by: $\frac{F}{L} = \mu_0\frac{I_1 I_2}{2\pi r}$ where r is the separation between the two wires.

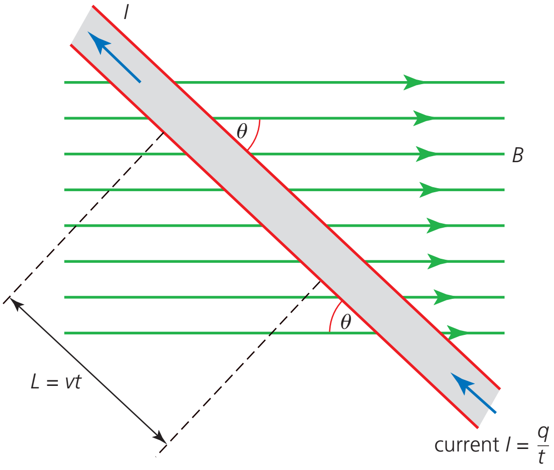


Figure D3.15 A wire carrying a current

In Topic B.5 we described the motion of *free electrons* constituting electric currents through conductors. Such electrons moving across a magnetic field in a current-carrying conductor will experience the same forces as if they were in an electron beam travelling across a vacuum.

Figure D3.15 shows a wire carrying a current I (= q/t) through a wire which is at an angle θ to a uniform magnetic field, B.

The electrons have an average (drift) speed of v through the wire, so that in time t they travel an average distance $L = vt$.

We can rewrite the equation ($F = qvB\sin\theta$) for the force as:

$$F = \left(\frac{q}{t}\right)(vt)B\sin\theta$$

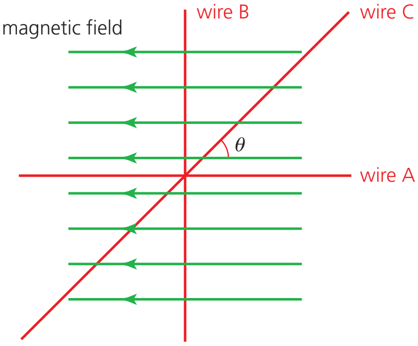


Figure D3.16 How force varies with the angle of the current to the magnetic field: there will be no force on wire A and the biggest force per unit length will be on wire B. Wire C will experience a force, but the force per unit length of wire C will be smaller than for wire B.

to show that:

> the force on a current I, passing through a conductor of length L across a uniform magnetic field B at an angle θ is given by:
>
> $$F = BIL \sin\theta$$

We briefly saw this equation in Topic D.2, where it was used to represent magnetic field strength as:

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

(when $\sin\theta = 1$)

Figure D3.16 illustrates how the magnetic force *per metre* depends on the angle of the wire to the magnetic field.

Note that the *total* force on the current in wire C would be the same as for wire B because there is a longer length in the field.

WORKED EXAMPLE D3.7

In Figure D3.17, a measured current is flowing in a wire across a small, uniform magnetic field.

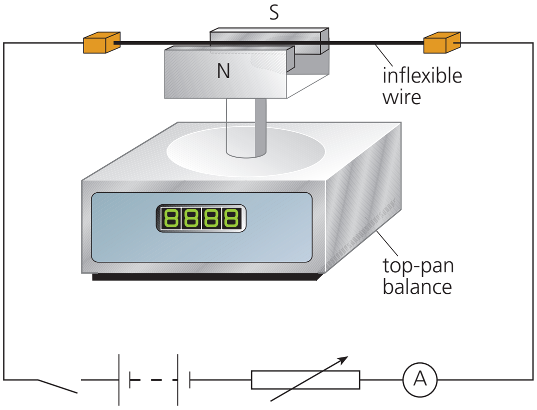


Figure D3.17 Current flowing in a wire across a small, uniform magnetic field

a State the direction in which the magnetic force is acting on the wire.

b In which direction is the force acting on the balance?

c When the current is flowing, the balance indicates that there is an extra mass of 4.20×10^{-2} g on the balance. Calculate the extra downwards force.

d If the current is 1.64 A and the length of the field is 8.13 cm, determine the strength of the magnetic field.

Answer

a Using the left-hand rule, the force is upwards.

b Using Newton's third law, the force is downwards.

c $F_g = mg = (4.20 \times 10^{-2} \times 10^{-3}) \times 9.81 = 4.12 \times 10^{-4}\,\text{N}$

d Using $F = BIL \sin\theta$, with $\sin\theta = 1$, gives:

$4.12 \times 10^{-4} = B \times 1.64 \times 0.0813$

$B = 3.09 \times 10^{-3}\,\text{T}$

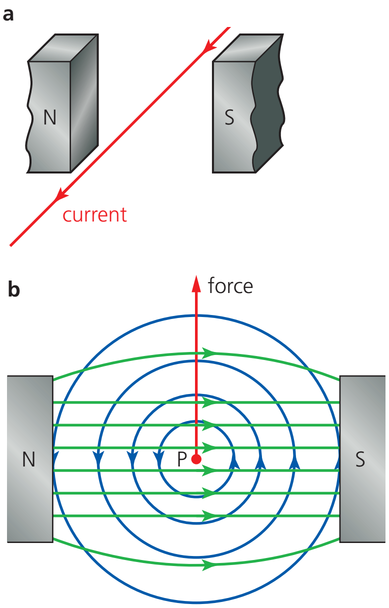


Figure D3.18 Comparing the directions of current, field and force

Figure D3.18 shows an alternative approach to understanding the magnetic force on a current in a magnetic field.

Figure D3.18a shows a wire carrying an electric current across a magnetic field. The current is perpendicular to the magnetic field from the permanent magnets. In Figure D3.18b the same situation is drawn in two-dimensional cross-section, with the wire represented by the point P and the magnetic fields from the magnets (shown in green) and from the current (shown in blue) included.

The two fields are in the same plane, so it is easy to consider the combined field that they produce. Above the wire, the fields act in opposite directions and they combine to produce a weaker field. Below the wire, the fields combine to give a stronger field. This difference in magnetic field strength on either side of the wire produces an upwards force on the wire, which can make the wire move (if it is not fixed in position).

Simple dc motor

The force acting on a wire crossing a magnetic field is commonly called the **motor effect** because it can be used to rotate a loop of wire, as shown in Figure D3.19.

◆ **Motor effect** Magnetic force on a current in a magnetic field, as used in electric motors.

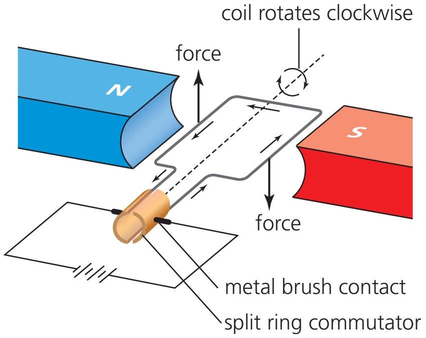


Figure D3.19 Essential parts of a dc motor

At the moment shown in Figure D3.19, the current on the right-hand side of the loop will experience a force downwards, while the current on the left-hand side will experience an upwards force because the current is flowing in the opposite direction. (Use Fleming's left-hand rule.)

It is not possible for a rotating loop to have fixed, permanent connections to an external power supply. The connection in Figure D3.19 is called a *commutator and brushes*. With this arrangement, the current will always enter the side of the loop on the right-hand side seen in the picture. In this way, the loop will experience the same forces every half rotation, which will keep it moving.

Increasing the current, strength of the magnetic field, or the number of turns in a coil will all make the motor spin faster. Winding the coil on an iron coil will also increase the rate of rotation.

Parallel current-carrying wires

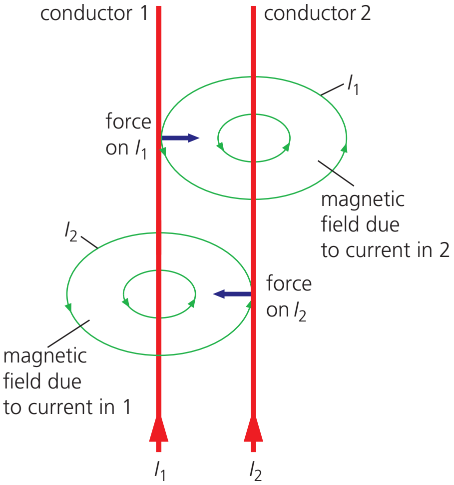


Figure D3. 20 Forces between parallel currents

Consider the two parallel wires carrying currents as shown in Figure D3.20.

In Topic D.2 we explained that the direction of the magnetic field created by a current in a single long straight wire can be determined using the *right-hand grip rule*.

If both wires are carrying a current, then each wire is in the magnetic field created by the current in the other. Both wires will experience a force and, using the left-hand rule, the forces will be attractive between the wires if the currents are in the same direction. The forces are equal and opposite (Newton's third law).

If the currents are in opposite directions, the wires will repel each other.

Figure D3.21 shows the combined magnetic fields produced in the two situations, looking down from above.

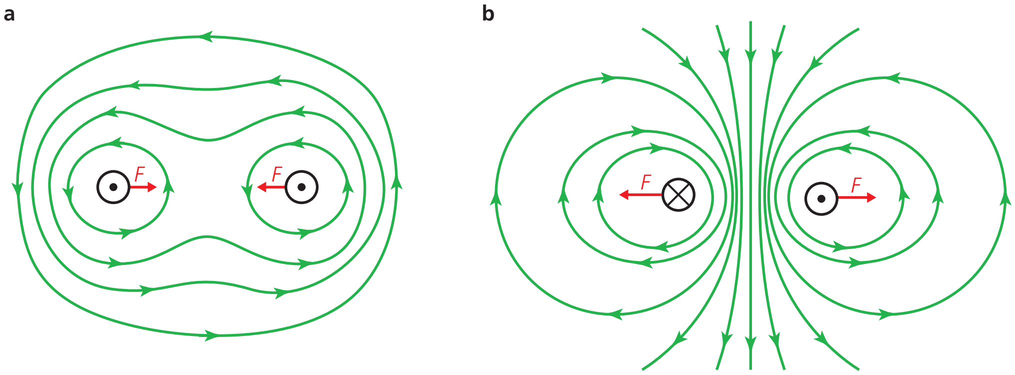


Figure D3.21 Magnetic fields around parallel currents in long wires : **a** currents in the same direction, and **b** currents in opposite directions.

The force on current I_1 can be determined from $F = B_2 I_1 L \sin\theta$, but in this case the field is perpendicular to the wire, so that $\sin\theta = 1$, which leads to:

force on unit length of conductor carrying current, $I_1 = \frac{F}{L} = B_2 I_1$

We saw in Topic D.2 that the field round a wire at a distance r can be determined from:

$$B = \frac{\mu_0 I}{2\pi r}$$

and in this case:

$$B_2 = \frac{\mu_0 I_2}{2\pi r}$$

so that the force per unit length between parallel currents:

$$\boldsymbol{\frac{F}{L} = \mu_0 \frac{I_1 I_2}{2\pi r}}$$

The same force acts on both currents.

This arrangement was, until recently, used to define the SI unit of current, the ampere. One ampere, 1 A, was defined as the current flowing in two infinitely long, straight, parallel wires that produced a force of exactly $2 \times 10^{-7}\,\mathrm{N\,m^{-1}}$ between the wires if they were exactly 1 m apart in a vacuum.

WORKED EXAMPLE D3.8

Two very long straight wires are 12 cm apart. One carries a current of 3.7 A, the other carries a current of 1.6 A in the opposite direction.

a Calculate the force exerted on 1.0 m of the 3.7 A current (magnitude and direction).

b State the force per metre acting on the other current.

Answer

a $\frac{F}{L} = \mu_0 \frac{I_1 I_2}{2\pi r} = \frac{((4\pi \times 10^{-7}) \times 3.7 \times 1.6)}{(2\pi \times 0.12)}$

$= 9.9 \times 10^{-6}$ N acting in a direction away from the other wire.

b The same: 9.9×10^{-6} N acting in a direction away from the other wire. That is, the two forces act in opposite directions.

16 Calculate the magnetic force per metre on a wire carrying a current of 1.2 A through a magnetic field of 7.2 mT if the angle between the wire and the field is:

a 30° **b** 60° **c** 90° **d** 0°.

17 a The Earth's magnetic field strength at a particular location has a horizontal component of 24 μT. Calculate the maximum force per metre that a horizontal cable carrying a direct current of 100 A could experience.

b State the direction in which the current needs to be flowing for this force to be vertically upwards.

c Discuss whether it is possible that such a force could support a cable.

18 A current is flowing in a horizontal wire perpendicularly across a magnetic field of strength 0.36 T. It experiences a force of 0.18 N, also horizontally.

a Draw a diagram to show the relative directions of the force, field and current.

b If the length of wire in the field is 16 cm, calculate the magnitude of the current.

19 a A current of 3.8 A in a long wire experiences a force of 5.7×10^{-3} N when it flows through a magnetic field of strength 25 mT. If the length of wire in the field is 10 cm, determine the angle between the field and the current.

b If the direction of the wire is changed so that it is perpendicular to the field, calculate the new force on the current.

20 Consider Figure D3.19. Figure D3.22 shows a side view of the same situation: the loop of wire and the magnetic poles.

The current in the loop of wire is 0.530 A, the horizontal magnetic field strength is 25.0 mT and the length of the right-hand side of loop in the magnetic field is 3.80 cm.

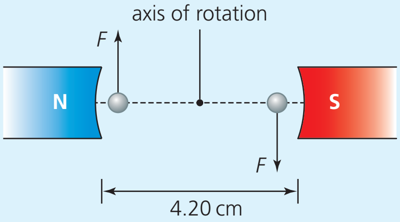


Figure D3.22 Forces between parallel currents; side view

a Determine the downwards force on the current in right-hand side of the loop.

b Calculate the torque applied to the loop by this force.

c What is the magnitude of the torque provided by the *couple* acting on the loop, and in what 'sense' is it acting?

d How will the magnitude of the torque change as the loop begins to rotate from its horizontal position (as shown)? Explain your answer.

e To make the loop rotate faster, the wire can be wound into a coil of many turns. Predict how many turns are needed to increase the maximum torque to 1.0×10^{-4} Nm.

21 Show that, when a current of 1.0 A flows in two infinitely long, straight, parallel wires, a force of exactly 2.0×10^{-7} N m^{-1} acts between them if they are exactly 1.0 m apart in a vacuum.

22 Two long straight wires are placed parallel to each other and 2.0 cm apart. One wire carries a current of 1.8 A.

a Determine what current in the other wire will result in a force of 4.7×10^{-5} N m^{-1} acting on it.

b State the magnitude of the force per metre on the other wire.

c If the currents are in opposite directions, in which directions will the forces act?

D.4 Induction

Guiding questions

- What are the effects of relative motion between a conductor and a magnetic field?
- How can the power output of electrical generators be increased?
- How did the discovery of electromagnetic induction effect industrialization?

Electromagnetic induction

As before, the word *induction* is being used to describe something being made to happen without physical contact. Previously, in Topic D.2, we have discussed *electrostatic* induction and *magnetic* induction.

◆ **Electromagnetic induction** Process in which an emf is produced across a conductor that is experiencing a changing magnetic field.

Whenever a conductor moves across a magnetic field, or a magnetic field moves across a conductor, an emf will be induced. This effect is called **electromagnetic induction**.

Reminder from Topic B.5: the *electromotive force* (emf) of a battery, or any other source of electrical energy, is defined as the total energy transferred in the source per unit charge passing through it. In simple terms, it is the potential difference across the source when there is no current flowing.

There are numerous important applications of electromagnetic induction, including:

- generating electricity
- transforming voltages
- using bank cards and smart cards
- metal detecting and security checks
- regenerative braking.

All examples of electromagnetic induction are produced by one of the following.

- A conductor moves across a permanent magnetic field.
- A permanent magnetic field is moved across a conductor.
- A changing current in a circuit produces a changing magnetic field which passes through a *separate* conductor (without any physical movement).
- A changing current in a circuit produces a changing magnetic field which passes through the *same* circuit (without any physical movement).

We will describe each of these in the rest of this topic.

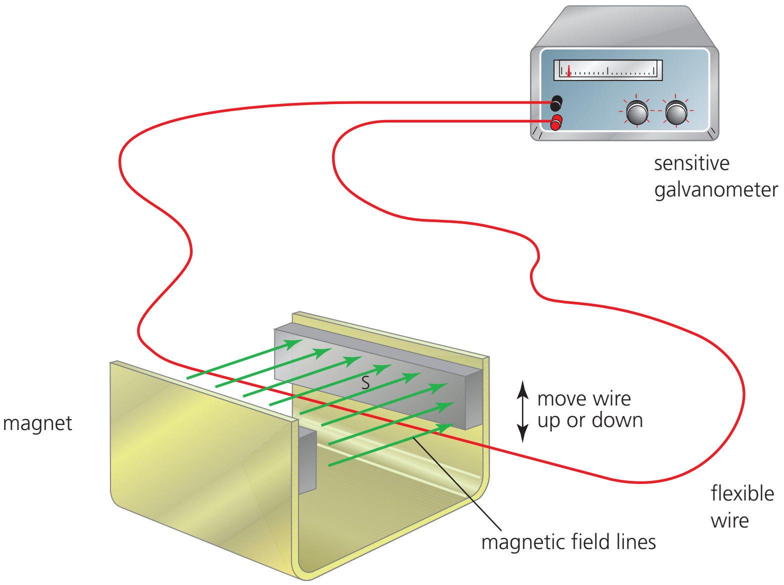
Electromagnetic induction by a conductor moving across a permanent magnetic field

SYLLABUS CONTENT

- ▶ A uniform magnetic field induces an emf in a straight conductor moving perpendicularly to it, as given by: $\varepsilon = BvL$.

◆ **Galvanometer** Ammeter that measures very small currents.

Figure D4.1 shows an experiment in which an emf is induced when a conductor (a metal wire) is moved across a permanent magnetic field. The induced emf can be detected because it makes a small current flow through a circuit containing a sensitive ammeter, called a **galvanometer**.

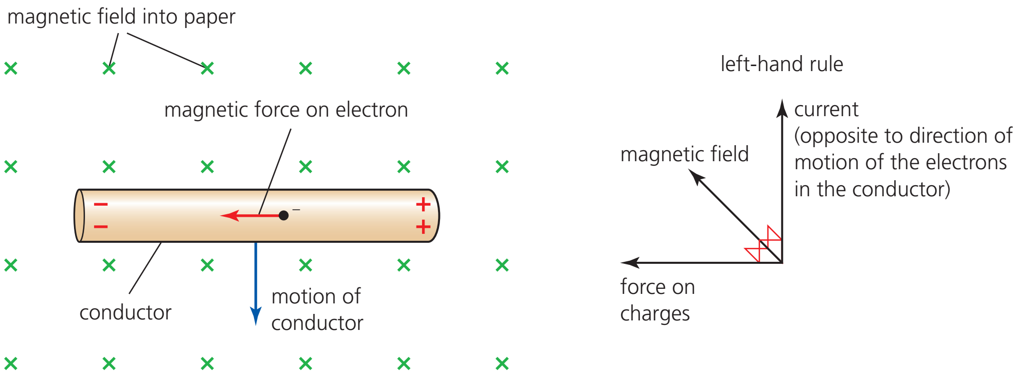


■ **Figure D4.1** Inducing an emf by moving a wire up or down across the magnetic field

The charged particles in the conductor experience forces because they are moving with the wire as it crosses the magnetic field (as discussed in Topic D.3). Because it is a conductor, the wire contains free electrons that can move along the wire under the action of these forces. Other charges (protons and most of the electrons) also experience forces but are not able to move along the conductor.

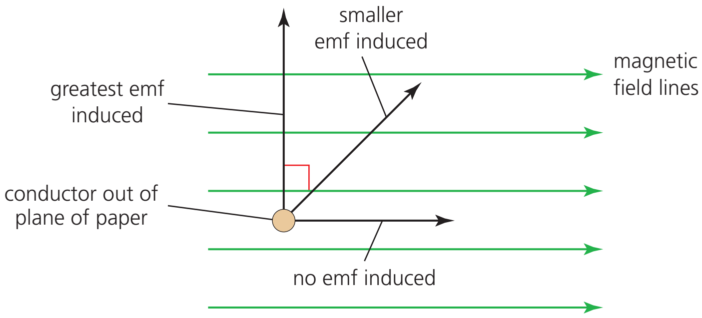
An emf is induced in a conductor because free electrons experience forces which make them move along the wire as it crosses a magnetic field.

Moving the wire, containing free electrons, is equivalent to a conventional current of positive charge in the opposite direction. We can use Fleming's left-hand rule (Topic D.3) to predict the direction of the forces on the electrons, as shown on the right in Figure D4.2. In this case the magnetic force pushes the electrons to the left, so the left-hand end of the conductor becomes negatively charged, while the other end becomes positively charged (because some electrons have flowed the other way). This charge separation produces a potential difference (emf) across the ends of the conductor.



■ **Figure D4.2** Magnetic force on electrons produces charge separation

If the motion or the magnetic field is reversed in direction, then the emf is also reversed. If both the motion and the magnetic field are reversed, then the direction of the emf is unchanged. If the conductor and the magnetic field are both moving, but with the same velocity, no emf is induced. For electromagnetic induction to occur there must be *relative* motion between the conductor and the magnetic field.



■ **Figure D4.3** The size of an induced emf depends on the direction of motion

In order to induce an emf, a conductor needs to move *across* a magnetic field. Magnetic fields are represented by field *lines* and the conductor needs to be moving so that it 'cuts' across (through) the field lines. There will be no induced emf if the conductor is moving in a direction that is parallel to the magnetic field lines. Consider Figure D4.3, which shows three possible movements of a straight conductor which remains perpendicular to the plane of the paper. For similar conductors moving at the same speed, the induced emf is highest if the motion is perpendicular to the magnetic field.

If the conductor shown in Figure D4.3 is moved in a direction which is parallel to its own axis, it will not cut field lines and no emf will be induced. If it is rotated in the position shown, emfs will be induced unless the plane of rotation is parallel to the magnetic field.

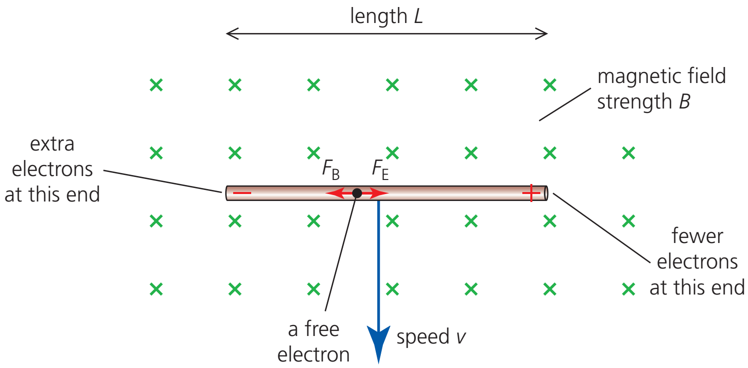
Experiments with the apparatus seen in Figure D4.1 can demonstrate that, for a conductor moving perpendicularly across the field, the emf, ε, induced can be increased by:

- increasing the speed of the movement, v
- using a magnetic field of greater strength, B
- increasing the length of the conductor in the magnetic field, L (which may mean increasing the extent of the magnetic field)
- winding the wire into a coil of many turns, N (with one side of the coil inside the magnetic field).

By considering the forces on free electrons, we can derive an equation for the emf, as follows.

Equation for an induced emf

Figure D4.4 shows a closer look at the situation seen in Figure D4.2. A conductor of length L is moving perpendicularly across a uniform magnetic field of strength B, with speed v. Free electrons in the conductor will each experience a magnetic force, F_B, given by the expression, $F = qvB\sin\theta$ (Topic D.3). In this perpendicular arrangement $\sin\theta = 1$. These forces tend to move free electrons towards the left of the conductor (as shown). As more electrons move along the conductor, the increasing amount of negative charge repels the motion of further electrons to that end. The right-hand end of the conductor, which has lost electrons, will become positively charged and act as an attractive force on the electrons.

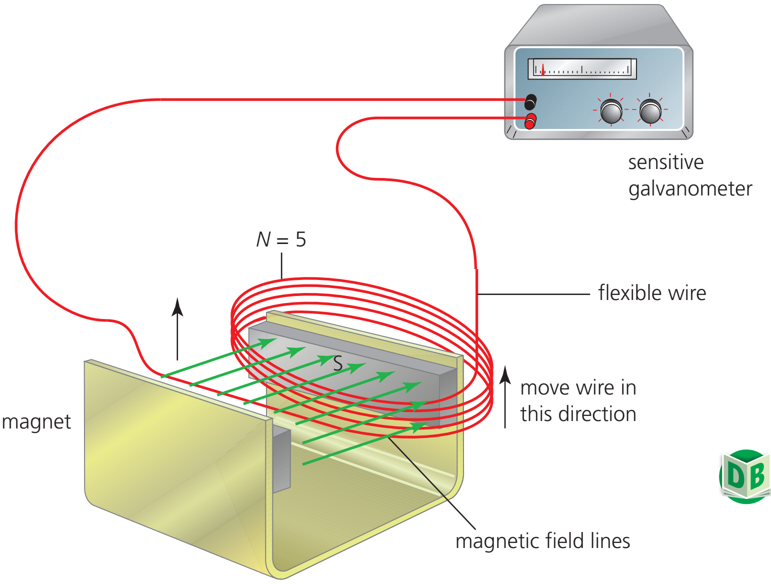


■ **Figure D4.4** Deriving $\varepsilon = BvL$

The charge separation produces an electric field along the conductor:

$$E = \frac{\varepsilon}{L}$$

where ε is the induced emf across the ends of the conductor.



■ **Figure D4.5** Electromagnetic induction with more than one turn

The maximum induced potential difference will occur when the force on each free electron due to the magnetic field, F_B, is equal and opposite to the force on the electron, F_E, due to the electric field.

$$\text{electric force, } F_E = \text{electric field} \times \text{charge} = \frac{\varepsilon q}{L}$$

At equilibrium, $F_E = F_B$

$$\frac{\varepsilon q}{L} = qvB$$

So that:

the induced emf when a straight conductor moves perpendicularly across a uniform magnetic field:

$$\boldsymbol{\varepsilon = BvL}$$

(If the field was not perpendicular to the wire, the component of the field in that direction would have to be used in the calculation.)

If a long length of wire is wound into a loose coil and one side (only) is moved in the magnetic field, as shown in Figure D4.5, each extra loop of wire will add an emf of the same value in series, similar to adding more cells to a battery. If there are N turns in the coil, the induced emf will become $\varepsilon = NBvL$.

The directions of the induced emf and current are important and will be discussed later.

WORKED EXAMPLE D4.1

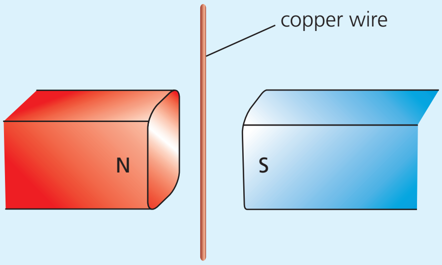
Calculate the induced emf produced across a 23.0 cm long conductor moving at 98.0 cm s^{-1} perpendicularly across a magnetic field of strength 120 µT.

Answer

$\varepsilon = BvL = (120 \times 10^{-6}) \times 0.98 \times 0.23 = 2.7 \times 10^{-5}\,\text{V}$

1 Explain why no emf is induced across a string made of plastic when it is moved through a magnetic field.

2 Figure D4.6 shows a copper wire between the poles of a permanent magnet. Describe the direction(s) in which the wire should be moved to induce:



■ **Figure D4.6** A copper wire between the poles of a permanent magnet

- a the highest emf
- b zero emf.
- c Explain why no current can be induced in this wire as shown.

3 When a straight conductor of length 90 cm moved perpendicularly across a uniform magnetic field of strength 4.5×10^{-4} T, an emf of 0.14 mV was induced. Calculate the speed of the conductor.

4 Determine the strength of magnetic field needed for a voltage of 0.12 V to be induced when a conductor of length 1.6 m moves perpendicularly across it at a speed of 2.7 m s^{-1}.

5 Show that the units of BvL are the same as for ε.

6 Consider Figure D4.5. Calculate the effective width of the uniform magnetic field if it has a strength of 7.8×10^{-3} T and an emf of 3.8 mV is induced when the side of the coil moves vertically with a speed of 1.8 m s^{-1}.

7 An aircraft is flying horizontally at a speed of $280\,\text{m}\,\text{s}^{-1}$ at a place where the vertical component of the Earth's magnetic field is $12\,\mu\text{T}$, as shown in Figure D4.7.

a Calculate the emf induced across its wing tips if its wingspan is 58 m.

b Suggest a possible reason why this voltage would be larger if the aircraft was flying close to the North, or South, Pole.

c Could the induced emf be used to do anything useful on the aircraft? Explain your answer.

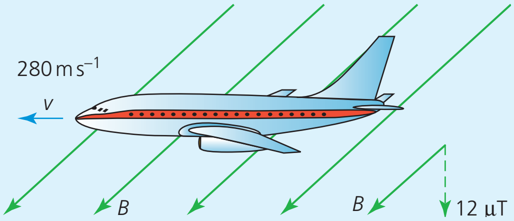


Figure D4.7 An aircraft flying horizontally

Electromagnetic induction by moving a permanent magnetic field across a conductor

Moving a conductor through a permanent magnetic field has a similar effect to keeping the conductor still and moving the field.

Figure D4.8 shows electromagnetic induction by moving a magnetic field (around a permanent magnet) through a conductor in the form of a loop of wire. Again, the induced emf and current will be very small in this basic example, but Figure D4.9 shows how the effects can be increased greatly by winding the conductor into a coil, or solenoid, with many turns. The direction of the induced current around the coil will be reversed if the magnet is reversed, or alternatively, if the motion of the magnet is reversed.

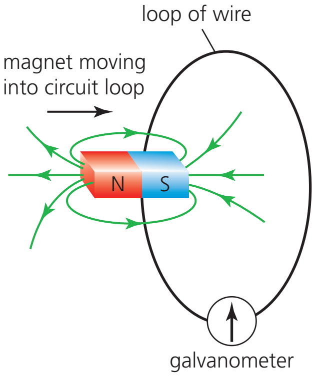


Figure D4.8 Moving a magnet to induce an emf and a current

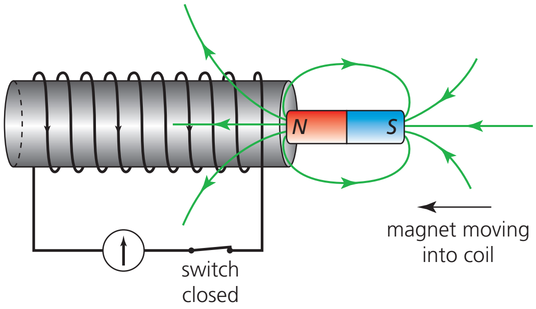


Figure D4.9 Inducing an emf and a current in a coil of wire

Figure D4.10 shows an electromagnetic induction experiment recorded on a data logger and computer. The data logger records the emf being induced at regular time intervals when a magnet is dropped through a coil, and then the data is used to draw a graph.

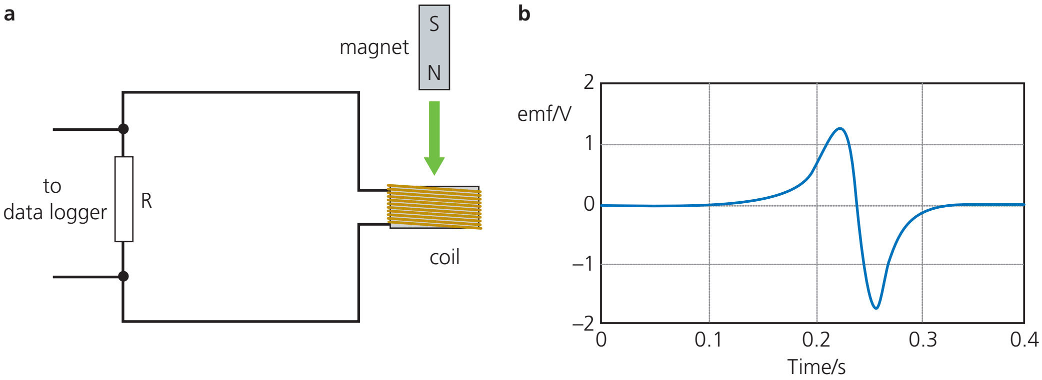


Figure D4.10 Inducing a current by dropping a magnet through a coil

WORKED EXAMPLE D4.2

Consider Figure D4.10. Describe how the graph would change if the:

a polarity of the magnet was reversed

b magnet was dropped from a greater height?

Answer

a The graph would be inverted.

b The peaks would be higher and their durations reduced.

8 In a demonstration of electromagnetic induction similar to that shown in Figure D4.8, the induced current was very small.

  a Suggest two ways of increasing the induced current while still using the same single loop of wire.

  b State two ways in which the current can be made to flow in the opposite direction around the circuit.

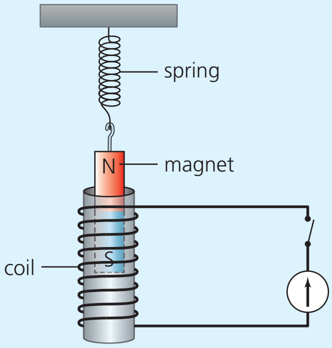
9 Draw a sketch similar to Figure D4.9 to show the current direction when the bar magnet comes out of the coil at the other end.

10 Suggest explanations for the shape of the graph shown in Figure D4.10b.

11 Figure D4.11 shows a magnet oscillating vertically on a spring. As it oscillates, with a frequency of 0.67 Hz, the end of the magnet passes in and out of a coil of wire which is in a circuit with a centre-reading galvanometer and a switch.

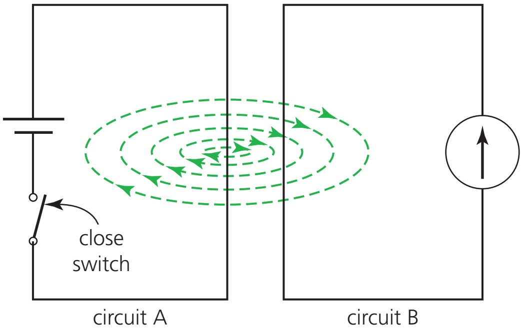
  a Describe how the pointer on the galvanometer will move while the switch is closed.

  b Sketch a graph of the induced current–time for 3.0 s.



■ **Figure D4.11** A magnet oscillating vertically on a spring

Electromagnetic induction without physical movement



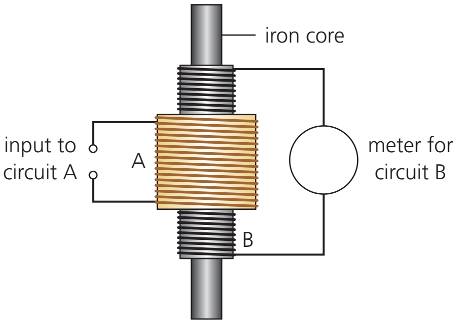
■ **Figure D4.12** When the switch is closed, a magnetic field passes from circuit A to circuit B

Emfs can also be induced, not by movement, but by changes in the current in one circuit affecting another, completely separate, circuit. Figure D4.12 represents the simplest example.

First consider circuit B at a time when the switch in circuit A is open – there is no power source and no changing magnetic field near B, so there is no current shown by the galvanometer. However, at the moment that the switch in circuit A is closed, a current starts to flow around circuit A and this sets up a magnetic field around it. This field spreads out and passes through circuit B.

The sudden *change* of magnetic field induces an emf and a current that is detected by the galvanometer in circuit B. The changing current produces a changing magnetic field in the same way as moving a magnet does.

This induced emf / current only lasts for a moment, while the switch in A is being turned on, because when the current in A is constant there is no *changing* magnetic field. When the switch is turned off, there is an induced emf / current for a moment in the opposite direction. As described so far, this is a very small (but important) effect. However, the induced emf can be increased greatly by winding the conducting wires in both circuits into coils of many turns (to increase the strength of the magnetic field) and placing them on top of each other with an iron core through the middle. This is shown in Figure D4.13. Remember from Topic D.2 that iron has high magnetic permeability and greatly increases the strength of the magnetic field.



■ **Figure D4.13** Making the induced emf larger by using an iron core and coils of many turns

◆ **Transformer** A device that transfers electrical energy from one circuit to another using electromagnetic induction between coils wound on an iron core. Transformers are used widely to transform one alternating voltage to another of different magnitude.

If, when a steady direct current is flowing around circuit A, it is suddenly switched off, the change in the magnetic field through circuit B can be so quick that a very large voltage can be momentarily induced if the coil in circuit B has a large number of turns. Used in this way, an *induction coil* can be both useful and dangerous.

If the voltage source in circuit A is changed from one that provides a *direct* current (dc) of constant value to a source of *alternating* current (ac), then the magnetic field in both circuits will change continuously and an alternating emf will be induced continuously. This has many useful applications, including **transformers**, as discussed below.

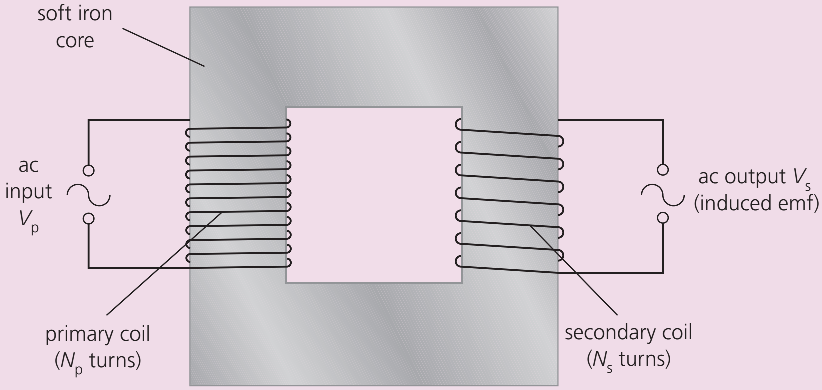
A changing current produces a changing magnetic field which can induce an emf without any physical movement. With alternating currents this effect is continuous.

Inquiry 1: Exploring and designing

Designing

Transformers

Figure D4.14 shows the basic components of a device known as a transformer. It functions in a similar way to the coils seen in Figure D4.13, where the coils are wound together. Transformers are used to change alternating voltages to lower, or higher, levels.



■ **Figure D4.14** Transformer

The alternating current in the *primary coil* creates a constantly changing magnetic field. The field is concentrated in the iron core and passes around to the *secondary coil*, where it induces an emf.

1 List the factors that will affect the value of the induced emf seen on the meter.

2 Design an experiment methodology (using two self-made coils) to investigate how the value of the induced emf depends on one of those variables. Pilot (try out) your design with the test coils that you have constructed.

■ **Figure D4.15** A transformer on a road-side pole

12 Consider Figure D4.12. Suggest two ways in which an emf momentarily induced in circuit B could be increased, without twisting the wires into coils.

13 Use a sketch graph to explain why an alternating current (which varies between the same maximum and minimum values) will induce a greater emf in a surrounding circuit when the frequency is greater.

14 See Figure D4.14. The output voltage, V_s, from a transformer can be calculated from:

$$V_s = V_p \times \left(\frac{N_s}{N_p}\right)$$

(You are not expected to remember this equation.)

- **a** Calculate the output voltage from a transformer which has an input of 230 V (ac), 350 turns on its primary coil and 18 turns on its secondary coil.
- **b** Another transformer is used to 'step-up' an alternating voltage from 50 V to 2000 V. If the primary coil has 40 turns, predict how many turns are needed for the secondary coil.

15 Explain why the kind of transformer seen in Figure D4.14 cannot transform steady voltages.

16 Suggest how induction between circuits is used in the operation of bank cards and transport cards. (See Figure D4.16 for an example.)

Figure D4.16 Using a transport card

◆ **Transmission of electrical power**
Electrical power is sent (transmitted) from power stations to different places around a country along wires (cables), which are commonly called transmission (or power) lines. These lines are linked together in an overall system called the transmission grid.

ATL D4A: Research skills

Evaluate information sources for accuracy, bias, credibility and relevance; use a single standard method of referencing and citation

A typical power station may produce electricity at a few hundred volts. Research and write a short report explaining the reasons why:

- Transformers are used to greatly increase this voltage before it is **transmitted** around the country.
- The currents are sent through aluminium cables.

Ensure you use reliable sources of information by carrying out credibility checks. In your report, be sure to provide clear references to the sources you used using the referencing and citation standard advocated by your school.

Figure D4.17 Transmission lines transfer electrical power around countries

Magnetic flux and magnetic flux linkage

SYLLABUS CONTENT

- ▶ Magnetic flux Φ as given by: $\Phi = BA\cos\theta$.

Magnetic flux

Nature of science: Models

Flux

The concept of *flux* has many applications. In general, the term is used to describe some kind of flow. A non-scientific example could be the (in)flux of people into a particular location, which could be recorded in terms of the number of people in a certain time. In physics we may refer to a flux of thermal energy, or light, or radiation, each of which could be measured in terms of the amount of energy flowing through a given area every second ($\mathrm{W\,m^{-2}}$). See Question 20.

Magnetic flux, as explained below, is slightly different, because no movement is implied, although the vector arrows seen, for example, in Figure D4.18, may suggest motion.

◆ **Magnetic flux, Φ** Defined as the product of an area, A, and the component of the magnetic field strength perpendicular to that area.

◆ **Weber, Wb** Unit of magnetic flux. $1\,\mathrm{Wb} = 1\,\mathrm{T\,m^2}$.

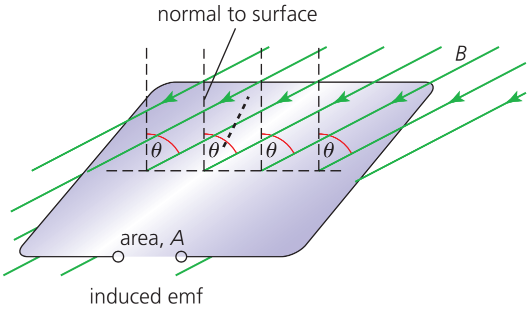
◆ **Magnetic flux density, B** The term more commonly used at Higher Level for magnetic field strength. $B = \frac{\Phi}{A}$.

Electromagnetic induction becomes easier to understand after the concept of **magnetic flux** has been introduced.

Suppose we want to induce an emf across a loop of wire. There are a number of possibilities, including:

- move the loop into, or out of, a permanent magnetic field
- rotate the loop in a permanent magnetic field
- keep the loop still and move a permanent magnetic field into, or out of, the loop
- keep the loop still in the changing magnetic field produced by a changing current in another circuit
- any combination of the above.

We will simplify the geometry of these situations to that shown in Figure D4.18, in which a magnetic field is acting into a loop of wire. To simplify the diagram, only a few field lines are seen, but we will assume that a uniform magnetic field is acting across the whole area of the loop. An emf can be induced by any of the changes listed above.



■ **Figure D4.18** Magnetic flux depends on field strength, area and angle

The size of the induced emf depends not only on the strength of the magnetic field, B, but also on the area, A, of the circuit over which it is acting, and the angle, θ , at which it is passing through the circuit.

Magnetic flux, Φ, (for a uniform magnetic field) is defined as the product of the area, A, and the component of the magnetic field strength which is perpendicular to that area:

$$\Phi = BA\cos\theta$$

If the field is perpendicular to the area, $\cos\theta = 1$ so the equation reduces to: $\Phi = BA$.

Common mistake

Note that the angle θ is the angle between the field and a normal to the surface, not the angle between the field and the surface.

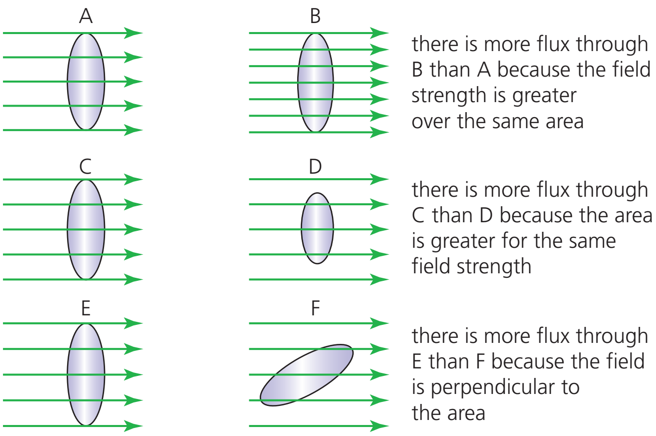
The SI unit of magnetic flux is the **Weber**, Wb. One weber is equal to one tesla multiplied by one metre squared ($1\,\mathrm{Wb} = 1\,\mathrm{T\,m^2}$).

We can rearrange the equation for flux to give:

$$B = \frac{\Phi}{A}$$

for B perpendicular to A; this shows us why magnetic field strength is widely known as **magnetic flux density** (flux / area). That is, 1 tesla = 1 weber per square metre.

The emf induced across a single loop of wire is proportional to the rate of change of magnetic flux through it (more details later).



■ **Figure D4.19** Magnetic flux explained in terms of field lines

Magnetic flux can be a difficult concept to understand and Figure D4.19 may help. It shows a non-mathematical interpretation of magnetic flux as the number of magnetic field lines that pass through the area.

It may be helpful to consider that the magnitude of an induced emf depends on the rate at which a conductor 'cuts' magnetic field lines (or the rate at which magnetic field lines cut a conductor).

Magnetic flux linkage

So far, we have been discussing electromagnetic induction using single loops of wire. But, if the wire is wound into a coil with N turns, each turn contributes the same emf (in series), so that the overall induced emf is multiplied by N. The concept of **magnetic flux linkage** takes this into account:

◆ **Magnetic flux linkage, $N\Phi$** The product of magnetic flux and the number of turns in a circuit (unit: Wb).

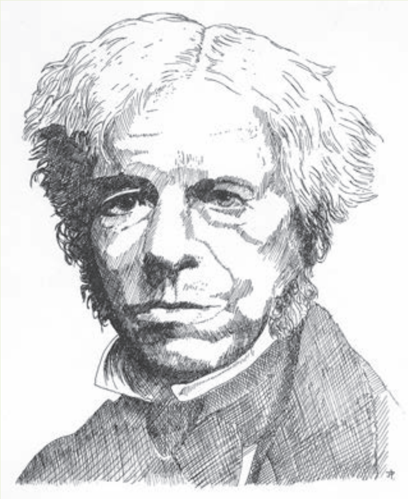
Magnetic flux linkage is defined as the product of magnetic flux and the number of turns in the coil. It does not have a widely used standard symbol:

$$\text{magnetic flux linkage} = N\Phi$$

The units of flux linkage are the same as flux (Wb), although sometimes Wb-turns is used.

Nature of science: Observations

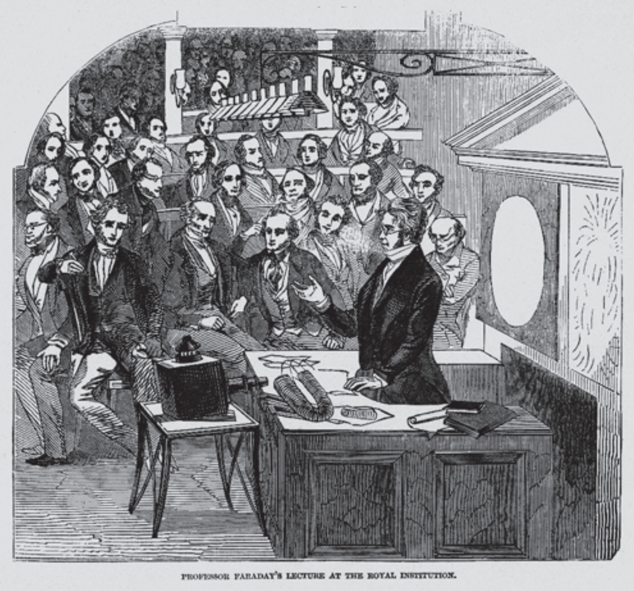
Unexpected or unplanned observations



■ **Figure D4.20** Michael Faraday (1791–1867) is considered to be one of the greatest scientists

In 1831 Michael Faraday (Figure D4.20) became the first person to demonstrate electromagnetic induction. See Figure D4.21.

The equipment available at the time made this a difficult phenomenon to observe, and some observers may have doubted its importance at the time, but its far-reaching consequences are now undeniable. In a similar way, the first transmission of radio waves more than 50 years later (by Heinrich Hertz) may have seemed trivial at the time, but both discoveries ultimately changed the world for ever.



■ **Figure D4.21** Michael Faraday giving a lecture at the Royal Institution in London

WORKED EXAMPLE D4.3

a Calculate the magnetic flux in a square loop of wire of sides 6.2 cm when it is placed at 45° to a magnetic flux density of 4.3×10^{-4} T.

b Calculate how many turns would be needed on a coil of the same dimensions to create a flux linkage of 8.4×10^{-4} Wb.

c Determine the magnetic flux linkage in the coil if only half of it was in the magnetic field.

Answer

a $\Phi = BA\cos\theta = (4.3 \times 10^{-4}) \times (6.2 \times 10^{-2})^2 \times \cos 45° = 1.2 \times 10^{-6}$ Wb

(1.16879… × 10^{-6} seen on the calculator display)

b $N\Phi = 8.4 \times 10^{-4}$

$$N = \frac{8.4 \times 10^{-4}}{1.16879 \times 10^{-6}} = 7.2 \times 10^2$$

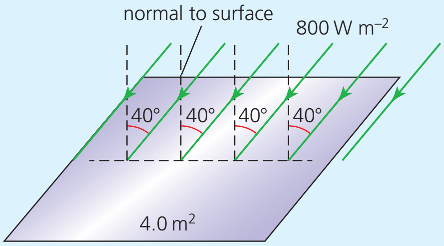
c The magnetic flux linkage would be reduced to half (4.2×10^{-4} Wb) because the area used in the calculation is the area of the coil in the magnetic field, not the total area of the coil.

17 Calculate the magnetic flux in a flat coil of area 48 cm^2 placed in a field of magnetic flux density 5.3×10^{-3} T if the field is at an angle of 30° to the plane of the coil.

18 A magnetic field of strength 3.4×10^{-2} T passes perpendicularly through a flat coil of 480 turns and area 4.4×10^{-5} m^2.
Determine the flux linkage.

19 A flat coil of 600 turns and area 8.7 cm^2 is placed where the magnetic flux density is 9.1×10^{-3} T. The axis of the coil was originally parallel to the magnetic field, but it was then rotated by 25°.
Calculate the *change* of flux linkage through the coil.

20 This question provides a solar radiation analogy to help understanding of the concept of flux. The 'solar flux density' arriving perpendicularly at the Earth's upper atmosphere is 1360 W m^{-2}. (This was called the Solar constant in Topic B.2.)
Suppose that near the Earth's surface this value has reduced to 800 W m^{-2}.
Calculate the power arriving at a horizontal solar panel of area 4.0 m^2 if the radiation arrives at an angle of 40° to the vertical (see Figure D4.22).



■ **Figure D4.22** Solar flux

Faraday's law of electromagnetic induction

SYLLABUS CONTENT

▶ A time-changing magnetic flux induces an emf ε as given by Faraday's law of induction: $\varepsilon = -N\frac{\Delta\Phi}{\Delta t}$.

We can now write down an equation which can be used to determine the value of an emf induced under *any* circumstances:

If a coil with N turns experiences a magnetic flux which changes by $\Delta\Phi$ in time Δt, the induced emf:

$$\boldsymbol{\varepsilon = -N\frac{\Delta\Phi}{\Delta t}}$$

◆ Faraday's law of electromagnetic induction The magnitude of an induced emf is equal to the rate of change of magnetic flux linkage, $\varepsilon = \frac{-N\Delta\Phi}{\Delta t}$. For an explanation of the negative sign, see Lenz's law.

The negative sign here is important and it is explained later in this topic.

We will now explain how **Faraday's law** can be applied to three different situations.

Induction because of motion of a conductor across a uniform magnetic field

Figure D4.23 shows a simple visualization of one type of electromagnetic induction: a metal rod is lying across two fixed parallel conducting rails. The rod is able to move horizontally which is perpendicularly across the magnetic field, but as it does so, it continues to complete an electrical circuit, as shown in the figure.

In this example of electromagnetic induction, the magnetic field is constant but the area of the circuit changes.

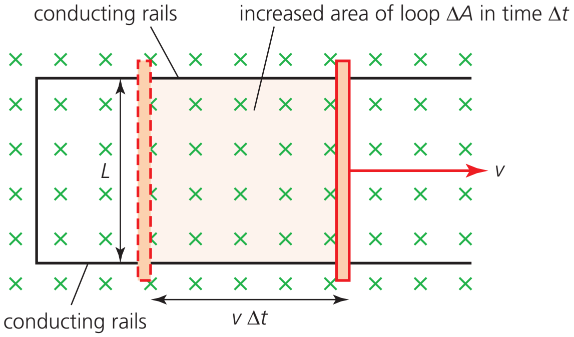


Figure D4.23 Inducing an emf with a moving conductor

Because there is only one loop, Faraday's law for the magnitude of the induced emf reduces to:

$$\varepsilon = \frac{\Delta\Phi}{\Delta t} = \frac{\Delta(BA)}{\Delta t} = B\left(\frac{\Delta A}{\Delta t}\right)$$

for a uniform magnetic field.

Suppose that the rod, of length L, moves to the right with a constant speed v. In time Δt it will move a distance, $v\Delta t$ perpendicularly across a uniform magnetic field of strength B.

The rate of change of area:

$$\frac{\Delta A}{\Delta t} = \frac{Lv\Delta t}{\Delta t} = vL$$

vL is often described as the 'area swept out' by the moving rod in time Δt; so that:

$\varepsilon = B\left(\frac{\Delta A}{\Delta t}\right)$ becomes:

$\varepsilon = BvL$ which is the same equation as we have seen earlier.

WORKED EXAMPLE D4.4

Two parallel and horizontal conducting rails, which are 44 cm apart, are placed in a uniform magnetic flux density of 8.7×10^{-4} T which is acting vertically downwards, as shown in Figure D4.23. The rod moves to the right with a speed of 48 cm s^{-1}.

a Determine the value of the emf induced across the loop.

b State the rate of change of magnetic flux.

c Determine how much extra magnetic flux passes through the circuit when the rod moves 25 cm.

Answer

a $\varepsilon = BvL = (8.7 \times 10^{-4}) \times 0.48 \times 0.44 = 1.8 \times 10^{-4}$ V

(1.83744×10^{-4} V seen on calculator display)

b 1.8×10^{-4} Wb s^{-1}

c Increase in area in $0.25 \times 0.44 = 0.11$ m^2

increase in magnetic flux = increase in area × magnetic flux density

$= 0.11 \times (8.7 \times 10^{-4}) = 9.6 \times 10^{-5}$ Wb

Alternatively:

increase in magnetic flux = rate of change of magnetic flux × time

$= (1.83744 \times 10^{-4}) \times \left(\frac{0.25}{0.48}\right) = 9.6 \times 10^{-5}$ Wb

Tool 3: Mathematics

Interpret areas under graphs

Consider again Figure D4.10b. What do the two areas between the curve and the horizontal axis represent? Explain why they are equal in magnitude.

Figure D4.24 Coil moving into a magnetic field

Induction because of motion of a coil into and out of a uniform magnetic field

Figure D4.24 shows a coil of wire being moved into a perpendicular magnetic field. As the right-hand side of the coil enters the field, an emf is induced and a current flows around the coil as shown. When all of the coil is moving inside the field, there will be no changing magnetic flux and so there is no induced emf or current. When the coil moves out of the right-hand side of the field, the emf and current are reversed from their directions when entering the field.

In this example of electromagnetic induction, the area (of the coil) is constant but the magnetic field through the coil changes. The same effect can also be produced by keeping the coil stationary and moving the magnetic field. Then, the magnitude of the induced emf given by

$$\varepsilon = N\frac{\Delta\Phi}{\Delta t}$$

becomes:

$$\varepsilon = NA \times \frac{\Delta B}{\Delta t}$$

WORKED EXAMPLE D4.5

Consider Figure D4.24.

a Determine the average magnitude of the induced emf when a coil of 40 turns and area 5.0 cm^2 is moved from completely outside to completely inside a uniform magnetic field of strength 0.34 T in 0.56 s.

b The coil is then turned upside-down (rotated 180°) in the same magnetic field, in a time of 0.29 s. Calculate the magnitude of the induced emf.

Answer

a $\varepsilon = N\frac{\Delta\Phi}{\Delta t} = NA \times \frac{\Delta B}{\Delta t}$

$= 40 \times (5.0 \times 10^{-4}) \times \left(\frac{0.34}{0.56}\right)$

$= 1.2 \times 10^{-2}\,\text{V}$

b The field changes from 0.34 T in one direction through the coil to 0.34 T in the opposite direction. Which is an overall change of 0.68 T.

$\varepsilon = N\frac{\Delta\Phi}{\Delta t} = NA \times \frac{\Delta B}{\Delta t}$

$= 40 \times (5.0 \times 10^{-4}) \times \left(\frac{0.68}{0.29}\right)$

$= 4.7 \times 10^{-2}\,\text{V}$

21 A train is travelling with a speed of 38 m s^{-1} through a region where the Earth's magnetic field is 42 µT, acting at 50° to the horizontal. An axle on the train has a length of 1.43 m.
- **a** Calculate the area swept out by the axle every second.
- **b** Determine the component of the magnetic field acting perpendicular to the axle.
- **c** Calculate the rate of change of magnetic flux experienced by the axle.
- **d** What was the magnitude of the induced emf across the axle?

22 The magnetic flux through a coil of 1200 turns increases from zero to 4.8×10^{-5} Wb in 2.7 ms. Calculate the magnitude of the average induced emf during this time.

23 A coil of area 4.7 cm^2 and 480 turns is in a magnetic field of strength 3.9×10^{-2} T.
- **a** Calculate the maximum possible magnetic flux linkage through the coil.
- **b** Determine the average induced emf (mV) when the coil is moved to a place where the perpendicular magnetic field strength is 9.3×10^{-3} T in a time of 0.22 s.

Induction between circuits

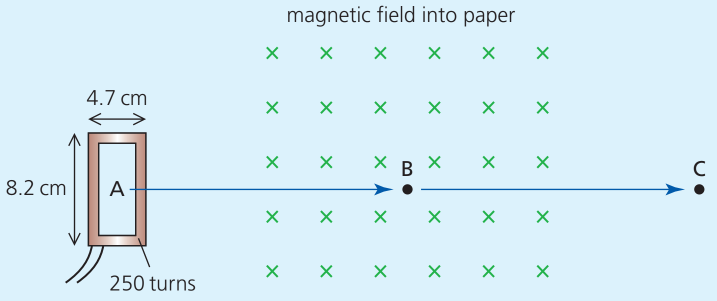
Reconsider Figure D4.13. Induction between circuits is called **mutual induction**. The changing magnetic flux passing through coil B depends on the rate at which the current, *I*, is changing in coil A.

◆ **Mutual induction** Electromagnetic induction between separate circuits.

For mutual induction in a fixed arrangement $\frac{\Delta \Phi}{\Delta t}$ and the induced emf, ε, are proportional to $\frac{\Delta I}{\Delta t}$.

You will *not* be expected to answer detailed quantitative questions on mutual induction.

24 Figure D4.25 shows a coil of 250 turns moving from position A, outside a strong uniform magnetic field of strength 0.12 T, to position B at the centre of the magnetic field in a time of 1.4 s.
- **a** Calculate the change of magnetic flux in the coil when it is moved.
- **b** State any assumption that you made in answering **a**.
- **c** Determine the change of magnetic flux linkage.
- **d** Calculate the average induced emf.
- **e** Sketch a graph to show how the induced emf changes as the coil is moved at constant speed from A to C (no values needed).



■ **Figure D4.25** A moving coil of 250 turns

25 Imagine you are holding a flat coil of wire in the Earth's magnetic field.
- **a** Draw a sketch to show how you would hold the coil so that there is no magnetic flux through it.
- **b** At a place where the magnitude of the Earth's magnetic field strength is 48 µT, what emf would be induced by moving a coil of 550 turns and area 17 cm^2 from being parallel to being perpendicular to the magnetic field in 0.50 s?

26 A small coil of area 1.2 cm^2 is placed in the centre of a long solenoid with a large cross-sectional area. A steady current of 0.50 A in the solenoid produces a magnetic field of strength 8.8×10^{-4} T.

a Determine how many turns would be needed on the coil if an induced emf of 2.4 mV was required when the current in the solenoid was increased to 2.0 A in a time of 0.10 s.

b Describe the position in which the coil needs to be placed.

ATL D4B: Thinking skills

Applying key ideas and facts in new contexts

Induction cookers

In an induction cooker, like that shown in Figure D4.26, there are coils of wire below the flat top surface. When a high-frequency current is passed through a coil in the cooker, a strong, rapidly oscillating magnetic field is created that will pass through anything placed on or near the cooker's surface, like a cooking pot. If the material of the pot is a conductor, emfs and currents will be induced in it. Currents circulating within solid conductors, rather than around wire circuits, are known as **eddy currents**.

The electrical energy in the eddy currents will be transferred to internal energy and the pan will get hotter. As in any circuit, the power generated can be calculated using $P = I^2R$. Thermal energy is then transferred by conduction to any food that is in the pan.

Apply your understanding of thermal physics and electromagnetic induction to explain the following:

- why the choice of metal for the pan is important
- why the cooker itself does not get noticeably hot (except for any thermal energy transferred from the pan)
- why induction cookers are more efficient at cooking food than conventional electric hot plates.

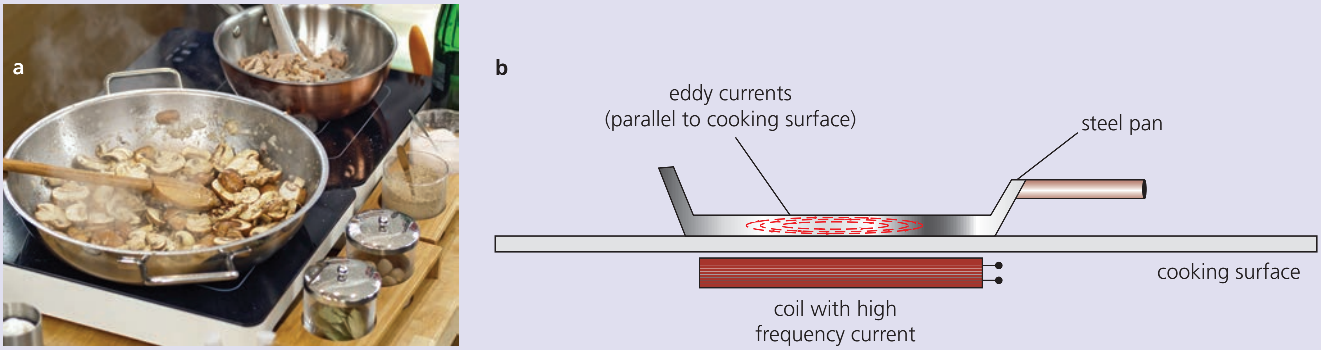


Figure D4.26 a A steel pan on an induction cooker and **b** How an induction cooker works

◆ Eddy currents Circulating currents induced in solid pieces of metal when changing magnetic fields pass through them.

Energy transfers during electromagnetic induction

SYLLABUS CONTENT

▶ The direction of an induced emf is determined by Lenz's law and is a consequence of energy conservation.

We will now explain why a negative sign appears in Faraday's law of electromagnetic induction.

If a current is generated from motion by electromagnetic induction, then energy must have been transferred from outside the circuit. We know this from an understanding of the *law of conservation of energy*. The origin of this energy is often the kinetic energy of the moving conductor or moving magnet. The moving object must therefore slow down as it loses some of its kinetic energy (unless there is an external force keeping it moving).

An induced electric current has had energy transferred to it from the process that induced it, for example from kinetic energy of motion.

Consider again Figure D4.11. To begin the experiment, a student pulled the magnet downwards, stretching the spring. The student supplied the energy. The magnet then oscillates vertically, interchanging elastic potential energy and kinetic energy (assuming that changes of gravitational potential energy are not significant). If there is no coil, or if the switch is open, the oscillations will continue for some time, although there will be a little energy dissipation (damping), as discussed in Topic C.2.

However, if the magnet oscillates into and out of a conducting coil, an emf will be induced across the coil because of the changing magnetic flux in it. A current will flow *if* the switch is closed and energy will be transferred in the coil. Energy is transferred to the current from the kinetic energy of the magnet, which means that the magnet must move more and more slowly.

Consider again Figure D4.9. A current is induced in the coil as the magnet moves towards it. That current makes the coil behave as an electromagnet, with one end a south magnetic pole and the other a north magnetic pole (Topic D.2). The induced emf across the coil is in the direction such that the induced current makes the right-hand side of the coil (as shown) a north pole. In this way there is a repulsive force between the magnet and the coil because two north poles are close together. Work has to be done to overcome this force and move the magnet into the coil. When this is done, the energy is transferred to the current in the coil.

If the motion of the magnet is reversed, an attractive force will be created and, again, work has to be done to move the magnet and induce a current in the coil.

If the magnet was already in the coil and then removed from the left-hand side, the induced current would set up a magnetic field to oppose that motion.

◆ **Lenz's law (of electromagnetic induction)** The direction of an induced emf is such that it will oppose the change that produced it. This is represented mathematically by the negative sign in the equation representing Faraday's law.

Whenever the magnet is moved in any way (in Figure D4.9) a current will be induced and the magnetic field of that current will tend to stop the movement. This application of the law of conservation of energy is known as **Lenz's law** and it is the reason why there is a negative sign in Faraday's law:

$$\varepsilon = -N\frac{\Delta\Phi}{\Delta t}$$

If the switch in Figure D4.9 is opened, there will still be an induced emf, but no current can flow. This means that there will be no magnetic field created and no force from the coil.

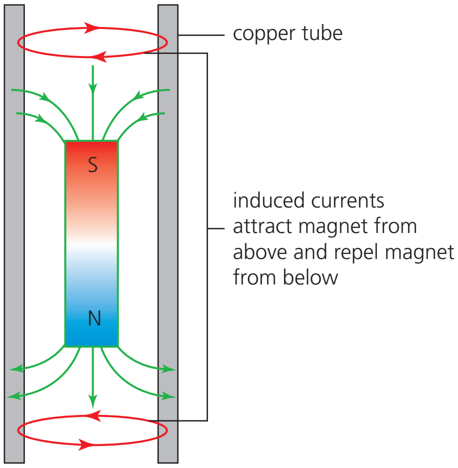
Lenz's law

Lenz's law states that the direction of any induced emf (and current) is always such that it will oppose the change that produced it.

That is, an induced emf will be in such a direction that any induced current will set up a magnetic field that resists the change.

Dropping a bar magnet through a vertical copper tube makes an interesting demonstration of Lenz's law. See Figure D4.27. The physics involved in this demonstration is similar to that shown in Figure D4.10.

As the magnetic field surrounding the falling magnet cuts through the copper tube, eddy currents are created. These currents produce their own magnetic fields which oppose the motion of the falling magnet. As a result, the magnet takes a surprisingly long time to reach the bottom of the tube.



■ **Figure D4.27** Magnet falling through a copper tube

27 Explain what positive and negative values of emfs represent.

28 Figure D4.28 shows what happens when a bar magnet is being moved away from a solenoid connected in a complete circuit. A galvanometer shows that a current *I* is flowing at that moment.

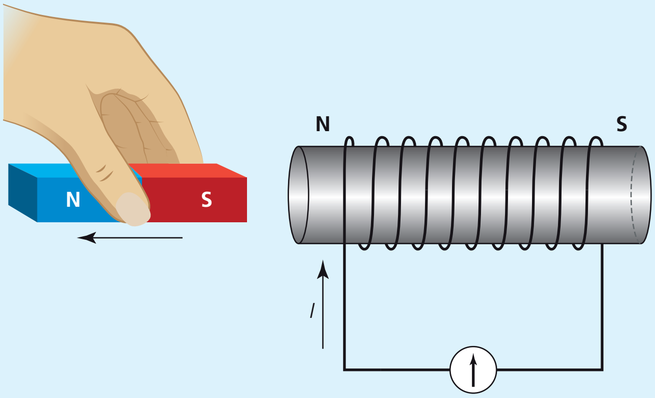


Figure D4.28 What happens when a bar magnet is being moved away from a solenoid connected in a complete circuit?

a Draw a similar diagram to represent what happens when the magnet is moved towards the solenoid.

b Under these circumstances, the coil is acting like an electromagnet. Make up rules to help another student predict how the polarity of the electromagnet is related to the motion of the magnet and the direction of the current.

29 Consider again a magnet falling through a metal tube (Figure D4.27).
Describe and explain what difference it would make if the copper tube was replaced by an aluminium tube of similar dimensions. (Aluminium has a higher resistivity than copper.)

30 Figure D4.29 shows an aluminium plate swinging as a pendulum through a magnetic field.
Sketch an appropriate graph to represent three complete oscillations.

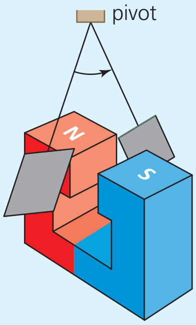


Figure D4.29 An aluminium plate swinging as a pendulum

ATL D4C RESEARCH SKILLS

Research into the uses and advantages of electromagnetic induction in ‘regenerative braking’.

Top tip!

In Question 30 there is no quantitative data provided, so there is no need to include any numbers on the sketch. The quantities being represented should be shown on the axes, preferably in words, although standard symbols are acceptable. Any important features of the graph should be labelled.

LINKING QUESTION

- Faraday’s law of induction includes a rate of change. Which other areas of physics relate to rates of change? (NOS)

TOK

The natural sciences

- Does the precision of the language used in the natural sciences successfully eliminate all ambiguity?

The terminology used in physics can often confuse people who have not studied the subject. The theory of electromagnetic induction is a good example. Does the use of specialized terminology make communicating scientific concepts to the public more difficult?

Faraday’s law states that ‘an induced emf is equal to the rate of change of magnetic flux linkage’. You should appreciate that this is an elegant and precise way of expressing a very important concept. However, to many non-scientists it may seem like a foreign language. Scientists aim to express ideas as briefly and as succinctly as possible, especially when communicating with other scientists. This involves the use of precise scientific terminology, including the introduction of new words for new ideas, or perhaps the use of common words in precise scientific ways.

It is certainly possible to write an explanation of Faraday’s law without using the phrases induced emf, rate of change and magnetic flux linkage, but it might require several pages instead of one line.

Electromagnetic induction in rotating coils

SYLLABUS CONTENT

- A uniform magnetic field induces a sinusoidal varying emf in a coil rotating within it.
- The effect on induced emf caused by changing the frequency of rotation.

Electromagnetic induction is used to generate most of the world's electrical energy.

Alternating current (ac) generators

Consider Figure D4.30, which shows a coil of wire between the poles of a magnet. For simplicity, only one loop is shown, but in practice there will be a large number of turns on the coil(s) of any practical ac generators.

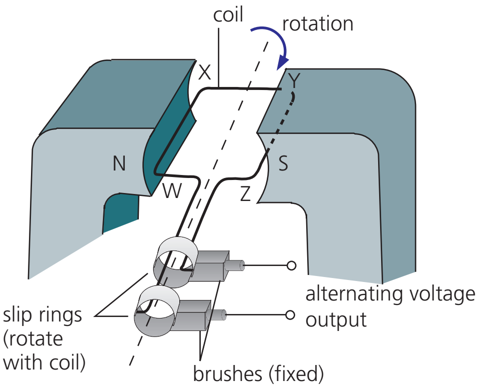


Figure D4.30 A simple ac generator

If the coil is rotating, there will be a changing magnetic flux passing through it and a changing emf will be induced. As side WX moves upwards, the induced emf will make a current flow into the page, if it is connected in a circuit. At the same time any induced current in YZ will flow out of the page, because it is moving in the opposite direction. In this way, the current flows continuously around the coil.

The emfs induced in opposite sides of a coil rotating in a magnetic field act in series to drive a current around the coil.

The connection between the coil and the external circuit cannot be fixed and permanent because the wires would become twisted as the coil rotated. Therefore carbon 'brushes' are used to make the electric contact with **slip rings** which rotate with the coil, so that the induced current can flow into an external circuit.

Figure D4.31 shows three views of the rotating coil from the side. In Figure D4.31a the plane of the coil is parallel to the magnetic field and, at that moment, the sides WX and YZ are cutting across the magnetic field at the fastest rate, so this is when the maximum emf is induced. In Figure D4.31b the sides WX and YZ are moving parallel to the magnetic field, so no emf is induced at that moment. In Figure D4.31c the induced emf is a maximum again, but the direction is reversed because the sides are moving in the opposite direction to Figure D4.31a.

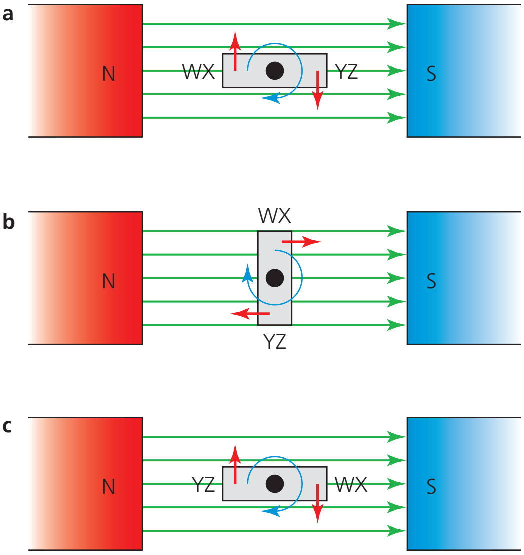


Figure D4.31 The sides of the coil cut through the magnetic field at different angles as they rotate, alternating the emf produced

The overall result, if the coil rotates at a constant speed in a uniform magnetic field, is to induce an emf that varies *sinusoidally*. This is shown by the green line in Figure D4.32.

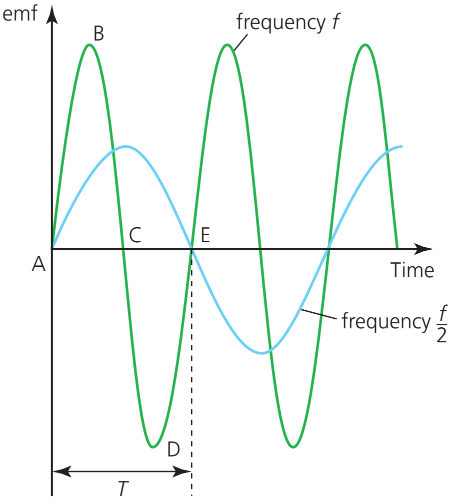


Figure D4.32 Comparing induced emfs at different frequencies

In positions B and D the plane of the coil is parallel to the magnetic field. At A, C and E the plane of the coil is perpendicular to the field. One complete revolution occurs in time T. Frequency, f, equals $1/T$. If the coil rotates at a slower frequency (fewer rotations every second), then there will be a smaller rate of change of magnetic flux through it and a smaller emf will be induced. For example, halving the frequency will halve the

◆ **Slip rings and brushes** In an ac generator these are used for connecting the rotating coil to the external circuit.

rate of change of magnetic flux linkage and therefore halve the induced emf. The time period is doubled. This is represented by the blue line in Figure D4.32. You should watch a computer simulation of an ac generator as the coil(s) rotates slowly to help your understanding.

Throughout the world, electrical energy is generated in this way using **turbines** with ac generators. A turbine is a device which transfers the kinetic energy of a moving fluid into useful rotation. Turbine blades can be made to rotate by, for example, forces from the wind, or from high-pressure steam produced from burning fossil fuels or nuclear reactions, or from falling water (hydroelectricity). See Figures D4.33 and D4.34.

◆ **Turbine** Device that transfers the energy from a moving fluid to do mechanical work and cause (or maintain) rotation.

◆ **Alternator** ac electrical generator.

In order to generate electricity, turbine blades are attached to the coils inside an ac generator. The coils – with many turns and cores with high magnetic permeability – are rotated in strong magnetic fields by the action of the turbine blades. Electricity can also be generated using the same principle, but with the magnetic field rotating inside the circuit, rather than the other way around. Such devices are commonly used in cars and they are often called **alternators**. Note that dc generators can be similar to ac generators in basic design, but the connections to the external circuit need to be modified.

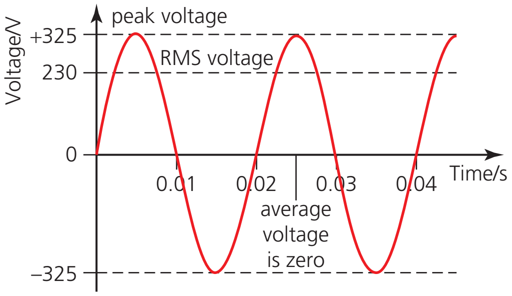
Mains electricity (also known as *utility power*) is the name given to the electrical power supply that is delivered from large power stations to homes and businesses. Figure D4.35 shows the symbol for an ac power supply.

In most of the world the electricity power supply is rated at 230 V ac 50 Hz. This is shown in Figure D4.36. 120 V 60 Hz is common in North America. There are a range of different designs of sockets and plugs used in different countries.

■ **Figure D4.33** An engineer working on a steam turbine

■ **Figure D4.34** Wind turbine blades are set in motion by the flow of wind past them

■ **Figure D4.35** Symbol for an ac power supply



■ **Figure D4.36** Mains voltage rated at 230 V

An ac electrical supply which is rated at 230 V actually varies between peaks of +325 V and −325 V. The average value is zero, but that is not really useful information. The supply is rated at 230 V because it delivers the same power as a steady 230 V dc would in the same circuit. (The effective voltage is described as the *RMS voltage* – root mean squared voltage. RMS voltage = peak voltage divided by $\sqrt{2}$. But you are *not* expected to remember this.)

Two wires are required to make a connection from the mains and deliver electrical energy to a circuit. One connection, called the *neutral wire*, is always kept at 0 V, while a varying voltage, as seen in Figure D4.36, drives a current backwards and forwards. This connection is commonly called the *live wire*.

LINKING QUESTION

- How is the efficiency of electricity generation dependent on the source of energy?

This question links to understandings in Topic B.4.

31 Figure D4.37 shows the output of an ac generator.

- **a** Determine the frequency of the output.
- **b** Estimate the approximate voltage rating of the supply.
- **c** Make a copy of Figure D4.37 and add a curve to represent the output if the frequency of the generator was reduced to 25 Hz.

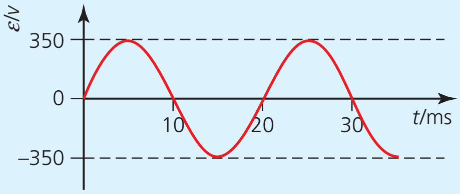


Figure D4.37 Output of an a.c. generator

32 a Sketch a voltage–time graph, with numerical values, for a 120 V 60 Hz rated power supply.

- **b** Estimate the maximum voltage in each cycle of the supply.

33 A rectangular coil of copper wire has 500 turns and sides of length 5.2 cm and 8.7 cm. The coil rotates at a constant frequency of 24 Hz about an axis that passes centrally through the shorter sides. A uniform magnetic field of 0.58 T acts on the coil in a direction perpendicular to the axis.

- **a** Draw a labelled sketch of this arrangement.
- **b** Calculate the linear speed of the sides of the coil.
- **c** Use $\varepsilon = NBvL$ to determine the emf that is induced across one of the longer sides at the instant that it is moving perpendicularly across the field.
- **d** What are the maximum and minimum induced emfs as the coil rotates?

34 What are the voltages on the live and neutral wires in your home?

35 Outline why some electrical sockets have three connections (rather than two).

Self-induction

◆ **Back-emf** An induced potential difference that opposes a change of current in the same circuit.

◆ **Self-induction** Electromagnetic induction within a single circuit.

When a current in any circuit changes, the magnetic flux associated with that current must also change. This means that an emf will be induced. So far, we have discussed induction in a *separate* circuit, but induction also occurs within the *same* circuit and then the induced emf opposes the change of current (Lenz's law), so that it acts in the reverse direction to the original emf producing the current. It is often called a **back-emf**.

In most simple circuits this effect will not be noticeable or important, but if a many-turned coil is involved, especially if it is wound on a core of high magnetic permeability, the effect will become significant when dealing with alternating currents. It is called **self-induction**.

Self-induction is the effect in which a change in the current in a circuit tends to produce an induced emf which opposes the change of current *in the same circuit.*

Figure D4.38 Treasure hunting

Self-induction becomes more important at higher frequencies, which usually involve greater values of $\Delta I/\Delta t$, and so greater rates of changing magnetic flux.

The magnitude of self-induction (or mutual induction) effects will change with the magnetic properties of different surrounding materials (especially metals). This has some interesting applications, including:

- The presence of cars waiting at traffic lights can be detected by changes to the self-induction of coils under the road surface.
- Metal detectors (for example at airport security checks) can use changes in self-inductance to identify the presence of metals. See Figure D4.38 for a similar application.

E.1 Structure of the atom

Guiding questions

- What is the current understanding of the nature of an atom?
- What is the role of evidence in the development of models of the atom?
- In what ways are previous models of the atom still valid despite recent advances in understanding?

The nuclear model of the atom

We have already briefly described the nuclear model of the atom, at the beginning of Topic B.5. As a reminder, Figure E1.9 shows another visualization of this model.

We now want to use knowledge about electric forces (Topic D.2) to explain how and why the *nuclear* model of the atom was first developed.

◆ **Subatomic particle** Any particle contained within an atom.

At the end of the nineteenth century, it had become clear that the atom was not an elementary, indivisible particle. However, only one **subatomic particle**, the negatively charged electron, had been identified, although it was not known for sure how many electrons were in each atom. Since atoms were not charged overall, it was clear that there must also be some part of atoms which were positively charged. This led to a model of the atom often described as the 'plum pudding model' (J.J. Thomson in 1904), see Figure E1.1.

Today, we may be more inclined to use a blueberry muffin visualization (Figure E1.2): an unknown number of individual blueberries (negative electrons) in a muffin of spread-out positive charge.

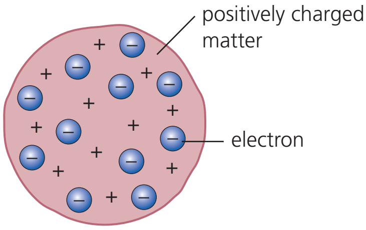


Figure E1.1 The 'plum pudding model' of the atom

Figure E1.2 Blueberry muffin

This model of the atom was far from being satisfactory and it raised many questions, but it was to be about seven years before it was improved, following the famous experiments of Geiger, Marsden and Rutherford.

TOK

The natural sciences

‘What is everything made of?’ is one of the most basic questions we can ask and there are records going back about 2500 years, to Greek philosophers (Democritus and others), asking just that. It was at that time that the concept of the ‘atom’ was first introduced: as tiny, solid spheres.

Moving forward about 2200 years, scientists of that time still had much the same ideas. The following is a quote from the famous physicist Isaac Newton in 1704.

‘All these things being considered, it seems probable to me that God in the Beginning form’d Matter in solid, massy, hard, impenetrable, moveable Particles, of such Sizes and Figures, and with such other Properties, and in such Proportion to Space, as most conduced to the End for which he form’d them; and that these primitive Particles being Solids, are incomparably harder than any porous Bodies compounded of them; even so very hard, as never to wear or break in pieces; no ordinary Power being able to divide what God himself made one in the first Creation.’

A lot has changed in the following three centuries. Chemical reactions were explained by the elements having different atoms which could be combined to form molecules (Dalton and others). However, the concept of the indivisible atom remained until the discovery (1897) of a constituent particle: the electron.

Rutherford’s proposal of a nuclear atom (and the existence of protons and neutrons), as explained below, was another paradigm shift in models of the atom, but not the last. The discovery of the wave properties of electrons (1924) meant that the model had to be significantly changed again.

◆ **Geiger–Marsden–Rutherford experiment** The scattering of alpha particles by a thin sheet of gold foil, which demonstrated that atoms consist of mostly empty space with a very dense positively charged core (the nucleus).

◆ **Alpha particle** A positively charged particle emitted by a radioactive nucleus.

◆ **Radioactive source** Radioactive substance used for the nuclear radiation it emits.

Geiger–Marsden–Rutherford experiment

SYLLABUS CONTENT

- ▶ The Geiger–Marsden–Rutherford experiment and the discovery of the nucleus.

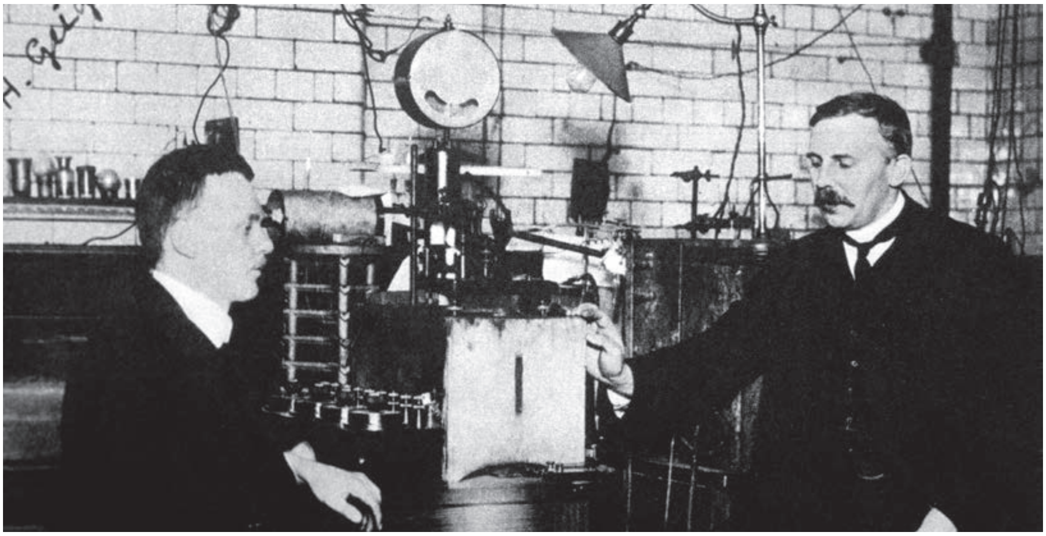


Figure E1.3 Geiger and Rutherford (right)

In 1909, Ernest Rutherford and two of his research students, Hans Geiger and Ernest Marsden, working at the University of Manchester, UK, directed a narrow beam of positively charged **alpha particles** (see below) from a **radioactive source** at very thin gold foil. A zinc sulfide detector was moved in a circle around the foil to determine the directions in which alpha particles travelled after striking the foil (Figure E1.4). The alpha particles had enough energy individually to be detected by a tiny flash of light when they were stopped by the zinc sulfide.

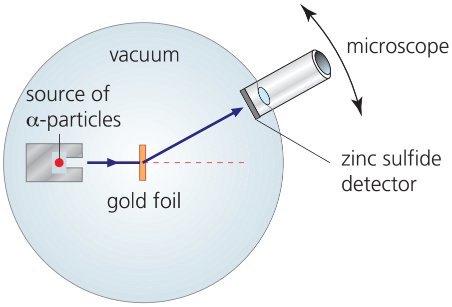


Figure E1.4 The alpha particle scattering experiment

Inquiry 1: Exploring and Designing

Exploring

What are alpha particles?

The nuclei of some atoms are unstable and they can emit particles and electromagnetic radiation. This is called **radioactivity** and the subject is covered in Topic E.3. Alpha particles are a common product of radioactivity and they are often used in school demonstrations. Research using a variety of relevant sources to find out:

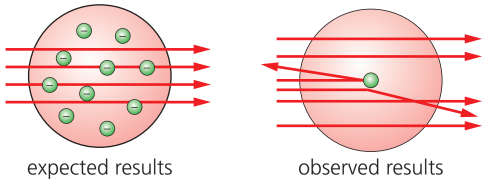
1 What subatomic particles are alpha particles made from?
2 What is the overall electrical charge of an alpha particle?

◆ **Radioactive decay (radioactivity)** Spontaneous emission of particles and / or radiation from unstable nuclei.

Alpha particles carry a *very* large amount of energy relative to their small size and, at that time (1909), it was expected that the alpha particles would not be affected much by passing through such thin gold. Gold foil can be made very thin (less than 10^{-6} m) and the foil may then only have about 6000 layers of atoms. Although alpha particles only travel about 4 cm in air, they would encounter many more molecules travelling that distance in air than passing through gold atoms in very thin foil.

But Geiger and Marsden's results were surprising. Rutherford published the results in 1911. He reported that:

- Most of the alpha particles passed through the foil with very little or no deviation from their original path (as was expected).
- A small number of particles (about 1 in 1800) were deviated through an angle of more than about 10° (see Figure E1.5).
- An extremely small number of particles (about 1 in 10 000) were deflected through an angle larger than 90°. Some particles were even deflected by 180°, returning in the direction from which they came.


expected results

observed results

Figure E1.5 Alpha scattering experiments produced unexpected results

The importance of this last point is emphasized by Rutherford's famous quote: *'It was quite the most incredible event that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch shell [large bullet] at a piece of tissue paper and it came back and hit you.'*

From alpha particle scattering experimental results Rutherford drew the following conclusions:

Most of the mass of an atom is concentrated in a very small volume at the centre of the atom. Most alpha articles would therefore pass through the foil undeviated (continuing in a straight line) because most of the atom was empty space.

The centre of an atom (he called it the nucleus) must be positively charged in order to repel the positively charged alpha particles. Alpha particles that pass close to a nucleus will experience a strong electrostatic repulsive force, causing them to change direction.

Only alpha particles that pass very close to the nucleus, striking, or almost striking it directly, will experience electrostatic repulsion large enough to cause them to deviate through large angles. The fact that so few particles did so, confirms that the nucleus is very small, and that most of the atom is empty space.

Figure E1.6 shows some of the possible trajectories (paths) of the alpha particles. Rutherford used his new nuclear model of the atom and Coulomb's inverse square law (covered in Topic D.2) to explain the repulsive force between the positively charged particles. He used the magnitudes of the forces to calculate the fraction of alpha particles expected to be deviated through various angles.

Rutherford's calculations agreed very closely with the results from the experiment, supporting his proposal of a nuclear model of the atom.

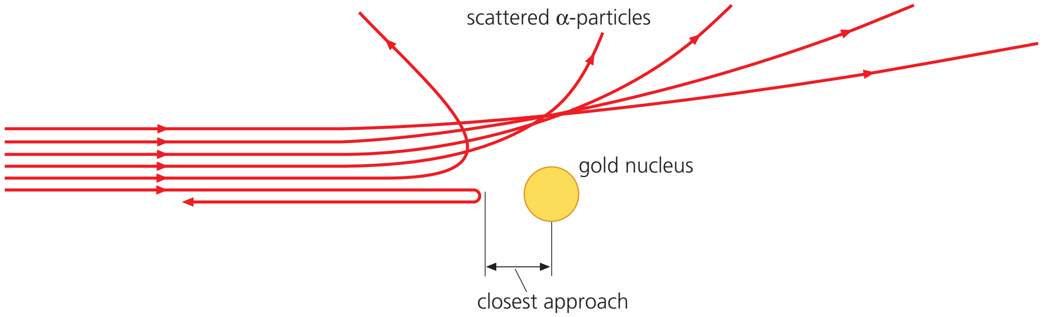


Figure E1.6 Alpha particle trajectories in the gold foil experiment

Figure E1.7 Alpha particle scattering analogue

Alpha particle scattering can be modelled using simple apparatus such as that shown in Figure E1.7, in which a small ball rolls down a wooden ramp onto a specially shaped metal 'hill'. The shape of the hill is made so that, when viewed from above, the ball moves as if it was being repelled from the centre of the hill by an inverse square law of repulsion. In other words, gravitational forces are used to model electric forces. Using this apparatus, it is possible to investigate how the direction in which a ball travels after leaving the hill (the scattering angle) depends on its initial direction ('aiming error') and/or its energy. In Geiger and Marsden's experiment it was not possible to observe the scattering paths of individual alpha particles, but the observed scattering pattern of large numbers of alpha particles in a beam is found to be in very close agreement with modelling based on individual balls rolling on hills.

From his results, Rutherford calculated that the diameter of the gold nucleus was of the order of 10^{-14} m, compared to the diameter of the whole atom, which was known to be of the order of 10^{-10} m.

LINKING QUESTION

- How have observations led to developments in the model of the atom? (NOS)

1 Explain in your own words (less than 100), without using a diagram, why Rutherford concluded that atoms contain a small, positively charged central nucleus.

2 Sketch the path of an alpha particle being scattering by a large angle by a positive nucleus. Label the 'aiming error' and the 'scattering angle'.

3 'Gold foil can be made very thin (less than 1×10^{-6} m) and the foil may then only have about 6000 layers of atoms.'

a Use this information to determine an approximate radius of a gold atom.

b State any assumptions that you made in answering part **a**.

4 a Calculate the forces acting between an alpha particle and a nucleus of charge $+81e$ when they are at their closest, separated by a distance of 2.0×10^{-14} m.

b If the alpha particle was scattered by an angle of about 40°, sketch its path, showing the forces that you calculated in part **a**.

5 Suggest what would have happened if neutrons had been used in Rutherford's experiment instead of alpha particles. Explain your answer.

WORKED EXAMPLE E1.1

a Make a sketch of an alpha particle being deflected through about 120° by a gold nucleus.

b On the same sketch draw the path of an alpha particle (approaching along the same path as before) being scattered by a copper nucleus (which has a much smaller charge).

c Show how an alpha particle of higher energy could be affected if it approached a gold nucleus along the same path.

Answer

a

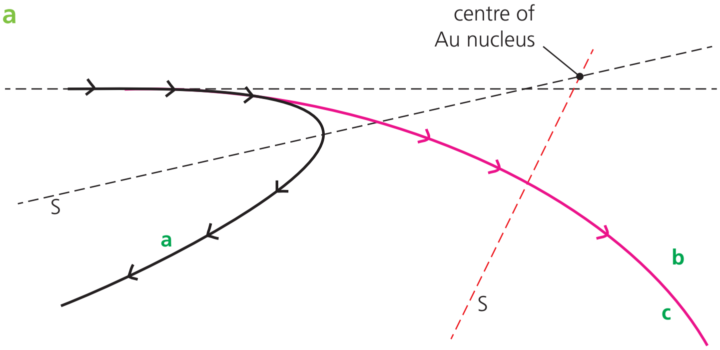


Figure E1.8 Answer to Worked example E1.1

The repulsive force from a copper nucleus is less than from a gold nucleus, so the alpha particle is scattered less. An alpha particle of greater energy will be scattered through a smaller angle than an alpha particle of less energy. For convenience, **b** and **c** have been shown with similar deflection, but that is unlikely. All paths are symmetrical about the dashed lines labelled S.

Composition of the nucleus

In the years that followed Rutherford's famous experiment, it was confirmed that a nucleus consists of *separate subatomic particles: protons and neutrons*, which contain almost all of the mass of the atom. The protons are positively charged and the neutrons are electrically neutral. The electrons are negatively charged but have very little mass in comparison to protons and neutrons. Atoms are electrically neutral because there are equal numbers of protons and electrons.

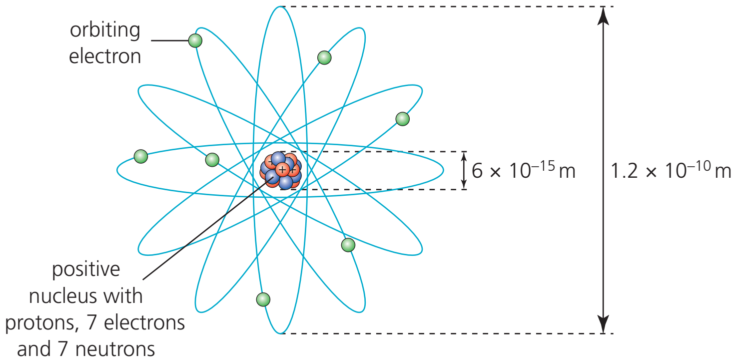
In this model of the atom, the electrons orbit the nucleus because of the centripetal force provided by the electrical attraction between opposite charges. If the electrons were not in circular motion, the electrostatic force of attraction would accelerate them towards the nucleus.

The vast majority of an atom is empty space, a vacuum. The properties of protons, neutrons and electrons are summarized in Table E1.1.

Table E1.1 Properties of subatomic particles

Name of particle	Approximate relative mass	Relative charge
proton	1	+1
neutron	1	0
electron	1/1840	−1

Figure E1.9 shows the features of the nuclear model of an atom as visualized in the years immediately following Rutherford's discovery. The example shown is a nitrogen atom. Although this model has been changed in important ways, some of which are explained below, it persists in popular culture and it continues to be a very useful starting point in the study of atomic structure in elementary science lessons.



■ **Figure E1.9** The orbital model of a nitrogen atom (not to scale)

Nature of science: Models

Power of visual models

The latest models of the atom cannot be drawn on paper or a computer screen. Instead, we need complex mathematical models (equations) to describe what we cannot see. (These are *not* included in the IB course.) It is unlikely that scientists will ever produce an accurate model of the atom that can be visualized, but we should have no expectation that the atomic-scale world behaves in any way similar to the world we see around us.

The visual model seen in Figure E1.9 can be understood to some extent by many people, but the complex mathematical modelling needed to explain the latest theories about the structure of matter will continue to be inaccessible to most people.

What holds the particles in the nucleus together?

After it was proposed that the nucleus was composed of separate particles (protons and neutrons), there was an obvious question to ask: what forces are there between these particles that holds them so closely together? In particular, it was known that there is a very large *repulsive* force between protons, as the following approximation demonstrates:

$$F = k\frac{q_1 q_2}{r^2}$$

(from Topic D.2)

$$F \approx \frac{(8.99 \times 10^9) \times (1.60 \times 10^{-19})^2}{(6 \times 10^{-15})^2} \approx 10\,\text{N (to an order of magnitude)}$$

which, on the atomic scale, is an extremely large force.

To oppose this repulsive force, we now know that there is a very *short-range* **strong nuclear force** (attractive) between the nuclear particles. However, detailed knowledge of this force is not required in the IB course.

The term **nucleon** is used to describe a particle in the nucleus of an atom which is either a proton or a neutron.

◆ **Strong nuclear force** Fundamental force that is responsible for attracting nucleons together. It is a short-range attractive force (the range is about 10^{-15} m), but for smaller distances it is repulsive, and hence it also prevents a nucleus from collapsing.

◆ **Nucleon** A particle in a nucleus, either a neutron or proton.

ATL E1A: Research skills

Use search engines and libraries effectively

For many years (until the 1960s) it was believed that protons and neutrons were **elementary particles**, meaning that it was thought that they were not composed of smaller particles. Scientists now know that protons and neutrons are *composite particles*, each consisting of three *quarks* with the strong nuclear force holding them together. You are not expected to have knowledge of quarks for examinations.

There are 17 known elementary particles, which can be arranged into four groups. Use the internet to find out the names of these particles. Do you find it surprising that there are 17? Explain your answer.

◆ **Elementary particles** Particles that have no internal structure. They are not composed of other particles. For example, electrons.

◆ **Proton number, *Z*** The number of protons in a nucleus.

◆ **Nucleon number, *A*** The total number of protons and neutrons in a nucleus.

◆ **Neutron number, *N*** The number of neutrons in a nucleus.

The number of protons in the nucleus of an atom determines which element it is. So, atoms of a particular element are identified by their **proton number** (sometimes called *atomic number*), which is given the symbol Z. The *periodic table* of the elements arranges the elements in order of increasing proton number.

The proton number, Z, is the number of protons in the nucleus of an atom.

Because atoms are electrically neutral, the number of protons must be equal to the number of electrons in the space around the nucleus. If electrons are added or removed from an atom, it is then described as an *ion* (of the same element).

The **nucleon number**, A, is defined as the total number of protons and neutrons in a nucleus.

The nucleon number represents the mass of an atom, because the mass of the electrons is (almost) negligible. (Nucleon number is sometimes referred to as the *mass number*.)

The difference between the nucleon number and the proton number gives the number of neutrons in the nucleus: $N = A - Z$.

The **neutron number**, N, is defined as the number of neutrons in a nucleus.

The number of neutrons in a nucleus is similar to the number of protons, although the ratio of the number of neutrons / number of protons generally increases with increasing proton number. (See Figure E3.22 in Topic E.3). As we will see, this ratio is an important factor when considering the stability of nuclei.

Nuclear notation

SYLLABUS CONTENT

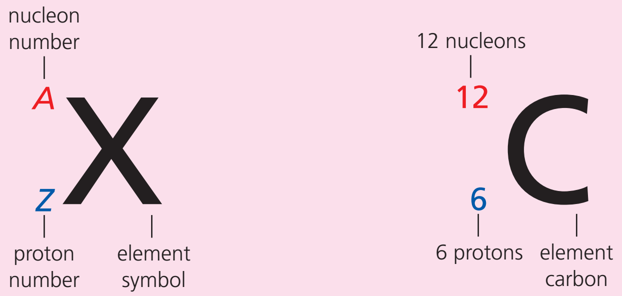
- Nuclear notation ${}^{A}_{Z}\mathrm{X}$ where A is the nucleon number Z is the proton number and X is the chemical symbol.

◆ **Nuclide** Term used to identify one particular species (type) of atom, as defined by the structure of its nucleus. A *radionuclide* is unstable and will emit radiation.

The term **nuclide** is used to specify one particular species (type) of atom, as defined by the structure of its nucleus.

All atoms with the same nucleon number and the same proton number are described as the same nuclide.

There is a standard notation used to represent a nuclide by identifying its proton number and nucleon number, as shown in Figure E1.10, which uses C-12 as an example.



■ **Figure E1.10** Standard notation for specifying a nuclide

Isotopes

◆ **Isotope** One of two or more atoms of the same element with different numbers of neutrons (and therefore different masses). A *radioisotope* is unstable and will emit radiation.

Two or more nuclides with the same proton number may have different numbers of neutrons. The atoms are of the same element and have identical chemical properties, but they have different nucleon numbers. These atoms are called **isotopes**.

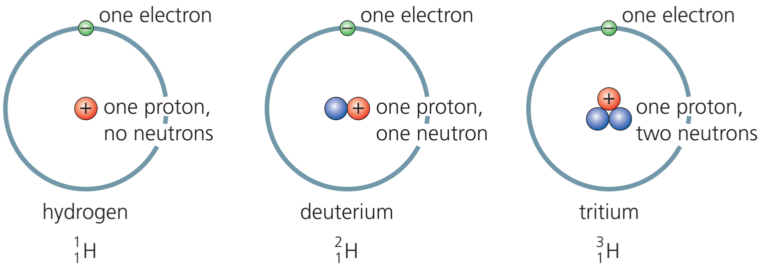
An isotope is one of two or more different nuclides of the same element (which have the same number of protons, but different nucleon numbers).

Some elements have many isotopes, but others have few or even one. For example, the most common isotope of hydrogen is hydrogen-1, $^{1}_{1}H$. Its nucleus is a single proton.

Hydrogen-2, $^{2}_{1}H$, is called deuterium; its nucleus contains one proton and one neutron.

Hydrogen-3, $^{3}_{1}H$, with one proton and two neutrons, is called tritium.

Hydrogen isotopes (Figure E1.11) are involved in nuclear fusion reactions (see Topic E.5).



■ **Figure E1.11** The three isotopes of hydrogen

As a further example, the following nuclides are three isotopes of carbon:

- $^{12}_{6}C$ (six protons, six neutrons)
- $^{13}_{6}C$ (six protons, seven neutrons)
- $^{14}_{6}C$ (six protons, eight neutrons).

◆ **Diffusion** Random movement of particles from a place of high concentration to places of lower concentration.

Samples of elements are often mixtures of isotopes. Isotopes cannot be separated by chemical means. Separation can only be achieved by processes that depend on the difference in masses of the isotopes, for example the **diffusion** rate of gaseous compounds.

The notation for describing nuclides can also be applied to the nucleons. For example, a proton can be written as $^{1}_{1}p$ and a neutron as $^{1}_{0}n$. An electron's charge is −1 compared to the +1 charge on a proton, so an electron can be represented by $^{0}_{-1}e$, remembering that the mass (number) of the electron is effectively zero compared to the proton and neutron.

WORKED EXAMPLE E1.2

A certain element has the proton number 17.

a Research which element this is.

b Suggest the nucleon numbers of two possible isotopes of this element.

c Chemists may say that the atomic mass (weight) of this element is 35.5. Explain what this number represents.

Answer

a Chlorine

b Chlorine has a large number of isotopes, each with a different nucleon number. We know that the number of neutrons is approximately equal to the number of protons, so $A = 33, 34, 35, 36, 37$ and so on are all reasonable guesses. (In fact, the two most common are chlorine-35 and chlorine-37.)

c 35.5 represents the *average* number of a large number of nucleons in a sample of chlorine (with a mixture of isotopes).

6 Explain the differences between an atom, a nuclide and an isotope.

7 The nuclides $^{129}_{53}\text{I}$, $^{137}_{55}\text{Cs}$ and $^{90}_{38}\text{Sr}$ were all formed during atomic weapons testing more than 40 years ago. State the number of neutrons, protons and electrons in the atoms of these nuclides.

8 State the electric charge of the nucleus $^{4}_{2}\text{He}$.

9 The number of electrons, protons and neutrons in an ion of sulfur, S, are equal to 18, 16 and 16, respectively. What is the correct nuclide symbol for this sulfur ion?

10 State the number of nucleons in one carbon-13 atom, $^{13}_{6}\text{C}$.

11 Chlorine, Cl, is an element that has 17 protons in its nucleus. The two most common isotopes of chlorine are chlorine-35 and chlorine-37. Write down the nuclide symbols for these two isotopes.

12 U-238 and U-235 are the two most common isotopes found in uranium ore. The more massive isotope has 146 neutrons in its nucleus.

- **a** Write down the nuclide symbols for these two isotopes.
- **b** Explain why it is difficult to separate these isotopes from each other.

Energy levels within atoms

SYLLABUS CONTENT

- ▶ Emission and absorption spectra provide evidence for discrete atomic energy levels.
- ▶ Emission and absorption spectra provide information on the chemical composition.

The total energy of an atom, such as the nitrogen atom represented in Figure E1.9, may be considered to be the sum of the kinetic energies of the electrons plus the electric potential energy of the system of negatively charged electrons moving around the positively charged nucleus (assumed to be stationary). Energies *within* the nucleus are for another discussion and they are not included in this topic.

Comparing this simplified model of an atom to a gravitational model of satellites orbiting a planet (Topic D.1), we define our zero of electric potential energy in the same way: at infinity.

The energies of electrons within atoms are given negative values because we would have to supply energy to remove them from the atom (to infinity), where they would then be considered to have zero electric potential energy.

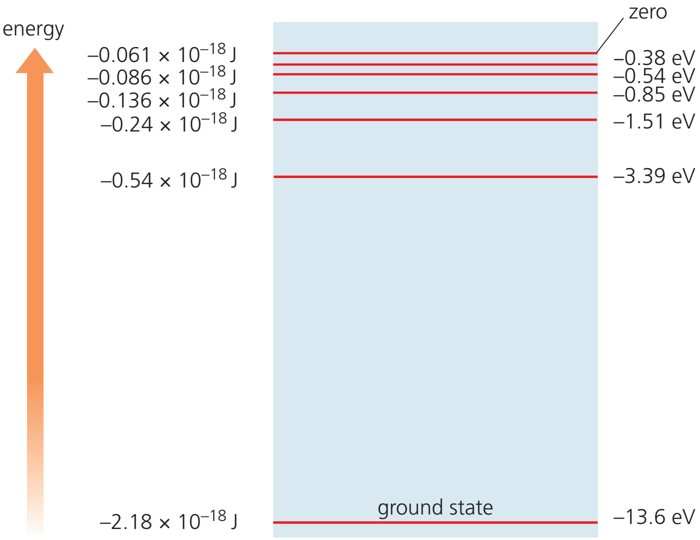
◆ **Atomic energy level** One of a series of possible discrete (separate) energy levels of an electron within an atom.

◆ **Ground state** The lowest energy state of an atom / electron (or nucleus).

◆ **Ionization energy** Amount of energy needed to remove an electron from an atom or molecule.

The orbiting electrons model of an atom has its uses, but is a long way from the whole truth, as we shall explain. Most significantly, orbiting satellites can, in principle, orbit at *any* height, with a *continuous* range of possible energies, but electron energies are very different: they can only have one of a range of very precise values. That is, their possible energies are *discrete*, *quantized*. We refer to the possible energies as **atomic energy levels**.

Figure E1.12 shows the simplest example: possible energy levels (of an electron) within the simplest atom, hydrogen. These levels will be discussed in detail later.



■ **Figure E1.12** The energy levels of the hydrogen atom

The following points should be noted about this important diagram:

- The energy levels are drawn to scale vertically, but the shape of this diagram has no physical meaning.
- The **ground state** is the lowest possible energy level. An electron in the ground state of any atom is the most difficult to remove. Atoms are usually in their ground states.
- *All* energies are negative (as explained above).
- The highest energy level shown is equivalent to removing the electron from the atom (to infinity, where it would then have zero energy – if it were not moving). That is, the **ionization energy** of hydrogen atoms is 13.6 eV (2.18×10^{-18} J).
- The energy levels have been given in both joules and electronvolts. Electron volts are widely used for atomic-scale energies.

Top tip!

A reminder: 1 eV (electronvolt) is the amount of energy transferred when unit charge *e* is accelerated, or decelerated, by a potential difference of one volt. ($W = qV$; $1\,\text{eV} = 1.6 \times 10^{-19}$ J) The electronvolt is a common unit of energy used throughout atomic physics, not just for accelerated charges.

13 Explain what it means if we say that the (first) ionization energy of an atom is 4.0×10^{-18} J.

14 Show that an energy of -0.136×10^{-18} J is equivalent to −0.85 eV (as shown in Figure E1.12).

15 110 eV is required to ionize an atom in its ground state. The four lowest energy levels above the ground state are −70 eV, −40 eV, −20 eV and −10 eV.

a Draw an energy level diagram for this atom.

b Determine the number of different transitions possible between these five levels.

We need to *explain* how physicists discovered that atoms had discrete energy levels. The evidence came from examining in detail the light that is emitted from (or absorbed by) atoms.

Evidence for energy levels within atoms

Emission and absorption spectra

Light is electromagnetic radiation that has been emitted from atoms. We can learn a lot about the energy inside atoms by examining the radiation (spectra) that atoms emit.

When elements (in the form of gases) are **excited** – given enough energy (by heating, or by an electrical current at high voltage), the spectra of the light that they *emit* are seen as a series of bright lines on black backgrounds– called *line spectra*. See Figure E1.13. Each line corresponds to a precise frequency. (In this Figure, a prism has been used to disperse the light. Alternatively, a diffraction grating could be used, as discussed in Topic C.3 for HL students.)

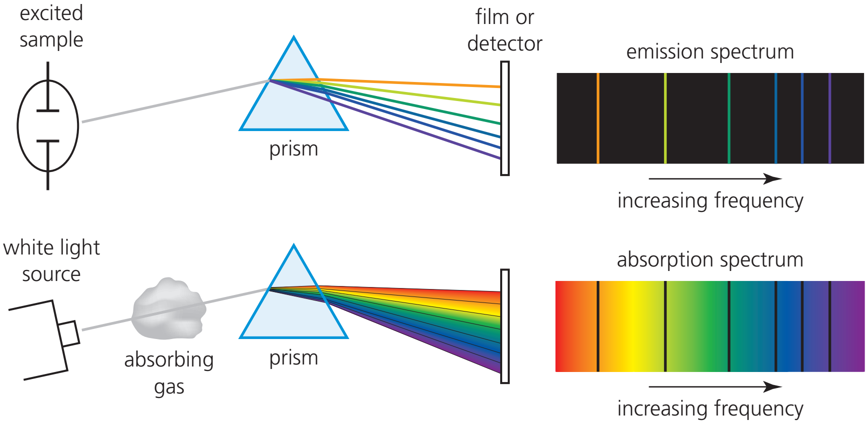


Figure E1.13 Emission and absorption spectra of the same element

When a continuous spectrum passes through a gas, the atoms in the gas will *absorb* the same frequencies as they would emit when given energy. This results in a spectrum with black absorption lines, also as seen in Figure E1.13 (lower diagram.) The atoms re-emit the energy, but in random directions.

Each line on an *emission spectrum* is explained by electrons moving to a lower energy level within the atom. Each line on an *absorption spectrum* is explained by electrons moving to a higher energy level within the atom.

Since the energy levels of atoms of different elements are different, emission and absorption spectra can be used to identify the elements involved.

◆ **Excitation** The addition of energy to a particle, changing it from its ground state to an excited state.

◆ **Emission spectrum** Line spectrum associated with the emission of electromagnetic radiation by atoms, resulting from electron transitions from higher to lower energy states.

◆ **Absorption spectrum** A series of dark lines across a continuous spectrum produced when white light passes through a gas at low pressure.

Inquiry 2: Collecting and processing data

Collecting data

The study of spectra is called (optical) **spectroscopy**, and instruments used to measure the wavelengths of spectra are called **spectrometers** (see Figure E1.14).

Ask your teacher, or otherwise find out, how a spectrometer is used to measure wavelengths of line spectra. Identify what precautions and methods must be taken to ensure accurate readings.

Figure E1.14 A spectrometer

LINKING QUESTION

- How can emission spectra allow for the properties of stars to be deduced?

This question links to understandings in Topic E.5.

Photons

SYLLABUS CONTENT

- Photons are emitted and absorbed during atomic transitions.
- The frequency of the photon released during an atomic translation depends on the difference in energy level as given by: $E = hf$.

◆ **Transition (between energy levels)** A photon is emitted when an atom (or nucleus) makes a transition to a lower energy level. The energy of the photon is equal to the difference in energy of the levels involved.

◆ **Photon** A quantum of electromagnetic radiation, with an energy given by $E = hf$.

◆ **Quantum** The minimum amount of a physical quantity that is quantized. Plural: quanta..

◆ **Planck's constant, *h*,** Fundamental constant of quantum physics which connects the energy and frequency of a photon.

The emission, transmission and absorption of light are not continuous processes. They are a very large number of separate events.

When an (excited) atom moves to a lower energy level, it emits an amount of energy equivalent to the *difference* in energy levels. Figure E1.15 shows a simplified example: an atom with four different energy levels that could have six possible energy **transitions between energy levels**.

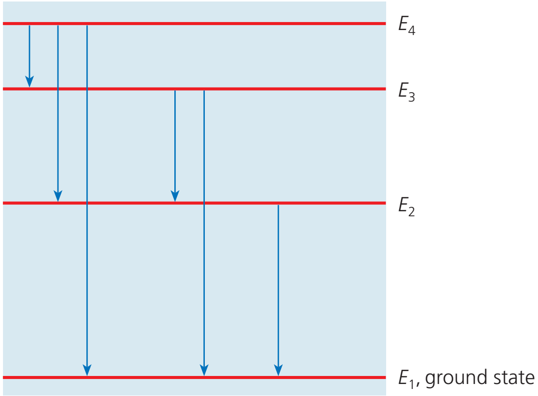


Figure E1.15 Energy transitions between four energy levels in an atom

If an atom receives electromagnetic energy equal to the difference between its energy level and a higher level, it can move to the higher level when the energy is absorbed. Typically, the energy is then quickly re-emitted as the atom returns to a lower energy level.

The 'bundles' of emitted electromagnetic energy are called **photons**. More generally, the term **quanta** (singular: **quantum**) is used to describe the smallest possible quantity of any entity that can only have discrete values. We can say that light is quantized.

The energy, E, carried by one photon of electromagnetic radiation depends only on its frequency, f, as follows:

$$\boldsymbol{E = hf}$$

h is a very important fundamental constant that controls the properties of electromagnetic radiations.

h is called **Planck's constant**. It has a value of $\mathbf{6.63 \times 10^{-34}\,Js}$.

Since we know from Topic C.2 that $c = f\lambda$, this equation is often rewritten as $E = \dfrac{hc}{\lambda}$.

(It is often convenient to know that $\boldsymbol{hc} = \mathbf{1.99 \times 10^{-25}\,Jm = 1.24 \times 10^{-6}\,eV\,m}$)

Nature of science: Measurement

Fundamental constants

Fundamental constants are the numbers that appear in the equations physicists use to describe the properties of force, mass and energy in the Universe around us (as in, for example, h in $E = hf$). They are believed to have exactly the same value at all places and for all time. Fundamental constants are determined experimentally and are not theoretical.

If any of the values of these constants were different, then the Universe would be very different.

In the IB Physics course, the list of fundamental constants used includes:

- gravitational constant, G
- speed of electromagnetic radiation in free space, c
- electric permittivity of free space, ε_0
- magnetic permeability of free space, μ_0 (connected to ε_0 and c)
- Planck's constant, h
- elementary charge, e

WORKED EXAMPLE E1.3

Calculate the energy carried by one photon of microwaves of wavelength 10 cm (as might be used in a mobile phone):

a in J

b in eV.

Answer

a $E = \frac{hc}{\lambda} = \frac{((6.63 \times 10^{-34}) \times (3.00 \times 10^{8}))}{0.10}$

$E = 2.0 \times 10^{-24}$ J

b $\frac{(1.989 \times 10^{-24})}{(1.60 \times 10^{-19})} = 1.3 \times 10^{-5}$ eV

Figure E1.16 shows the visible emission line spectrum of hydrogen.

Figure E1.16 Lines on the hydrogen spectrum

Common mistake

In examination questions when you are asked to *show* that a given value is valid, you should *show* the result of your calculation to more significant figures than given in the question.

Taking one line as an example: 434.2 nm

We can use $E = \frac{hc}{\lambda}$ to determine the energy of a photon with this wavelength:

$$E = \frac{hc}{\lambda} = \frac{((6.63 \times 10^{-34}) \times (3.00 \times 10^{8}))}{(434.2 \times 10^{-9})} = 4.58 \times 10^{-19}\text{ J (equal to 2.86 eV)}$$

Referring back to Figure E1.12, we can see that this amount of energy is equivalent to the difference between the energy levels of −0.54 eV and −3.39 eV.

Here, we are using one line of the hydrogen spectrum as an example but, similarly, all spectral lines can be directly related to specific transitions between the discrete energy levels of the atoms of different elements. In practice, the reverse is also true:

spectral lines were used to determine atomic energy levels.

WORKED EXAMPLE E1.4

a Show that the frequency of a photon emitted by a transition in the hydrogen atom between its two lowest energy levels is approximately 2×10^{15} Hz.

b State in which part of the electromagnetic spectrum this radiation occurs.

Answer

a Consider Figure E1.12. The transition is from -0.54×10^{-18} J down to -2.18×10^{-18} J $= -1.64 \times 10^{-18}$ J

$E = hf$

$1.64 \times 10 = 6.63 \times 10^{-34} \times f$

$f = 2.5 \times 10^{15}$ Hz

b This frequency is in the ultraviolet part of the spectrum.

16 a Show that when an electron in an energy level of -1.36×10^{-18} J moved to a level of -0.74×10^{-18} J, a photon of energy approximately 4 eV was involved.

b Was the photon emitted or absorbed?

17 a Determine the frequency of electromagnetic radiation which has photons of energy 1.0×10^5 eV.

b State the name we give to that kind of radiation.

18 a A microwave oven uses electromagnetic photons of energy 1.6×10^{-24} J. What is the wavelength of this radiation?

b Use the internet to find out why this wavelength is used.

19 One of the electromagnetic frequencies absorbed by the greenhouse gas carbon dioxide is 1.4×10^{14} Hz.

a Calculate how much energy is carried by the absorbed photons.

b In what part of the electromagnetic spectrum is this radiation?

20 A particular visible line in the spectrum of oxygen has a wavelength of 5.13×10^{-7} m. Determine the energy (eV) transferred by one photon of this radiation.

21 Light has a typical wavelength of 5×10^{-7} m, and X-rays have a typical wavelength of 5×10^{-11} m.

a Draw a small square of sides 2 mm to represent the energy carried by a light photon.

b Assuming that photon energy is represented by the area, draw another square to represent the energy carried by an X-ray photon.

c Suggest why X-rays are more dangerous than light.

22 A light bulb emits light of power 7.0 W. Estimate the number of photons emitted every second.

23 An atom has six energy levels. What is the maximum possible number of transitions between these levels?

24 Consider Figure E1.17, which shows some of the energy levels in a mercury atom.

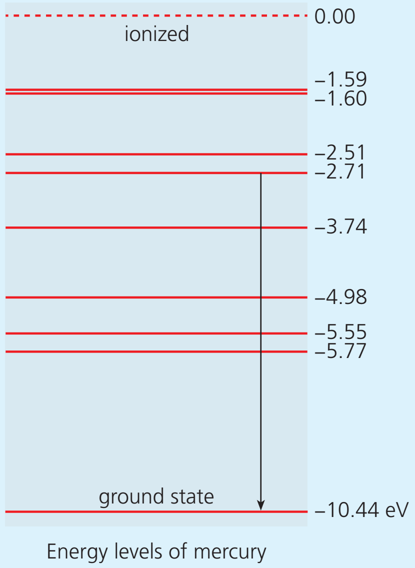


Figure E1.17 Some of the energy levels of mercury

a Determine the wavelength of radiation emitted by the transition shown.

b State in which part of the electromagnetic spectrum this radiation occurs.

c When radiation of frequency 1.18×10^{15} Hz passes through cool mercury vapour it is absorbed. Identify the transition involved in this process.

d Determine the longest wavelength of radiation that could be emitted by a transition between the levels shown.

25 When the spectrum emitted by the Sun is observed closely using a spectrometer, by looking at a white surface – **not** the Sun directly, it is found that light of certain frequencies is missing and, in their place, are dark lines.

a Explain how the cooler outer gaseous atmosphere of the Sun is responsible for the absence of these frequencies.

b Suggest how an analysis of the solar absorption spectrum could be used to determine which elements are present in the Sun's atmosphere.

A mathematical understanding of the Geiger–Marsden–Rutherford experiment

SYLLABUS CONTENT

- The distance of closest approach in head-on scattering experiments.
- The relationship between the radius and the nucleon number for a nucleus as given by: $R = R_0 A^{\frac{1}{3}}$ and the implications for nuclear densities.

LINKING QUESTION

- How is the distance of closest approach calculated using conservation of energy?

This question links to understandings in Topics A.3 and D.2.

Refer again to Figure E1.6. As an alpha particle approaches a positive nucleus, it loses kinetic energy, E_k, because it is being repelled, but the same amount of energy is transferred to electric potential energy, E_p, as shown in more detail in Figure E1.18.

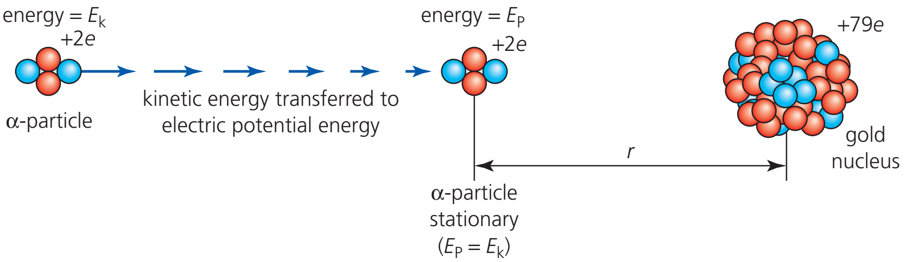


Figure E1.18 As an alpha particle approaches a positive nucleus, energy is transferred to electric potential energy, E_p

An alpha particle moving *directly* towards a nucleus will have lost *all* of its initial kinetic energy at the moment it is stationary, before returning back in the opposite direction. Then, assuming that the gold nucleus does not gain any significant kinetic energy:

initial kinetic energy of the alpha particle = the maximum electric potential energy momentarily stored in the system when the alpha particle is at its closest to the nucleus, with a separation of r. ($E_p = kq_\alpha q_n/r$)

WORKED EXAMPLE E1.5

Determine the closest distance from a gold nucleus that is possible for an alpha particle with kinetic energy of 5.0 MeV. Charge on alpha particle = $(2 \times 1.60 \times 10^{-19}\,\text{C})$, charge on gold nucleus = $(79 \times 1.60 \times 10^{-19}\,\text{C})$

Answer

Kinetic energy of alpha particle, $E_k = k\dfrac{q_\alpha q_n}{r}$

$$(5.0 \times 10^6) \times (1.60 \times 10^{-19}) = \frac{((8.99 \times 10^9) \times (2 \times 1.60 \times 10^{-19}) \times (79 \times 1.60 \times 10^{-19}))}{r}$$

$r = 4.5 \times 10^{-14}\,\text{m}$

Considering that the alpha particle has a very large amount of kinetic energy (for its small mass) and would therefore be expected to be able to get close to a nucleus, this type of calculation was the first to provide some evidence for the possible size of a nucleus. However, a gold nucleus is smaller than the value shown above. (The actual radius of a gold nucleus is about 0.7×10^{-14} m.)

Nuclear radii

◆ **Nuclear radius, R** R is proportional to the cube root of the nucleon number. $R = R_0A^{\frac{1}{3}}$, where R_0 is called the **Fermi radius**.

◆ **Rutherford scattering** Sometimes called Coulomb scattering. The scattering of alpha particles by nuclei, which can only be explained by the action of an inverse square law of electric repulsion. When high-energy particles are used they might enter the nucleus, so that strong nuclear forces are also involved and then the scattering will no longer follow the same pattern.

◆ **Nuclear density** All nuclear densities are similar in magnitude and are extremely large.

Rutherford scattering experiments (and similar) have shown that the radius, R, of any nucleus is proportional to the cube root of its nucleon number, A:

radius of nucleus $R = R_0A^{\frac{1}{3}}$

This has implications for nuclear density, as seen below.

The constant R_0 **= 1.20 × 10^{-15} m** is called the **Fermi radius**, which is the assumed radius of a nucleus with only one proton ($A = 1$).

WORKED EXAMPLE E1.6

Show that the radius of a $^{197}_{79}$Au (gold) nucleus is 'about 7.0 × 10^{-15} m'.

Answer

$R = R_0A^{⅓} = (1.20 \times 10^{-15}) \times 197^{⅓} = 6.98 \times 10^{-15}$ m

Nuclear density

We can use $\rho = m/V$ to estimate **nuclear density**. The mass of a nucleus will be approximately equal to Au, where u can be considered as the average mass of a nucleon, 1.661 × 10^{-27} kg. (u is called the *atomic mass unit*. It is explained in Topic E.3.)

$$\rho = \frac{Au}{\frac{4}{3}\pi\left(R_0A^{\frac{1}{3}}\right)^3} = \frac{3u}{4\pi R_0^{\ 3}}$$

Importantly, this result shows us that nuclear densities do *not* depend on the radius of the nucleus, or the number of nucleons.

The densities of all nuclei are approximately the same.

$$\rho = \frac{(3 \times (1.661 \times 10^{-27}))}{(4\pi \times (1.20 \times 10^{-15})^3)} = 2.3 \times 10^{17}\,\text{kg}\,\text{m}^{-3}$$

This is an extremely large density! If the electrons in an atom are considered to have negligible mass compared to the nucleons, and the radius of an atom is typically 10^5 times larger than a nucleus, then an order of magnitude density for atoms would be:

$$\frac{10^{17}}{(10^5)^3} \approx 10^2\,\text{kg}\,\text{m}^{-3}$$

which is comparable to everyday observations of the density of matter, as expected. The only macroscopic objects with densities comparable to nuclear densities are collapsed massive stars, known as neutron stars and black holes.

26 Determine the closest distance that an alpha particle of energy 1.37 MeV could approach to:

a a gold nucleus

b a copper nucleus.

Copper has a proton number of 29.

27 An alpha particle nearly collides with a gold nucleus and returns along the same path.

Sketch a graph showing how the electric potential energy and kinetic energy possessed by the alpha particle vary with the distance of the alpha particle from the gold nucleus (assumed to be stationary).

28 a Calculate the velocity at which an alpha particle (mass of 6.64 × 10^{-27} kg) should travel directly towards the nucleus of a gold atom (charge +79e) in order to get within 2.7 × 10^{-14} m of it. Assume the gold nucleus remains stationary.

b Calculate the energy (MeV) of an alpha particle with this velocity.

29 Explain why alpha particles were used in the Geiger–Marsden–Rutherford experiments.

30 Discuss whether it is reasonable to assume that when an alpha particle approaches a gold atom in thin foil:

- **a** the repulsive force from the alpha particle is the only force acting on the gold nucleus
- **b** the gold atom remains stationary.

31 a Estimate the radius of:

  - **i** a gold nucleus ($A = 197$)
  - **ii** an oxygen nucleus ($A = 16$).

b The measured radius of a gold-197 nucleus is 6.87×10^{-15} m. How does this compare with the value calculated in part **a**?

32 Show that the largest possible nuclear radius (of a naturally occurring element) is only about six times the smallest.

33 Gold is considered to be a dense element. Estimate what fraction of the volume of a gold ring is actually occupied by the subatomic particles. Assume that the radius of a gold atom is 1.5×10^{-10} m. State any other assumptions that you make.

34 The mass of the Sun is 2.0×10^{30} kg and its radius is 7.0×10^{8} m.

- **a** Estimate the radius of a neutron star that has twice the mass of the Sun. Assume that a neutron star has the same density as a nucleus.
- **b** Compare your answer to the radius of the Sun.

The strong nuclear force

SYLLABUS CONTENT

- Deviations from Rutherford scattering at high energies.

Up to this point, we have assumed that the only force acting between the alpha particle and a nucleus is a repulsive electric force between positive charges. Rutherford scattering is sometimes called Coulomb scattering because Coulomb's law can be used to describe it.

However, there is another field force acting between individual nucleons when they are close together: the strong nuclear force. This is the attractive force that overcomes the repulsive forces between positively charged protons as mentioned earlier in this topic.

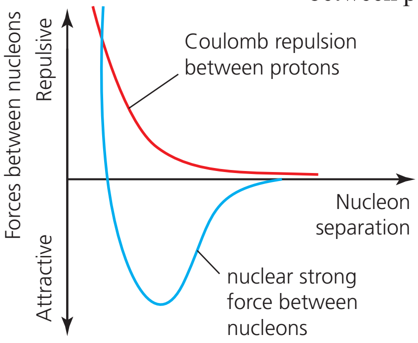


Figure E1.19 How the strong nuclear force varies with distance between two protons

A very energetic alpha particle can get close enough to the nucleons that it is affected by the attractive strong nuclear force as well as the repulsive electric force.

If this happens, the scattering can no longer be explained simply by Coulomb's law.

Figure E1.19 approximately compares the strong nuclear force to the electric force between two protons. The electric repulsion force dominates for separations greater than about 3×10^{-15} m, whereas the strong nuclear force is 'short range' and only becomes significant for separations less than about 1.5×10^{-15} m. At that separation, the strong force attracts the protons together, but if they get much closer the force becomes repulsive.

The Bohr model of the hydrogen atom

SYLLABUS CONTENT

- The discrete energy levels in the Bohr model for hydrogen as given by: $E = \frac{-13.6}{n^2}$ eV.
- The existence of quantized energy and orbits arise from the quantization of angular momentum in the Bohr model for hydrogen as given by: $mvr = \frac{nh}{2\pi}$.

HL ONLY

◆ **Energy levels of hydrogen** Because hydrogen is the atom with the simplest structure, scientists were very interested in determining the energy levels of the electron within the atom by examining hydrogen's line spectrum. They were able to show that the energy levels could be predicted by the empirical equation $E = 13.6n^2$.

◆ **Principal quantum number, *n*** Number used to describe the energy level of an atom. The lowest energy level is called the ground state, with $n = 1$, the next level has $n = 2$, and so on.

◆ **Bohr model** A theory of atomic structure that explains the spectrum of hydrogen atoms.

The hydrogen atom is the simplest atom and, as such, it was at the centre of attention when physicists were beginning to understand atomic structure. The ideas in this section can be expanded to include other atoms, but we will restrict discussion to hydrogen.

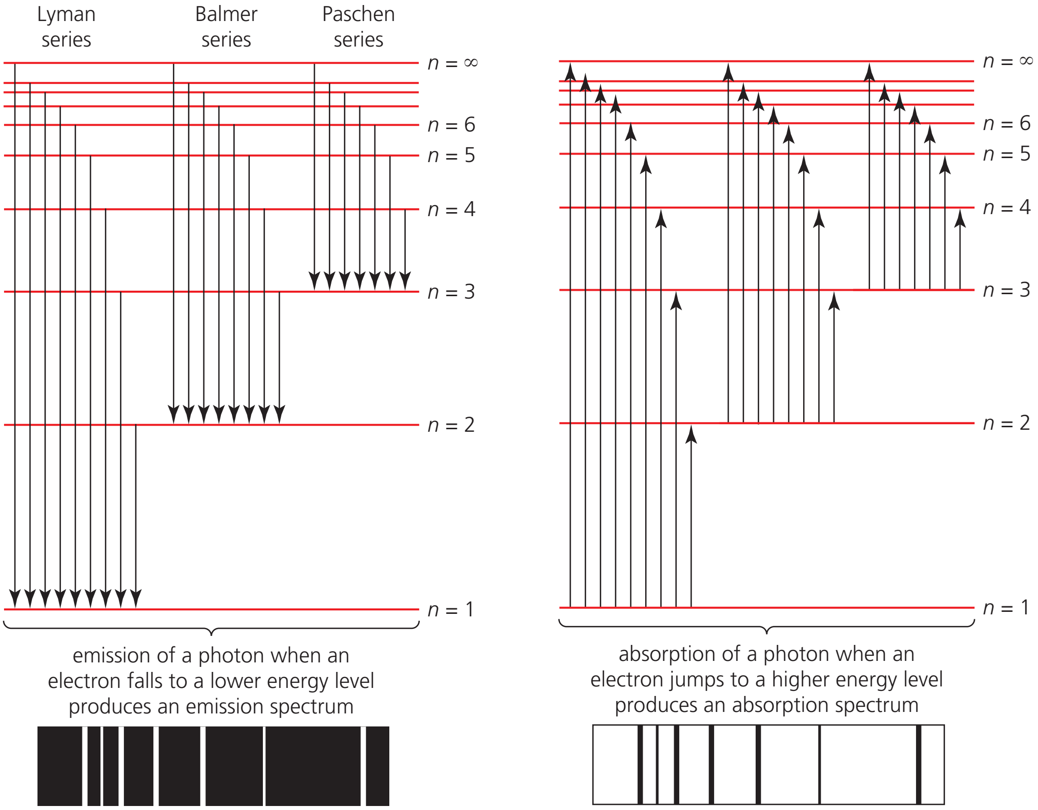
Figure E1.20 shows the **energy levels of hydrogen** again, this time with all the possible energy transitions, arranged into three groups.

All the possible levels are shown and they are numbered, beginning with the lowest level, the ground state as $n = 1$. n is called the **principal quantum number**.

The energy levels in hydrogen atoms get closer together as n increases. This enables the possible transitions to be grouped as shown. All transitions down to the ground state ($n = 1$) are larger than all transitions down to the level $n = 2$, but note that Figure E1.20 is not drawn to scale. All transitions down to level $n = 2$ are larger than all transitions down to the level $n = 3$. (You do not need to remember the names of these three series. The transitions of the Lyman series all produce photons of ultraviolet radiation, the transitions of the Balmer series all produce photons of visible light, the transitions of the Paschen series all produce photons of infrared radiation.)

It is easily confirmed that all the discrete energy levels of hydrogen can be predicted by:

energy levels of hydrogen $\boldsymbol{E} = \dfrac{\mathbf{-13.6}}{\boldsymbol{n}^{\mathbf{2}}}$ eV (electronvolts)



■ **Figure E1.20** The Lyman, Balmer and Paschen series of the hydrogen atom

WORKED EXAMPLE E1.7

Calculate a value for the fifth energy level of the hydrogen atom ($n = 5$) in:

a electronvolts **b** joules.

Answer

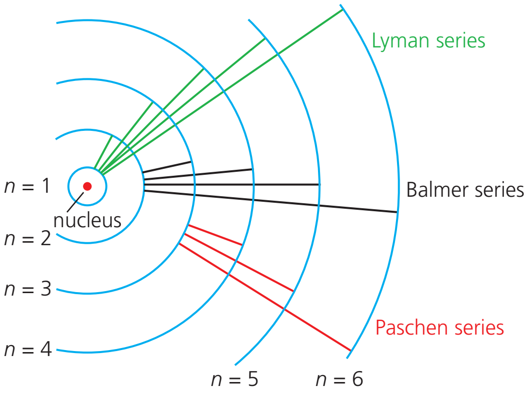
a $E = \frac{-13.6}{n^2}\text{ eV} = \frac{-13.6}{5^2} = -0.544\text{ eV}$

b $-0.544 \times 1.60 \times 10^{-19} = -8.70 \times 10^{-20}\text{ J}$

Figure E1.21 Niels Bohr

The equation highlighted above was empirical, not based on any theory, when it was first discovered by the Danish physicist Neils Bohr (Figure E1.21). Of course, scientists wanted an explanation of *why* atoms had discrete energy levels and why the energy levels of hydrogen were predicted by a simple equation.

The **Bohr model** of the atom, first proposed in 1913, has electrons orbiting around the nucleus due to the centripetal force provided by electric attraction between opposite charges. But the essential feature of the Bohr model was that it restricted the orbits to only certain distances from the nucleus and, most importantly, because they remained in that orbit, they did not emit electromagnetic radiation, lose energy and spiral inwards. It was known that accelerated charges emit electromagnetic radiation, remembering that moving in a circle involves a centripetal acceleration.



not to scale

Figure E1.22 The Bohr model explains the spectrum of Hydrogen using possible orbits of different radii for the electron

In the Bohr model, each electron orbit had a definite and precise energy, and intermediate energies were not possible. Photons were emitted or absorbed when electrons moved between these energy levels. See Figure E1.22, in which the radii of the orbits are not drawn to scale.

Bohr showed that the quantized radii, r, of possible orbits of an electron of mass, m, moving with speed, v, in a hydrogen atom can be calculated from the equation:

$$mvr = \frac{nh}{2\pi}$$

The product of linear momentum and radius, mvr, is the *angular momentum, L*, of the electron about the nucleus (see Topic A.4). h, as before, is Planck's constant.

WORKED EXAMPLE E1.8

Calculate three possible values for the angular momentum of an electron in a hydrogen atom.

Answer

Angular momentum, $L = \frac{nh}{2\pi}$

For $n = 1$, $L = 1 \times \frac{(6.63 \times 10^{-34})}{(2 \times 3.14)} = 1.06 \times 10^{-34}\text{ kg m}^2\text{ s}^{-1}$

For $n = 2$, $L = 2.11 \times 10^{-34}\text{ kg m}^2\text{ s}^{-1}$

For $n = 3$, $L = 3.17 \times 10^{-34}\text{ kg m}^2\text{ s}^{-1}$

As the following mathematics shows, the Bohr model combined the classical physics of circular motion and the force of electric attraction with quantum concepts in order to predict the radii of orbits and the energy levels in hydrogen atoms.

Equating the centripetal force on the electron to the electric attraction between it and the proton in a hydrogen atom, remembering that both the electron and the proton have charges of the same magnitude, e, we get:

$$\frac{mv^2}{r} = \frac{kee}{r^2}$$

(the signs of the charges are not relevant here); which leads to:

$$v = \sqrt{\left(\frac{ke^2}{mr}\right)}$$

Putting this expression for v in the equation for angular momentum (highlighted above), we get:

$$\sqrt{ke^2mr} = \frac{nh}{2\pi}$$

and rearranging enables us to obtain an expression for r (which need not be remembered):

$$r = \frac{n^2h^2}{4\pi^2ke^2m}$$

Putting in various values for n enables us to correctly predict the radii of possible electron orbits within the hydrogen atom.

WORKED EXAMPLE E1.9

Determine a value for the radius of the electron's orbit when the hydrogen atom is in its ground state ($n = 1$).

Answer

$$r = \frac{n^2h^2}{4\pi^2ke^2m} = \frac{\left(1^2 \times (6.63 \times 10^{-34})^2\right)}{\left(4 \times \pi^2 \times (8.99 \times 10^9) \times (1.60 \times 10^{-19})^2 \times (9.110 \times 10^{-31})\right)}$$

$$= 5.3 \times 10^{-11}\ \text{m}$$

Then, using equations for electrical potential energy and kinetic energy, the total energy associated with the ground state of the atom can be calculated as follows.

The total energy of the hydrogen–electron system, a hydrogen atom, can be determined as follows:

$$E_{\text{total}} = E_{\text{k}} \text{ of electron} + E_{\text{k}} \text{ of proton} + \text{electric potential energy}$$

assuming that the electron (charge $-e$, mass m) is orbiting at speed v in a circular orbit of radius r around a proton (charge $+e$) which is effectively stationary:

total energy of a hydrogen atom, $E_{\text{total}} = \frac{1}{2}mv^2 + \left(-\frac{ke^2}{r}\right)$

But, we know that the magnitude of the centripetal force:

$$\frac{mv^2}{r} = \frac{ke^2}{r^2} \text{ (see above), or } mv^2 = \frac{ke^2}{r}$$

Top tip!

The calculation on the right has determined a value for the ground state energy level of the hydrogen *atom*. It has clearly involved the electric potential energy of the electron–proton system. Nevertheless, it is common to refer to the energy levels of *electrons* within hydrogen (and other) atoms. This is understandable because we commonly visualize electrons moving between orbits when an atom receives or emits energy.

Leading to:

$$E_{\text{total}} = \frac{1}{2}\frac{ke^2}{r} + \left(-\frac{ke^2}{r}\right) = -\frac{1}{2}\frac{ke^2}{r}$$

For the ground state, $r = 5.3 \times 10^{-11}$ m (as above), so that:

$$E_{\text{total}} = -\frac{1}{2}\frac{ke^2}{r} = -2.2 \times 10^{-18}\ \text{J}$$

So, Bohr's quantization of angular momentum equation leads directly to a correct calculation of the hydrogen atom ground state (and other energy levels).

35 Consider Figure E1.20.

- **a** Calculate the lowest frequency of the Balmer series.
- **b** In which part of the electromagnetic spectrum is this radiation?

36 Determine the energy (J) of the level with $n = 8$ in the hydrogen atom.

37 Determine the angular momentum of an electron in a hydrogen atom if it has a principal quantum number of four:

- **a** in terms of h/π
- **b** in SI units.

38 Calculate the radius of the first electron orbit above the ground state of a hydrogen atom.

39 Determine the total energy of a hydrogen atom if its electron is in the energy level which has a principal quantum number of three.

LINKING QUESTION

- Under what circumstances does the Bohr model fail? (NOS)

Although Bohr's quantized model of the atom was a big step forward in understanding the structure of the atom, and it was very accurate in predicting the energy levels of one-electron atoms such as a hydrogen atom or a helium ion, it was less successful with atoms containing more electrons. Furthermore, the reasons for the existence of energy levels were still not understood. The Bohr model still remains an important initial step for students learning about quantization in atoms, but the discovery of the wave properties of electrons (Topic E.2) quickly led to dramatic changes in physicists' understanding of the atom, but there are not included in the IB course.

◆ **Quantum mechanics** The mathematical aspects of quantum physics

Quantum mechanics is the name given to the important branch of physics that deals mathematically with events on the atomic and subatomic scales that involve quantities that can only have discrete (quantized) values. In the quantum world the laws of classical physics are often of little use; trying to apply knowledge and intuition gained from observing the macroscopic world around us often only leads to confusion.

E.2 Quantum physics

Guiding questions

- How can light be used to create an electric current?
- What is meant by wave–particle duality?

Nature of science: Theories

What is quantum physics?

A *quantum* is the general term used to describe the minimum amount of any physical quantity that can only exist in discrete quantities (which are all basic multiples of one quantum).

On the subatomic scale, we have seen that charge is quantized, and in Topic E.1 we discussed the quantization of atomic energy levels, angular momentum and electromagnetic radiation. In a more general sense, matter itself could be considered as quantized because it is made of discrete particles, rather than being continuous.

Quantum ideas are so fundamental to understanding the behaviour of subatomic particles and waves (which, of course, also affects everything in our macroscopic world), that this branch of science has become generally known as quantum physics, and its detailed mathematic treatment is called quantum mechanics.

A famous quote from Neils Bohr: *'Everything that we call real is made of things that cannot be regarded as real. If quantum mechanics has not profoundly shocked you, you haven't understood it yet.'*

The photoelectric effect

SYLLABUS CONTENT

- The photoelectric effect as evidence of the particle nature of light.
- Photons of a certain frequency, known as the threshold frequency, are required to release photoelectrons from the metal.
- Einstein's explanation using the work function and the maximum kinetic energy of the photoelectrons as given by: $E_{max} = hf - \Phi$, where Φ is the work function of the metal.

◆ **Photoelectric effect** Ejection of electrons from a substance by incident electromagnetic radiation.

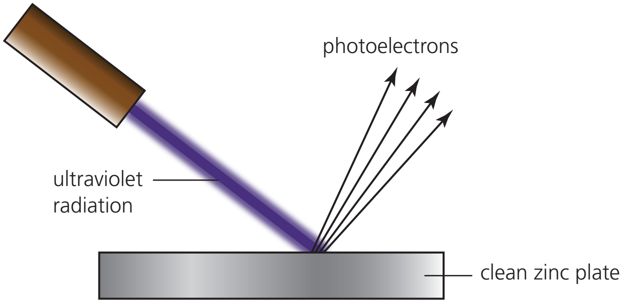
◆ **Photoelectrons** Electrons ejected in the process of the photoelectric effect.

The **photoelectric effect** (described below) was first discovered by Hertz in 1887. Eighteen years later, in 1905, Einstein expanded ideas about quantized energy ($E = hf$) proposed by Max Planck five years earlier, to propose that light and other electromagnetic radiation consisted of individual bundles of energy, thereby explaining the photoelectric effect. These events marked the beginnings of quantum physics. We will start by giving details of the photoelectric effect.

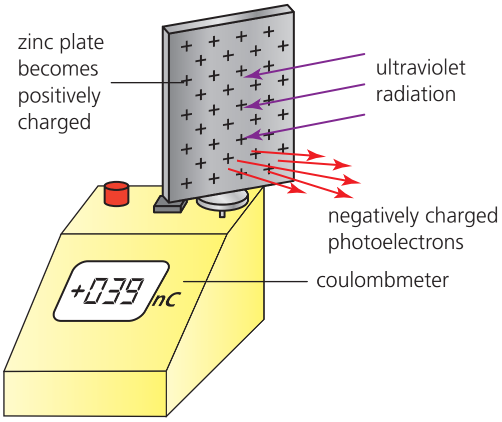
When electromagnetic radiation is directed onto a clean surface of some metals, electrons may be ejected. This is called the **photoelectric effect** and the ejected electrons are known as **photoelectrons**. The principle is shown in Figure E2.1. The importance of this effect lies in understanding why photoelectrons are emitted under some circumstances, but not others.

Under suitable circumstances, the photoelectric effect can occur with visible light, X-rays or gamma rays, but it is most often demonstrated with ultraviolet radiation and zinc. A possible arrangement is shown in Figure E2.2.

Ultraviolet radiation is shone onto a zinc plate attached to a coulombmeter (an instrument which measures very small quantities of charge). The ultraviolet radiation causes the zinc plate to become positively charged because some negatively charged electrons on the (previously neutral) zinc plate have gained enough kinetic energy to escape from the surface.



■ **Figure E2.1** The photoelectric effect – a stream of photoelectrons is emitted from a metal surface illuminated with ultraviolet radiation



■ **Figure E2.2** Demonstration of the photoelectric effect

◆ **Threshold frequency, f_0** The minimum frequency of a photon that can eject a photoelectron from the surface of a metal.

Investigations of the photoelectric effect show a number of important observations.

- If the intensity of the radiation is increased, the charge on the plate increases more quickly because more photoelectrons are being released every second.
- There is no time delay between the radiation reaching the metal surface and the emission of photoelectrons. The release of photoelectrons from the surface appears to be instantaneous.
- The photoelectric effect can only occur if the frequency of the radiation is above a certain minimum value. The lowest frequency for emission is called the **threshold frequency, f_0**. (Alternatively, we could say that there is maximum wavelength above which the effect will not occur.) If the frequency used is lower than the threshold frequency, the effect will not occur even if the intensity of the radiation is greatly increased. The threshold frequency of zinc, for example, is 1.04×10^{15} Hz, which is in the ultraviolet part of the spectrum. Visible light will not release photoelectrons from zinc (or other common metals).
- For a given incident frequency, the photoelectric effect occurs with some metals but not with others. This is because different metals have different threshold frequencies.
- The photoelectrons emitted from a particular metal by monochromatic radiation of a known frequency have a range of kinetic energies, up to a well-defined maximum.

Explaining the photoelectric effect: the Einstein model

If we tried to use the wave theory of radiation to make predictions about the photoelectric effect, we would expect the following. (1) Radiation of any frequency will cause the photoelectric effect if the intensity is made high enough. (2) There may be a delay before the effect begins because it needs time for enough energy to be provided (similar to heating water until it boils).

These predictions are wrong, so an alternative theory is needed. Einstein realized that we cannot explain the photoelectric effect without first understanding the quantum nature of radiation.

The Einstein model explains the photoelectric effect using the concept of photons.

When a photon in the incident radiation interacts with an electron in the metal surface, it transfers all of its energy to that electron. It should be stressed that a *single* photon interacts with a *single* electron and that this transfer of energy is instantaneous; there is no need to wait for a build-up of energy. If a photoelectric effect is occurring, increasing the intensity of the radiation only increases the number of photons and photoelectrons, not their individual energies.

Einstein realized that some of the energy carried by the photon was used to overcome the attractive forces that normally keep an electron within the metal surface. The remaining energy is transferred to the kinetic energy of the newly released photoelectron. Using the principle of conservation of energy, we can write:

energy carried by photon
= work done in removing the electron from the surface + kinetic energy of photoelectron

But the energy required to remove different electrons from the same surface is not always the same. It will vary with the position of the electron with respect to the surface. Electrons closer to the surface will require less energy to remove them. However:

there is a well-defined minimum amount of energy needed to remove an electron from the surface of any particular metal and this is called the **work function**, Φ, of the metal.

◆ **Work function, Φ** The minimum amount of energy required to free an electron from the attraction of ions in a metal's surface.

◆ **Photoelectric equation** The maximum kinetic energy of an emitted photoelectron is the difference between the incident photon's energy and the work function, Φ: $E_{max} = hf - \Phi$.

Different metals have different values for their work functions. For example, the work function of a clean zinc surface is 4.3 eV. This means that at least 4.3 eV (= 6.9×10^{-19} J) of work must be done to remove an electron from zinc.

To understand the photoelectric effect, we need to compare the photon's energy, hf, to the work function, Φ, of the metal:

If $hf < \Phi$

If an incident photon has less energy than the work function of the metal, the photoelectric effect cannot occur. Radiation that may cause the photoelectric effect with one metal may not have the same effect with another (which has a different work function).

If $hf\,(= hf_0) = \Phi$

At the threshold frequency, f_0, the incident photon has exactly the same energy as the work function of the metal. We may assume that the photoelectric effect occurs, but any released photoelectrons will have zero kinetic energy.

If $hf > \Phi$

If an incident photon has more energy than the work function of the metal, the photoelectric effect occurs and a photoelectron will be released. Photoelectrons produced by different photons (of the same frequency) will have a range of different kinetic energies because different amounts of work will have been done to release them.

It is important to consider the situation in which the minimum amount of work is done to remove an electron (equal to the work function):

energy carried by photon = work function + *maximum* kinetic energy of photoelectrons

Or in symbols: $hf = \Phi + E_{max}$.

Or:

$$E_{max} = hf - \Phi$$

This equation is often called Einstein's **photoelectric equation**.

Because $hf_0 = \Phi$, we can also write this as: $E_{max} = hf - hf_0$.

Figure E2.3 shows a graphical representation of how the maximum kinetic energy of the emitted photons varies with the frequency of the incident photons. The equation of the line is $E_{max} = hf - \Phi$, as above.

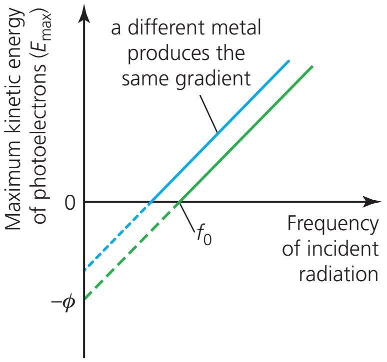


Figure E2.3 Theoretical variation of maximum kinetic energy of photoelectrons with incident frequency (for two different metals)

Tool 3: Mathematics

Interpret features of graphs including gradient, intercepts

Any straight line on an x–y graph can be represented by the equation $y = mx + c$, where m is the gradient ($\Delta y/\Delta x$) and c is the intercept on the y-axis ($y = c$, when $x = 0$). See Figure E2.4. The value of the intercept on the x-axis: $x = -c/m$ when $y = 0$.

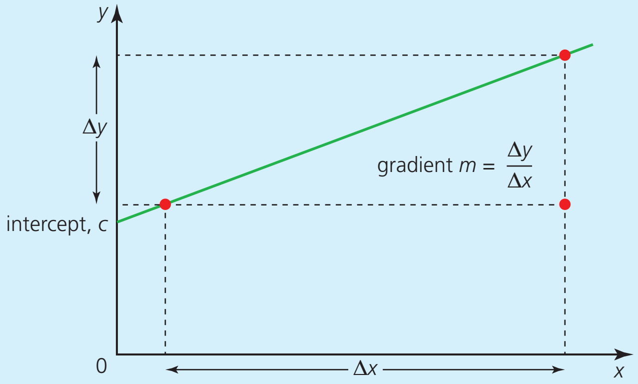


Figure E2.4 The line $y = mx + c$

We can take the following measurements from graphs of the form seen in Figure E2.3:

- The gradient of the line is equal to Planck's constant, h. The gradient will be the same for all circumstances because it does not depend on photon frequencies, or the metal used.
- The intercept on the frequency axis gives us the value of the threshold frequency, f_0.
- A value for the work function can be determined from: when $E_{max} = 0$, $\Phi = hf_0$; or when $f = 0$, $\Phi = -E_{max}$.

WORKED EXAMPLE E2.1

Radiation of wavelength 5.59×10^{-8} m was incident on a metal surface that had a work function of 2.70 eV.

a Calculate the frequency of the radiation.

b Determine how much energy is carried by one photon of the radiation.

c Calculate the value of the work function expressed in joules.

d Explain whether the photoelectric effect occurs under these circumstances.

e Determine the maximum kinetic energy of the photoelectrons.

f Calculate the threshold frequency for this metal.

g Sketch a fully labelled graph to show how the maximum kinetic energy of the photoelectrons would change if the frequency of the incident radiation was varied.

Answer

a $f = c/\lambda = \dfrac{(3.00 \times 10^{8})}{(5.59 \times 10^{-8})} = 5.37 \times 10^{15}\,\text{Hz}$

b $E = hf = (6.63 \times 10^{-34}) \times (5.37 \times 10^{15}) = 3.56 \times 10^{-18}\,\text{J}$

c $2.70 \times (1.60 \times 10^{-19}) = 4.32 \times 10^{-19}\,\text{J}$

d Yes, because the energy of each photon is greater than the work function.

e $E_{max} = hf - \Phi = (3.56 \times 10^{-18}) - (4.32 \times 10^{-19}) = 3.13 \times 10^{-18}\,\text{J}$

f $\Phi = hf_0$

$$f_0 = \frac{4.32 \times 10^{-19}}{6.63 \times 10^{-34}} = 6.52 \times 10^{-14}\,\text{Hz}$$

g The graph should be similar to Figure E2.3, with numerical values provided for the intercepts.

1 Repeat Worked example E2.1 but for radiation of wavelength 6.11×10^{-7} m incident on a metal with a work function of 2.21 eV. Omit part **e**.

2 a Outline how Einstein used the concept of photons to explain the photoelectric effect.
  b Explain why a wave model of electromagnetic radiation is unable to explain the photoelectric effect.

3 The threshold frequency of a metal is 7.0×10^{14} Hz. Calculate the maximum kinetic energy of the electrons emitted when the frequency of the radiation incident on the metal is 1.0×10^{15} Hz.

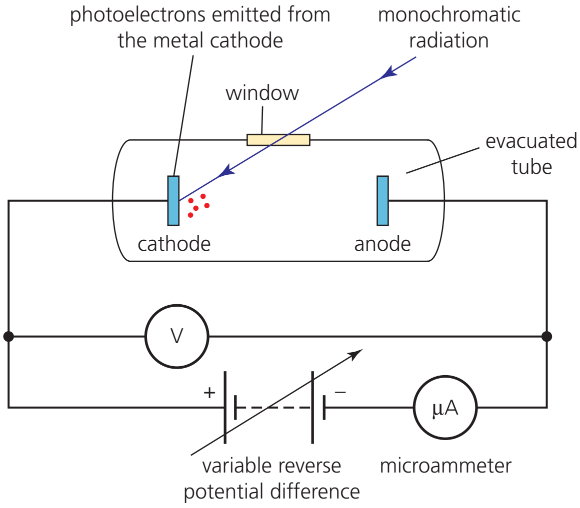
4 a The longest wavelength that emits photoelectrons from potassium is 550 nm.
  Calculate the work function (in joules).
  b Determine the threshold wavelength for potassium. What is the name given to this kind of radiation?
  c State one colour of visible light that will not produce the photoelectric effect with potassium.

5 When electromagnetic radiation of frequency 2.90×10^{15} Hz is incident on a metal surface, the emitted photoelectrons have a maximum kinetic energy of 9.70×10^{-19} J.
Calculate the threshold frequency of the metal.

Experiments to test the Einstein model

Investigating stopping voltages (potential differences)

To test Einstein's equation (model) for the photoelectric effect, it is necessary to determine the maximum kinetic energy of the photoelectrons emitted under a variety of different circumstances. In order to do this the kinetic energy must be transferred to another (measurable) form of energy.



■ **Figure E2.5** Experiment to test Einstein's model of photoelectricity

The kinetic energy of the photoelectrons can be transferred to electrical potential energy if they are repelled by a negative voltage. This experiment was first performed by the American physicist Robert Millikan and a simplified version is shown in Figure E2.5.

Ideally *monochromatic* radiation should be used, but it is also possible to use a narrow range of frequencies such as those obtained by using coloured filters with white light.

When radiation is incident on a suitable emitting surface, photoelectrons will be released with a range of different energies, as explained previously. Because it is emitting negative charge, this surface can be described as a *cathode* (the direction of conventional current flow will be out of a cathode and around the circuit). Any photoelectrons that have enough kinetic energy will be able to move across the tube and reach the other electrode, the *anode*. The tube is evacuated (the air is removed to create a vacuum) so that the electrons do not collide with air molecules during their movement across the tube.

The most important thing to note about this circuit is that the (variable) source of potential difference is connected the 'wrong way around'. We say that it is supplying a *reverse* potential difference across the tube. This means that there is a negative voltage on the anode that will repel the photoelectrons. Photoelectrons moving towards the anode will have their kinetic energy reduced as it is transferred to electrical potential energy. (Measurements for positive voltages can be made by reconnecting the battery the 'correct' way.)

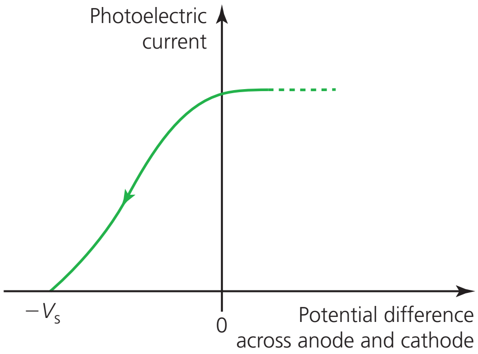


Figure E2.6 Increasing the reverse potential difference decreases the photoelectric current

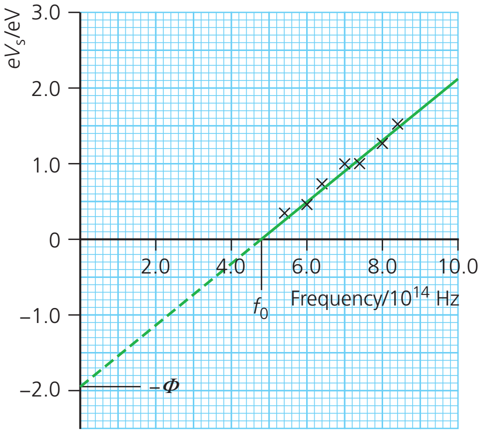


Figure E2.7 Experimental results showing variation of maximum energy (eV_s) of photoelectrons with incident frequency

Any flow of charge across the tube and around the circuit can be measured by a sensitive microammeter. When the reverse voltage on the anode is increased from zero, more and more photoelectrons will be prevented from reaching the anode and this will decrease the current. (Remember that the photoelectrons have a range of different energies.) Eventually the reverse potential difference will be large enough to stop even the most energetic of photoelectrons, and the current will fall to zero (Figure E2.6).

The potential difference across the tube needed to *just* stop all photoelectrons reaching it is called the **stopping voltage** (p.d.), V_s.

Because, by definition, potential difference = energy transferred / charge, after measuring V_s we can use the following equation to calculate values for the maximum kinetic energy of photoelectrons under a range of different circumstances: $E_{max} = eV_s$.

For convenience, it is common to quote all energies associated with the photoelectric effect in electronvolts (eV). In which case, the maximum kinetic energy of the photoelectrons is numerically equal to the stopping voltage. That is, if the stopping voltage is, say, 3 V, then $E_{max} = 3$ eV.

Einstein's equation ($E_{max} = hf - \Phi$) can be rewritten as: $eV_s = hf - \Phi$.

By experimentally determining the stopping voltage for a range of different frequencies, the theoretical graph shown previously in Figure E2.3 can now be confirmed by plotting a graph from actual data, as shown in Figure E2.7.

◆ **Stopping voltage** The minimum voltage required to reduce a photoelectric current to zero.

WORKED EXAMPLE E2.2

Use Figure E2.7 to determine:

a the threshold frequency

b the work function

c a value for Planck's constant.

Answer

a f_0 can be determined from the intercept on the frequency axis: $f_0 = 4.8 \times 10^{14}$ Hz

b from the intercept on the eV_s axis: 1.9 eV

c h can be determined from the gradient (remembering to convert electronvolts to joules):

$$h = \frac{(2.1 \times 1.60 \times 10^{-19})}{((10 - 4.8) \times 10^{14})} = 6.5 \times 10^{-34}\ \text{Js}$$

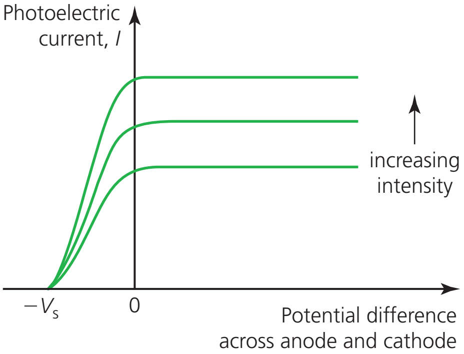
Common mistake

If two sources of monochromatic radiation have the same intensity, but different frequencies, the photons from the source with the lower frequency will carry less energy, so there must be more photons emitted every second from that source.

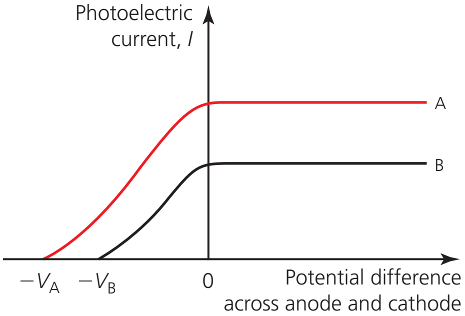
Investigating photoelectric currents

Using apparatus similar to that shown in Figure E2.5, it is also possible to investigate quantitatively the effects on the photoelectric current of changing the:

- intensity
- frequency
- metal used in the cathode.



■ **Figure E2.8** Variation of photoelectric current with potential difference for radiation of three different intensities (same frequency)



■ **Figure E2.9** Variation of photoelectric current with potential difference for radiation of two different frequencies

Intensity

Figure E2.8 shows the photoelectric currents produced by monochromatic radiation of the same frequency with three different intensities.

For positive p.d.s, each of the photoelectric currents remain constant because the photoelectrons are reaching the anode at the same rate as they are being produced at the cathode, and this does not depend on the size of the positive voltage on the anode. Greater intensities (of the same frequency) produce higher photoelectric currents because there are more photons releasing more photoelectrons (with the same range of energies). Because the maximum kinetic energy of photons depends only on frequency and and work function, but not intensity, all these graphs have the same value for stopping voltage, V_s.

Frequency

Figure E2.9 shows the photoelectric currents produced by radiation from two monochromatic sources of different frequencies, A and B, incident on the same metal.

The individual photons in radiation A must have more energy (than B) and produce photoelectrons with a higher maximum kinetic energy. We know this because a greater reverse voltage is needed to stop the more energetic photoelectrons produced by A. No conclusion can be drawn from the fact that the current for A has been drawn higher than for B, because the intensities of the two radiations are not known. In the unlikely circumstances that the two intensities were equal, the maximum current for B would have to be higher than for A because the radiation from B must have more photons, because each photon has less energy than in A.

Metal used in the cathode

Experiments confirm that when different metals are tested using the same frequency, the photoelectric effect is observed with some metals but not with others (those metals for which their work function is higher than the energy of the photons).

6 Calculate the maximum kinetic energy of photoelectrons emitted from a metal if the stopping voltage was 2.4 V. Give your answer in electronvolts and in joules.

7 Make a copy of Figure E2.6 and sketch lines to show the results that would be obtained with:

- **a** the same radiation, but with a metal of higher work function (assume that the photoelectric effect still occurs)
- **b** the original metal and the same frequency of radiation but using radiation with a greater intensity.

8 In an experiment using monochromatic radiation of frequency 7.93×10^{14} Hz with a metal that had a threshold frequency of 6.11×10^{14} Hz, it was found that the stopping voltage was 0.775 V.
Calculate a value for Planck's constant from these results.

9 Make a copy of Figure E2.6 and sketch the results that would be obtained using radiation of a higher intensity (of the same frequency) incident on a metal that has a smaller work function.

10 Make a copy of Figure E2.6. Add to it a line showing the results that would be obtained with radiation of a higher frequency but with same number of photons every second incident on the same metal.

11 a Select **five** different metallic elements and then use the internet to research their work functions.

- **b** Calculate the threshold frequencies of the five metals.

Inquiry 2: Collecting and processing data

Processing data

Light emitting diodes (LEDs) can be used to determine an approximate value for Planck's constant. Each LED emits photons of a precise frequency. The energy of each photon ($E = hf$) is transferred when an individual electron is accelerated by the voltage across the LED. That is, $eV = hf$.

If the voltage across the LED that *just* results in the LED emitting light is measured, then this equation can be used to determine a value for h if the frequency of the radiation is known. However, it is better to draw a graph.

1 Draw a voltage–frequency graph of the results shown in Table E2.1 and use the gradient to determine a value for Planck's constant.

■ **Table E2.1**

Colour	Frequency / 10^{14} Hz	Voltage / V
red	4.54	1.91
amber	5.01	2.06
yellow	5.10	2.12
green	5.37	2.21
blue	6.37	2.65

The wave nature of matter

SYLLABUS CONTENT

- Diffraction of particles as evidence of the wave nature of matter.
- The de Broglie wavelength for particles as given by: $\lambda = \frac{h}{p}$.

The fact that light and other electromagnetic waves could behave as particles (photons) raises an obvious question: can particles behave like waves?

In 1924 the French physicist Louis de Broglie proposed that electrons, which were thought of as particles, might also have a wave-like character. He later generalized his hypothesis to suggest that:

All moving particles have wave-like properties.

◆ **de Broglie's hypothesis** All particles exhibit wave-like properties, with a de Broglie wavelength, $\lambda = \frac{h}{p}$

◆ **Hypothesis** A suggested explanation of a phenomenon (but not proven).

According to **de Broglie's hypothesis**, the wavelength, λ, of a moving particle was inversely proportional its momentum, p, as represented by:

$$\text{wavelength of a moving particle } \boldsymbol{\lambda} = \frac{\boldsymbol{h}}{\boldsymbol{p}}$$

Once again, we can see the importance of Planck's constant, h, in predicting the size of quantum phenomena. The very small value of Planck's constant shows us that wave properties of particles are only significant for those with tiny momenta, as Worked example E2.3 illustrates.

Nature of science: hypotheses and theories

A **hypothesis** uses limited information to suggest a possible outcome or explanation. A hypothesis is not assumed to be true until it has been tested. For example, student investigations will usually begin with a hypothesis about what they *think* will happen in their experiments. Many scientific advances are preceded by hypotheses.

An explanation that has been confirmed from much-repeated experiments / observations is usually described as a *theory*.

It is reported that Isaac Newton was strongly opposed to the use of hypotheses, and that he believed that theories should be inferred directly from observations. But experimentation was not such a major feature of science at that time.

Despite the fact that it has been confirmed that all moving particles have wave properties, de Broglie's work is still usually described as a 'hypothesis' rather than a theory. Suggest why.

ATL E2A: Social skills

Working collaboratively to achieve a common goal

Working in small groups, discuss one of the following, and then summarize your conclusions for the rest of your class.

- Are there circumstances under which it may not be reasonable to require a hypothesis before a student's investigation?
- Are there examples of major scientific advances which were not preceded by hypotheses?
- Should governments provide funds for research projects which are not aimed at producing a useful result?
- Will the research of private science-based companies always be aimed at financial profit and, if so, what are the implications?

WORKED EXAMPLE E2.3

a Calculate the momentum of a moving particle that has a de Broglie wavelength of 200 pm (1 pm = 1 × 10^{-12} m).

b Determine the wavelength associated with an electron moving with a speed of five million metres per second.

c If we were to suppose that de Broglie's hypothesis extends to macroscopic objects, estimate the wavelength of a moving tennis ball.

Answer

a $p = \frac{h}{\lambda} = \frac{6.63 \times 10^{-34}}{200 \times 10^{-12}} = 3.32 \times 10^{-24}\,\text{kg m s}^{-1}$

b $\lambda = \frac{h}{p} = \frac{6.63 \times 10^{-34}}{(9.110 \times 10^{-31}) \times (5.0 \times 10^{6})} = 1.5 \times 10^{-10}\,\text{m}$

c Estimate momentum, $p = mv = 0.06 \times 20 \approx 1\,\text{kg m s}^{-1}$

$\lambda = \frac{h}{p} = \frac{6.63 \times 10^{-34}}{1} \approx 7 \times 10^{-34}\,\text{m}$

If a ball had this wavelength, it would be too small to measure.

12 A neutron has a de Broglie wavelength of 80 pm.
Calculate the velocity of the neutron.

13 Show that a potential difference of about 4000 V is needed to accelerate an electron from rest so that it has a de Broglie wavelength of 2.0 × 10^{-11} m.

14 Which is associated with a de Broglie wavelength of longer wavelength – a proton or an electron travelling at the same velocity? Explain your answer.

15 Explain why a moving car has no detectable wave properties.

LINKING QUESTION

- What are the defining features and behaviours of waves?

This question links to understandings in Topics C.2 and C.3.

Evidence for the wave nature of matter

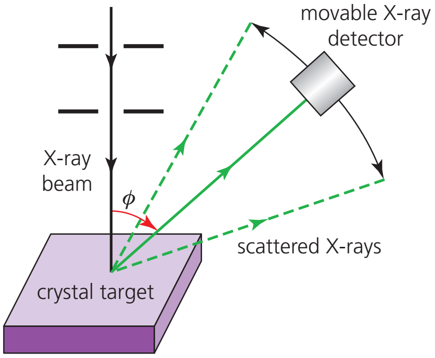
Superposition (interference and diffraction) is behaviour that is characteristic of waves, but not particles.

In order to verify de Broglie's hypothesis, it was necessary to observe and measure the diffraction of a beam of particles (electrons).

◆ **Davisson–Germer experiment** Experiment that verified the wave properties of matter by showing that a beam of electrons is diffracted by a crystal (at an angle dependent upon the velocity of the electrons).

A reminder: we have seen in Topic C.3 that the diffraction of light by a diffraction grating can be represented by the equation $n\lambda = d\sin\theta$. Knowledge of d (the spacing of lines on the grating) and the measurement of the diffraction angle θ (for $n = 1$), can lead to a determination of an unknown wavelength, λ. An important example of this is the determination of spectral wavelengths and frequencies (Topic E.1).

In principle, other electromagnetic waves can be similarly diffracted into patterns. The diffraction / scattering of X-rays is of especial importance because a typical X-ray wavelength is comparable to the separation of atoms / ions (approximately 10^{-10} m). This is the necessary condition for significant diffraction effects.



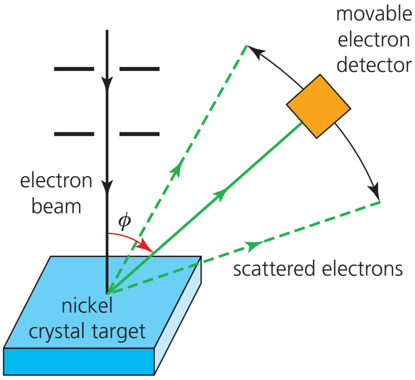
■ **Figure E2.10** Diffraction/ scattering of X-rays by a crystal

The regular three-dimensional arrangement of atoms in a crystal has similarities with the regular two-dimensional arrangement of lines on a diffraction grating. Figure E2.10 shows a much-simplified arrangement. (Knowledge of X-ray diffraction is *not* required for the IB Physics examination.)

When X-rays are diffracted / scattered by parallel layers of atoms / ions in a crystal, knowledge of n, λ and θ can lead to a determination of d, the separation of the layers.

De Broglie's hypothesis proposed that electrons can have wavelengths similar to X-rays and atomic separations ($\approx 10^{-10}$ m) as shown in Worked example E2.3, so it was anticipated that electrons would also be diffracted / scattered by regular arrangements of atoms / ions.

The diffraction of an electron beam was first achieved in the **Davisson–Germer experiment** as seen in Figure E2.11, using a nickel crystal target.

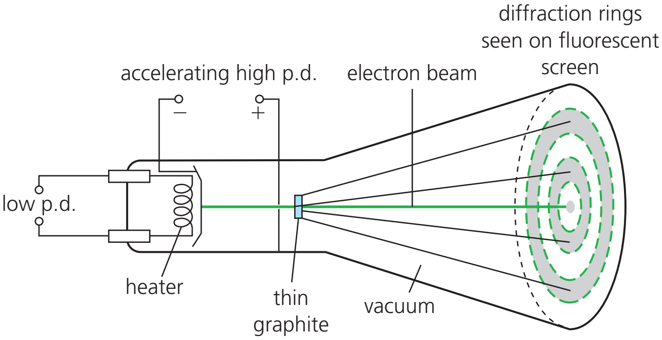


■ **Figure E2.11** Davisson-Germer experiment

A scattered beam of electrons was detected at an angle which confirmed that diffraction was occurring of electron waves with a wavelength consistent with de Broglie's hypothesis.

When the electrons were accelerated to greater speeds (using a larger voltage), they were diffracted through a smaller angle because the wavelength was less.

Figure E2.12 shows the type of modern apparatus used in schools to demonstrate the diffraction of electrons. Two prominent diffraction rings are seen, representing constructive interference of electron waves from two different sets of layers of carbon atoms in the graphite target. (The graphite has many sets of layers with different orientations.)



■ **Figure E2.12** An electron diffraction apparatus

If the voltage accelerating the electrons is increased, they will have greater speed and momentum when they arrive at the graphite. This means that the electron wavelength and diffraction angle will be reduced.

Tool 1: Experimental techniques

Recognize and address safety, ethical or environmental issues in an investigation

What safety precautions are necessary when using the apparatus shown in Figure E2.12? Make a list of safety precautions for other experimenters to follow.

WORKED EXAMPLE E2.4

Using apparatus similar to that seen in Figure E2.11, electrons were accelerated by 5.0 kV. Calculate the:

a kinetic energy of the electrons (J)

b speed of the electrons

c momentum of the electrons

d wavelength of the electrons using de Broglie's hypothesis

e diffraction angle for these electrons, assuming that:

- layers causing the diffraction had a separation of 1.4×10^{-10} m
- diffraction can be modelled by the equation $\lambda = 2d\sin\theta$.

Answer

a $W = qV = (1.60 \times 10^{-19}) \times 5000 = 8.0 \times 10^{-16}\,\text{J}\ (= 5000\,\text{eV})$

b $E_k = \frac{1}{2}mv^2$

$8.0 \times 10^{-16} = \frac{1}{2} \times (9.110 \times 10^{-31}) \times v^2$

$v = 4.2 \times 10^7\,\text{m s}^{-1}$

c $p = mv = (9.110 \times 10^{-31}) \times (4.2 \times 10^7) = 3.8 \times 10^{-23}\,\text{kg m s}^{-1}$

d $\lambda = \dfrac{h}{p} = \dfrac{6.63 \times 10^{-34}}{3.8 \times 10^{-23}} = 1.7 \times 10^{-11}\,\text{m}$

e $\lambda = 2d\sin\theta$

$(1.7 \times 10^{-11}) = 2 \times 1.4 \times 10^{-10} \times \sin\theta$

$\theta = 3.6°$

◆ **X-ray diffraction (crystallography)** Investigating the arrangements of atoms and molecules in matter by detecting how X-rays are diffracted by crystalline materials.

ATL E2B: Social skills

Appreciate the diverse talents of others

Ask an IB chemistry student (or teacher) to give a short presentation to your class about X-ray crystallography (diffraction), as it is explained in the chemistry course. Find out if electron diffraction is used similarly.

16 If the accelerating voltage in Figure E2.12 was doubled, determine by what factor the following would change:

a electron kinetic energy

b momentum of electrons

c wavelength of electrons

d sine of the diffraction angle?

17 In Figures E2.10 and E2.11, the beams can be seen to pass between two slits. Suggest a reason why the slits are needed.

18 Outline the differences and similarities between electrons and X-rays.

Wave–particle duality

SYLLABUS CONTENT

▶ Matter exhibits wave–particle duality.

In principle, all particles have wave properties and all electromagnetic waves have particle (photon) properties. This is widely known as **wave–particle duality**.

◆ **Wave–particle duality** Theory that all particles have wave properties and that all electromagnetic waves have particle properties.

Nature of science: Theories

Two theories for the same thing

We may choose to believe that light, for example, is composed of waves (because it interferes and diffracts), or we may choose to think of light as a stream of particles (photoelectric and Compton effects), but we have difficulty believing the confusing truth that light can behave as both waves and particles, depending on the circumstances.

The following two famous quotations reflect the situation:

'It seems as though we must use sometimes the one theory and sometimes the other, while at times we may use either. We are faced with a new kind of difficulty. We have two contradictory pictures of reality; separately neither of them fully explains the phenomena of light, but together they do.' (Albert Einstein)

'God runs electromagnetics by wave theory on Monday, Wednesday, and Friday, and the Devil runs them by quantum theory on Tuesday, Thursday, and Saturday.' (William Lawrence Bragg)

Maybe we are tempted to believe that there is some, yet unknown, discovery which will solve this paradox, but it is much more likely that we have to accept, once again, that the world of quantum physics does not match the 'reality' we observe in everyday life.

Adding to the confusion, we use the frequency of a wave to determine the energy carried by an individual particle ($E = hf$).

◆ **Compton effect (scattering)** The increase in wavelength (decrease in energy) of high-frequency photons when they interact (collide) with electrons. Important evidence for the particle nature of electromagnetic radiation.

Compton scattering

SYLLABUS CONTENT

- Compton scattering of high-frequency photons by electrons as additional evidence of the particle nature of light.
- Photons scatter off electrons with increased wavelength.
- The shift in photon wavelength after scattering off an electron as given by: $\lambda_f - \lambda_i = \Delta\lambda = \frac{h}{m_e c}(1 - \cos\theta)$.

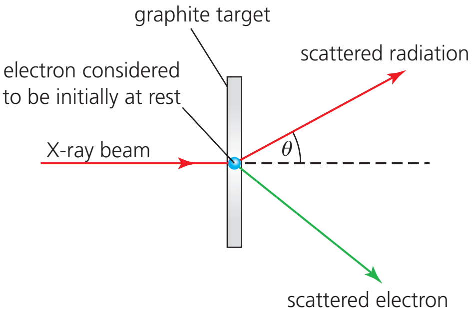


Figure E2.13 Compton scattering

As well as the photoelectric effect, the interaction of electromagnetic radiation with matter was also investigated in the **Compton effect (scattering)**, named after Arthur Compton, the American physicist who won the Nobel prize in 1927 for his work. The effect is most significant with shorter wavelength radiation, such as X-rays and gamma rays. (Compton used wavelengths typically smaller than those used in X-ray diffraction experiments.)

The experiment examined the interaction of monochromatic X-rays with electrons. The electrons were the outer electrons in carbon atoms in a small graphite target placed in the X-ray beam. See Figure E2.13.

The essential feature of Compton scattering is that it cannot be explained by considering the incident X-rays to be waves. A 'particle model' is needed: an individual photon (with momentum, $p = \frac{h}{\lambda}$) 'collides' with an individual electron.

The laws of conservation of energy and momentum can be applied to the scattering. The electron gains kinetic energy, so the photon must lose energy. Since for the X-ray photon, $E = hf$, if it loses energy, it must change to a smaller frequency (equivalent to a change to a greater wavelength: λ_i to λ_f). This was confirmed by experiment, although the change in wavelength was relatively small.

Applying the laws of conservation of momentum and energy leads to the following equation, which accurately predicts the change in wavelength, $\Delta\lambda$, of the photons. (You do *not* need to know how this equation is derived.)

Compton scattering equation:

$$\lambda_f - \lambda_i = \Delta\lambda = \frac{h}{m_e c}(1 - \cos\theta)$$

θ is the scattering angle, as shown in Figure E2.13 and m_e is the mass of the electron, which is assumed to be effectively stationary (at rest) before being scattered.

WORKED EXAMPLE E2.5

Determine the change in wavelength for photons scattered at an angle of 30° in the Compton effect.

$\left(\frac{h}{m_e c}\right.$ is a constant, called the Compton wavelength, and it has a value of 2.426×10^{-12} m.$\left.\right)$

Answer

$$\Delta\lambda = \frac{h}{m_e c}(1 - \cos\theta) = (2.426 \times 10^{-12}) \times (1 - \cos 30°) = 3.25 \times 10^{-13}\,\text{m}$$

This is a small change of wavelength and requires excellent experimental techniques and equipment to measure.

Photons scattered at different angles will have different wavelengths, as is represented by the equation.

LINKING QUESTIONS

- How can particles diffract?
- Why is Compton scattering more convincing evidence for the particle nature of light than that from the photoelectric effect? (NOS)

These questions link to understandings in Topic C.3.

Interestingly, the change of wavelength in Compton scattering depends only on the scattering angle, it does *not* depend on the energy, or wavelength of the incident photon. It is also relatively small: for example, if the incident X-ray photon had a wavelength of 2.25×10^{-10} m, a change of 3.25×10^{-13} m would only be 0.14%.

19 Arthur Compton's scattering experiment is considered to be one of the classic physics investigations of the early twentieth century. Explain why.

20 Calculate the change in photon frequency that occurred in Worked example E2.5.

21 Compton's original experiment used X-rays of wavelength 0.0709 nm.
Calculate the change of wavelength of photons scattered through an angle of 45°.

22 At what scattering angle would you expect to observe the largest change in photon wavelength? Explain your answer.

23 **a** Calculate a typical energy (eV) of:
  i a photon of visible light
  ii a photon of X-rays.
b It requires about 10 eV to remove an electron from a carbon atom. Suggest why this was ignored in the previous discussion of the Compton effect.
c State **two** reasons why Compton scattering of visible light is not observed.

24 An X-ray photon with initial wavelength of 4.700×10^{-11} m was scattered through 34°.
a Determine the change in photon:
  i wavelength **ii** energy.
b State the resulting kinetic energy of the electron with which it collided.
c State any assumption(s) you made in answering part **b**.

25 Outline the main evidence for:
a the wave nature of electromagnetic radiation
b the particle nature of electromagnetic radiation.

E.3

Radioactive decay

Guiding questions

- Why are some isotopes more stable than others?
- In what ways can a nucleus undergo change?
- How do large, unstable nuclei become more stable?
- How can the random nature of radioactive decay allow for predictions to be made?

What is radioactivity?

Isotopes

The nuclei of some atoms are unstable. Spontaneous changes within an unstable nucleus can result in the emission of a particle and/or a high-energy photon. This process is called **radioactivity**. When particles are emitted, the proton number of the atom will change, so that it becomes a different element. This is called **transmutation** or radioactive decay.

A material involved in the process of radioactivity is described as being **radioactive**, while an atom with an unstable nucleus may be referred to as a **radioisotope** or **radionuclide**.

The term isotope was explained in Topic E.1. As a reminder: an *isotope* is one of two or more different nuclides of the same element (which have the same proton numbers, but different nucleon numbers).

Radioactive decay should not be confused with chemical or biological decay. The decay of a radioactive material will not usually involve any obvious change in appearance.

Most of this topic is concerned with explaining radioactivity, but a straightforward example now will help you to begin to understand all these terms, which will become more familiar as your understanding develops.

Atoms of $^{235}_{92}\text{U}$ have unstable nuclei, so we can describe the material as being *radioactive*. The element uranium has several *isotopes* which are all unstable / radioactive. They can all be described as *radioisotopes*. In the last two sentences, we can replace 'isotope' with 'nuclide' if we wish to stress that we are discussing nuclei.

At some (uncertain) time in the future, any $^{235}_{92}\text{U}$ nucleus may emit an alpha particle, $^{4}_{2}\text{He}$, and when this happens, we say that the nucleus has *decayed* or *transmuted*. In this example $^{231}_{90}\text{Th}$ (the element thorium) is formed and it may be called the *decay product* or the **daughter product**.

◆ **Radioactivity** Spontaneous transmutation of an unstable nucleus, accompanied by the emission of ionizing radiation in the form of alpha particles, beta particles or gamma rays.

◆ **Transmutation** When a nuclide changes to form a different element after emitting a particle.

◆ **Radioactive** Describes a substance which contains unstable nuclei which will emit radiation.

◆ **Radioisotope** or **radionuclide** Isotope / nuclide with an unstable nucleus which emits radiation.

◆ **Daughter product** The resulting nuclide after a radionuclide ('parent') emits a particle.

TOK

The natural sciences

- Does the precision of the language used in the natural sciences successfully eliminate all ambiguity?

New terminology

The last section illustrates a recurring theme in science education: so many new words to learn! When acquiring new scientific knowledge, is the introduction of new terms unavoidable? Does it help, or discourage, a student? Is science different from other areas of knowledge in this respect?

Radioactivity experiments

SYLLABUS CONTENT

- Effect of background radiation on count rate.

Figure E3.1 shows a typical experimental arrangement for investigating radioactivity in a school laboratory.

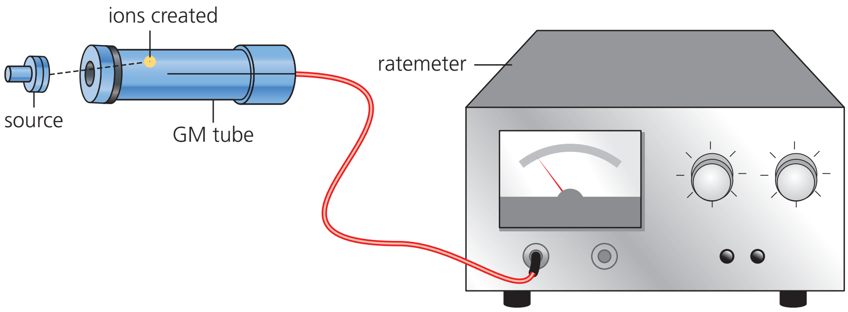


Figure E3.1 Basic components of a radioactivity experiment

◆ **Geiger–Muller tube** Apparatus used with a counter or ratemeter to measure the radiation from a radioactive source.

◆ **Count rate (radioactivity)** The number of nuclear radiation events detected in a given time (per minute or per second).

◆ **Ratemeter** Meter which is connected to a Geiger–Muller tube (or similar) to measure the rate at which radiation is detected.

A tiny amount of a radioactive nuclide is contained in the ‘source’. When nuclear radiation emitted by the source enters the **GM (Geiger–Muller) tube** through the end ‘window’, it causes ionization of the gas inside and a sudden tiny burst of current. These events are ‘counted’ by an electronic ‘counter’, or **ratemeter**, and the results are expressed as a *radioactive count,* or a count per second, or per minute. (The tube and counter together are often described as a Geiger counter.) For example, if over a period of five minutes a total count of 7200 was detected, this would probably be recorded as a **count rate** of 1440 min^{-1} or 24 s^{-1}. Many radiation detectors display a count rate directly, as seen in Figure E3.1.

Top tip!

If a radioactive count is repeated, it will probably *not* give the same result. This is because of the random nature of radioactive decays and not because of uncertainty in the measurement. For example, if a repeated count had an average of 9, it probably varied between 6 and 12, which means that a single measurement could have been unreliable. Larger counts are better. For example, if a repeated count had an average of 900, it probably varied between 870 and 930.

Tool 1: Experimental techniques

Recognize and address safety, ethical or environmental issues in an investigation

Nuclear radiation can be hazardous to humans and animals. Any experiment with radioactive materials must follow safety precautions, which include the following.

- The radioactive sources must be well marked and stored securely in lead-lined boxes. They should be used for as short a time as possible.
- All experiments should be done by, or supervised by, a teacher experienced with the appropriate procedures.
- Sources should be handled with tongs and never pointed directed towards anybody.
- Students watching a demonstration should be a safe distance away. (Nuclear radiation from a point source will be absorbed to some extent in air (depending on the type of radiation) and it will also spread out.)

However, radiation sources used in schools emit very low levels of radiation.

Figure E3.2 Radiation hazard sign

There are a number of things that could be investigated with the apparatus seen in Figure E3.1, including:

- How does the count rate vary when the distance between the GM tube and the source is changed?
- How is the count rate affected by placing various materials between the source and the GM tube?
- Does the count rate change with time?
- Is any count detected if the source is removed?
- Are the radiations affected by passing through electric or magnetic fields?
- How much radiation is emitted by the source every second?

◆ **Background radiation** Radiation from radioactive materials in rocks, soil and building materials, as well as cosmic radiation from space and any radiation escaping from artificial sources.

◆ **Background count** Measure of background radiation.

It is important to understand that there are tiny amounts of radioactive materials in almost everything around us (and in our bodies). These materials emit very low amounts of nuclear radiation which we are all *unavoidably* exposed to everyday. Under most circumstances, this **background radiation** is low enough to be considered completely harmless.

Because of background radiation, a GM tube and ratemeter, such as seen in Figure E3.1 will record a **background count**, even when there is no obvious source of radiation present. A typical count might be 0.25 to 0.5 s^{-1}. If an experiment is measuring low counts, the effect of this background count is significant and it should be deducted from all readings before they are processed.

WORKED EXAMPLE E3.1

In a radioactivity experiment, a count of 42 was recorded from a source in one minute. If the background count rate in that location was 0.44 s^{-1}, what was the value of the count from the source after it had been adjusted for background radiation?

Answer

$42 - (0.44 \times 60) = 16\,\text{min}^{-1}$

1 Give **two** reasons why it is better to use larger count rates in radioactivity experiments.

2 a A ratemeter recorded an average 400 counts per minute from repeated measurements. Use information from the previous ‘Top tip’ box to predict the range of the count rates detected.

  b Discuss which is better:
  - determine a count rate over ten minutes, or
  - calculate an average of ten one-minute measurements.

3 Research on the internet to find possible sources of background radiation.

4 In 15 minutes, a count of 5486 was measured when the GM tube was directed towards a radioactive source. It was known that the background count at that location was 18 per minute. Calculate the average count rate, per second, due to nuclear radiation coming directly from the source.

5 At a location where the background count was 22 min^{-1}, in separate experiments, count rates of 50 min^{-1} and 5000 min^{-1} were measured.
Compare the significance of the background counts in these experiments.

Alpha particles, beta particles and gamma rays

SYLLABUS CONTENT

- The penetration and ionizing ability of alpha particles, beta particles and gamma rays.
- The changes in the state of the nucleus following alpha, beta and gamma radioactive decay.
- The radioactive decay equations involving α, β^{-}, β^{+}, γ.
- The existence of neutrinos v and antineutrinos $\overline{v}$.

◆ **Beta particle** A high-speed electron that is released from a nucleus during beta negative decay, or a high-speed positron released during beta positive decay.

In a school laboratory we can detect three different kinds of radiation emitted from radionuclides:

- alpha particles: fast-moving helium-4 nucleus (2 protons and 2 neutrons tightly bound together), released from a nucleus during alpha decay
- **beta particles**
- gamma rays (usually associated with alpha or beta emission).

Atoms of the same radionuclide always emit the same types of radiation

ATL E3A: Communication skills

Using terminology, symbols and communication conventions consistently and correctly

Nuclear equations

The particle(s) before the reaction are shown on the left and the products shown on the right.

Nuclear equations must balance: the sum of the nucleon numbers and proton numbers must be equal on both sides of the equation.

Alpha particles

The composition of an alpha particle is the same as a helium-4 nucleus: the combination of two protons and two neutrons, which is very stable. It has a nucleon number of 4 and a proton number of +2. Alpha particles can be represented by the symbols $^{4}_{2}\alpha$ or $^{4}_{2}\text{He}$.

Clearly the emission of an alpha particle results in the loss of two protons and two neutrons from a nucleus, so that the proton number of the nuclide decreases by two and a new element is formed (transmutation). This is represented in a generalized **radioactive decay equation** as follows:

$$^{A}_{Z}\text{X} \longrightarrow {}^{A-4}_{Z-2}\text{Y} + {}^{4}_{2}\alpha$$

parent nucleus ⟶ daughter nucleus + alpha particle

◆ **Nuclear equation** An equation representing a nuclear reaction. The sum of nucleon numbers (A) on the left-hand side of the nuclear equation must equal the sum of the nucleon numbers on the right-hand side of the equation. Similarly with proton numbers (Z).

◆ **Radioactive decay equation** Balanced nuclear equation which shows a radionuclide and its decay products.

As an example, the decay of radium-226 results in the emission of an alpha particle:

$$^{226}_{88}\text{Ra} \longrightarrow {}^{222}_{86}\text{Rn} + {}^{4}_{2}\alpha$$

The change to a more stable nucleus is equivalent to a decrease in nuclear potential energy. This energy is transferred to the kinetic energy of the alpha particle (and a lesser amount to the daughter nucleus).

All alpha particles from the decay of radium-226 have exactly the same (kinetic) energy: 4.7 MeV, or 7.5×10^{-13} J. (Some radionuclides emit alpha particles with different, but discrete, energies. This is explained later in this topic for HL students.)

Assuming that there are only two particles after the decay, they must move (recoil) in exactly opposite directions. This is because of the law of conservation of momentum, which also predicts that the alpha particle will have more kinetic energy and a much faster speed, because it is the less massive particle.

One mole of radium (226 g) would release a total energy of: $6.02 \times 10^{23} \times 4.71 = 2.83 \times 10^{24}$ MeV. This is a lot of energy (4.53×10^{11} J) from a relatively small mass, but the energy will be released over a very long time (because the *half-life* of radium-226 is about 1600 years – the concept of half-life is explained later).

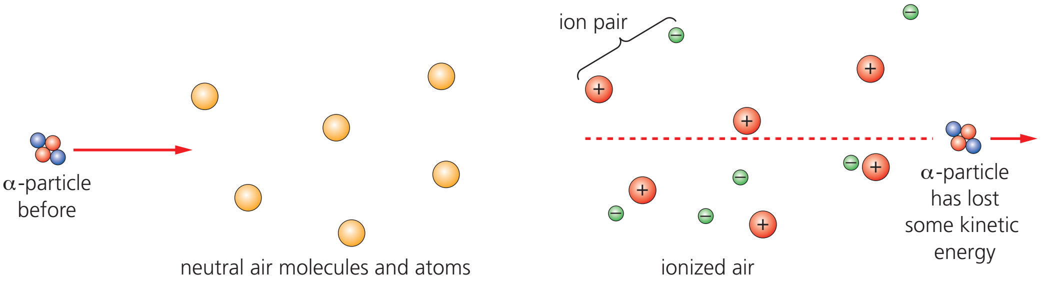
Radionuclides are not generally used to transfer large amounts of energy because they are both low power and expensive, but they can provide energy for a long period of time. Alpha sources can be used to generate small amounts of electrical energy, or in places that are difficult to access, such that replacing a power source would be problematic. This includes some uses on satellites and space probes. But, as we will see later, another kind of nuclear reaction is capable of transferring large amounts of energy quickly – nuclear fission.

◆ **Penetrating power** The penetrating power of nuclear radiation depends upon the ionizing power of the radiation. The radiation continues to penetrate matter until it has lost (nearly) all of its energy. The greater the ionization per cm, the less penetrating power it will possess.

◆ **Ionizing ability** A measure of how much ionization is caused when a particular type of radiation passes through a material.

Penetrating power and ionizing ability

We have already noted in discussing the Geiger–Muller–Rutherford experiment (Topic E.2), that alpha particles have considerable kinetic energy (for a subatomic particle) however they have limited penetration of matter (**penetrating power**). This is because they transfer significant amounts of energy in collisions with other atoms / molecules. (Kinetic energy transfer is greatest when the colliding particles have comparable masses, as discussed in Topic A.3.) See Figure E3.3.



■ **Figure E3.3** Formation of ion pairs by alpha particles from molecules in the air

The collisions transfer the energy needed to ionize a large number of atoms / molecules in the material through which the alpha particles are passing. After most of their kinetic energy has been transferred, the alpha particles are effectively absorbed (as tiny amounts of helium). See Question 14 for a numerical example.

Typically, all the alpha particles emitted from a source will be absorbed by a few centimetres of air, or a sheet of paper (although they will mostly pass through much thinner gold foil). Alpha particles would be absorbed in the outer layers of skin, so that radioactive sources that only emit alpha particles are not considered to be dangerous outside of the human body. However, sources of alpha radiation that have been taken into the body (by eating, drinking or breathing) are a significant health hazard.

◆ **Antiparticle** Every particle has an antiparticle with opposite physical properties.

◆ **Antimatter** Matter consisting of antiparticles.

◆ **Antineutrino** Low-mass, uncharged and very weakly interacting particle emitted during beta-negative decay. Antiparticle of neutrino.

◆ **Annihilation** When a particle and its antiparticle interact, their mass is totally converted to electromagnetic energy.

◆ **Neutrino** Low-mass, and very weakly interacting uncharged particle emitted during beta-positive decay. Antiparticle of antineutrino.

■ Beta-negative particles

In an unstable nucleus it is possible for an uncharged neutron to be converted into a positive proton and a negative electron. This also involves the creation of another particle called an **antineutrino**, $\overline{v}$. An antineutrino is an example of an **antiparticle** (**antimatter**).

$$^{1}_{0}n \longrightarrow {}^{1}_{1}p + {}^{0}_{-1}e + \overline{v}$$

ATL E3B: Communication skills

Clearly communicating complex ideas in response to open-ended questions

Find out what physicists mean by ‘antimatter’.

When particles of matter and antimatter collide, they destroy (**annihilate**) each other, with an enormous release of energy. Since we live in a universe which is made of matter, if any antimatter is created, it very quickly annihilates.

Antineutrinos (and **neutrinos**) are very small particles with no charge, which travel at speeds close to the speed of light, so they are *very* penetrating and *very* difficult to detect. (They cannot be detected in a school experiment.)

After the nuclear reaction shown in the equation above occurs, it is not possible for the newly formed electron to remain within the nucleus and it is ejected from the atom at a very high speed (close to the speed of light). It is then called a beta-negative particle and it is represented by the symbol ${}^{0}_{-1}\beta^-$ or ${}^{0}_{-1}e$.

When **beta-negative decay** occurs, the number of nucleons in the nucleus remains the same, but the number of protons increases by one, so that a new element is formed. This can be represented in a radioactive decay equation of the general form:

$$^{A}_{Z}\mathrm{X} \longrightarrow {}^{A}_{Z+1}\mathrm{Y} + {}^{0}_{-1}\beta^- + \overline{\nu}$$

Top tip!

You will *not* be expected to remember the names of elements from their proton numbers.

◆ **Beta-negative decay** Radioactive decay resulting in the emission of an electron (and an antineutrino)

◆ **Beta-positive decay** Radioactive decay resulting in the emission of a positron (and a neutrino).

◆ **Positron** Antiparticle of the electron; released during beta-positive decay.

◆ **Gamma radiation / ray** Electromagnetic radiation (photons) emitted from some radionuclides and having an extremely short wavelength.

A typical example is the decay of a strontium-90 nuclide:

$$^{90}_{38}\mathrm{Sr} \longrightarrow {}^{90}_{39}\mathrm{Y} + {}^{0}_{-1}\beta^- + \overline{\nu}$$

The beta particles in this decay have a range of energies up to 0.55 MeV.

Beta particles (unlike alpha particles) are emitted with a continuous range of different energies, but there is a well-defined maximum energy from any particular source, typically about 1 MeV.

Penetrating power and ionizing ability

Beta particles are considerably less massive than alpha particles, which means that they transfer less energy in ionizing collisions with atoms and molecules. Therefore, they travel further before they lose their kinetic energy and become absorbed.

Beta particles travel typically about 30 cm in air, although more energetic particles may go as far as one metre. They will mostly pass through a sheet of paper easily and their absorption in other materials is usually characterized by saying that a sheet of aluminium of thickness 3 mm will just absorb them all.

Sources of beta radiation can be dangerous if they enter the body, but they should also be considered as a possible health hazard if they are outside the body.

Beta-positive particles

In a similar process to beta-negative decay, called **beta-positive decay**, a proton in a nucleus can be converted into neutron and a positively charged electron, called a **positron** (another example of antimatter), which is then ejected from the atom, after which it is called a beta-positive particle and it is represented by the symbol ${}^{0}_{1}\beta^+$ or ${}^{0}_{1}e$.

A neutrino, ν, is created at the same time.

$$^{1}_{1}p \longrightarrow {}^{1}_{0}n + {}^{0}_{1}\beta^+ + \nu$$

The following equation represents a typical beta-positive decay:

$$^{23}_{12}\mathrm{Mg} \longrightarrow {}^{23}_{11}\mathrm{Na} + {}^{0}_{1}\beta^+ + \nu$$

Gamma rays

Gamma rays are high-frequency, high-energy electromagnetic radiation (photons) released from unstable nuclei. A typical wavelength is about 10^{-12} m. This corresponds to an energy of about 1 MeV (use $E = hc/\lambda$). Gamma rays are usually emitted after an unstable nucleus has emitted an alpha or beta particle. Gamma rays are represented by the symbol ${}^{0}_{0}\gamma$.

For example, when a thorium-234 nucleus is formed from a uranium-238 nucleus by alpha decay, the thorium nucleus contains excess energy and is said to be in an *excited state*. The excited thorium nucleus (shown by the symbol * in an equation) returns to a more stable state by emitting a gamma ray:

$$^{234}_{90}\mathrm{Th}^* \longrightarrow {}^{234}_{90}\mathrm{Th} + {}^{0}_{0}\gamma$$

Because gamma rays have no mass or charge, the composition of the emitting nucleus does not change. There is no transmutation.

Penetrating power and ionizing ability

Gamma rays cause less ionization, so that they have much greater penetrating power than alpha particles or beta particles.

We usually assume that gamma rays are not significantly absorbed in air, but if a beam is spreading out, its intensity falls with distance, following an inverse square law (assuming that they come from a point source). At least a two centimetres thickness of solid lead is needed to absorb most gamma rays. Because they are so penetrating, gamma rays are less easy to detect than alpha particles and beta particles rays.

However, because all of their energy can be transferred in one interaction, gamma rays can cause significant chemical and biological changes when absorbed in the human body. Because they are so penetrating, sources outside the body can be as dangerous as sources inside the body.

LINKING QUESTION

- Are there differences between the photons emitted as a result of atomic versus nuclear transitions?

Summary of the properties of alpha, beta and gamma nuclear radiations

Table E3.1 Summary of properties of alpha, beta and gamma radiations

Property	**Alpha (α)**	**Beta negative (β^-)**	**Beta positive (β^+)**	**Gamma (γ)**
relative charge	+2	−1	+1	0
relative mass	4	1/1840	1/1840	0
typical range in air	4 cm	30 cm	very quickly annihilates	very little absorption in air
composition	helium nucleus	electron	positron	electromagnetic wave / photon
typical speed	$\approx 10^7\,\text{m s}^{-1} = 0.1c$	$\approx 2.5 \times 10^8\,\text{m s}^{-1} \approx 0.9c$	$\approx 2.5 \times 10^8\,\text{m s}^{-1} \approx 0.9c$	$3.00 \times 10^8\,\text{m s}^{-1} = c$
notation	^4_2He or $^4_2\alpha$	$^{0}_{-1}e$ or $^{0}_{-1}\beta^-$	$^{0}_{+1}e$ or $^{0}_{+1}\beta^+$	γ or $^0_0\gamma$
ionizing ability	very high	low	very quickly annihilates	very low
absorbed by	thick piece of paper	3 mm aluminium	very quickly annihilates	intensity halved by about 2 cm lead

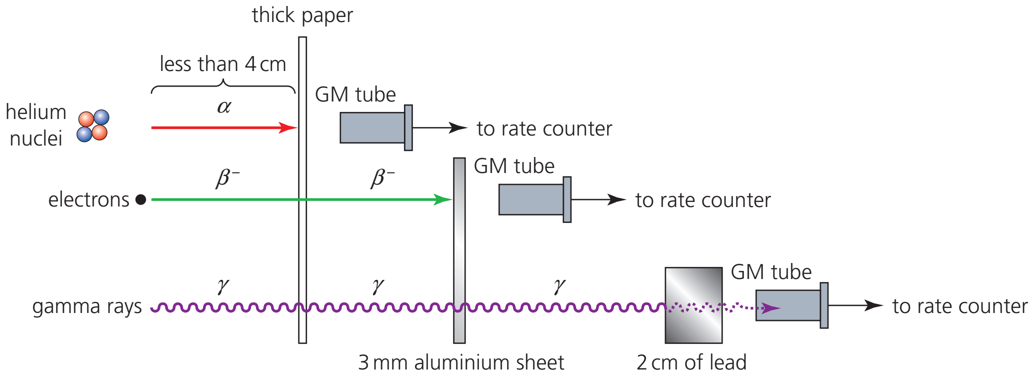


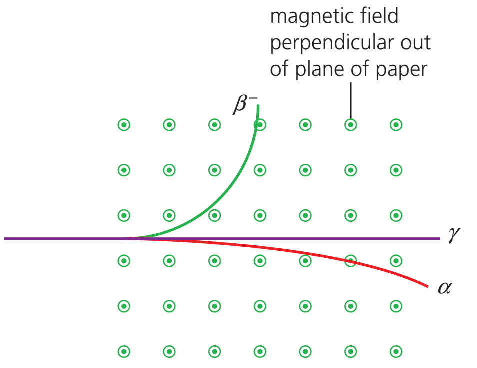
Figure E3.4 Absorption of ionizing radiations

Deflection of nuclear radiations in electric and magnetic fields

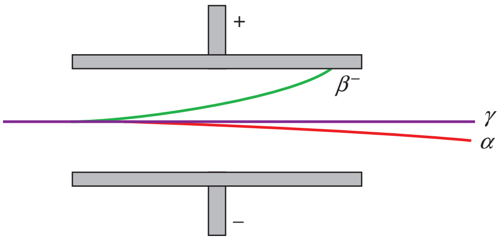
Alpha and beta radiation will be emitted in random directions from their sources, but they can be formed into narrow beams (collimated) by passing the radiation through slits.

Because a beam of alpha particles, or beta particles, is a flow of charge they will be deflected if they pass across a an electric field or magnetic field (as discussed in Theme D).

Gamma radiation is uncharged, so it cannot be deflected in this way.



■ **Figure E3.5** Behaviour of ionizing radiations in a magnetic field



■ **Figure E3.6** Behaviour of ionizing radiations in an electric field

Figure E3.5 shows the passage of the three types of ionizing radiation perpendicularly across a strong magnetic field. Fleming's left-hand rule can be applied to confirm the deflection of the alpha and beta particles into circular paths, the magnetic force providing the centripetal force. The radius of the path of a charged particle moving perpendicularly across a magnetic field can be calculated from:

$$r = \frac{mv}{qB} \text{ (Topic D.3).}$$

An alpha particle has twice the magnitude of charge and about 8000 times the mass of a beta particle, although a typical beta particle may be moving ten times faster. Taking all three factors into consideration, we can predict that the radius of an alpha particle's path may be about 400 times the radius of a beta particle in the same magnetic field: it is deflected much less. (Note that observation of the deflection of alpha particles will require a vacuum.)

Alpha and beta radiation can also be deflected by electric fields, as shown in Figure E3.6. Alpha particles are attracted to the negative plate; beta particles are attracted to the positive plate. The combination of constant speed in one direction, with a constant perpendicular force and acceleration, produces a parabolic trajectory. This is similar to the projectile movement discussed in Topic A.1. The deflection of the alpha particles is small in comparison to beta particles, due to the same factors as discussed for magnetic deflection.

6 **a** Assuming they have energies of 1.0 MeV, calculate the speeds of alpha particles (mass = 6.64×10^{-27} kg).

b What potential difference would be needed to accelerate doubly charged helium ions to the same energy from rest?

7 Explain why the distance before the thick paper in Figure E3.4 is labelled as 'less than 4 cm'.

8 Alpha particles usually carry more energy than beta particles, or gamma rays but, paradoxically, they are less penetrating. Explain why.

9 Explain why a source of alpha radiation outside the human body may be considered to be very low risk (for example, they are used in smoke detectors), but a source inside the body is considered dangerous.

10 Explain how a beam of beta particles can be distinguished experimentally from alpha particles and gamma rays.

11 Explain why gamma rays are considered to be particularly dangerous.

12 Calculate the amount of energy carried by a gamma ray photon of wavelength 2.6×10^{-12} m in **i** J and **ii** eV.

13 An adjusted count rate of 45 min^{-1} was detected from a gamma ray source when the GM tube was 20 cm from the source.

Predict what average count rate would be detected at a distance of:

a 40 cm **b** 100 cm **c** 10 m.

14 Alpha particles lose about 2.2×10^{-18} J of kinetic energy in each collision with an atom or molecule in the air.

An alpha particle travelling through air makes 7×10^4 ionizing collisions with molecules or atoms in the air for each centimetre of travel.

Calculate the approximate range of an alpha particle if the particle begins with an energy of 7.0×10^{-13} J.

15 Represent in a drawing a magnetic field acting perpendicularly into the paper.

Then draw a straight line down the page to represent the original direction of a beam containing alpha particles, beta particles and gamma rays passing through the field.

Finally, show in your drawing, what happens to the three different types of radiation as they pass through the magnetic field.

16 Discuss why beta particles are usually affected more than alpha particles as they pass through electric and magnetic fields.

17 Some beta-negative particles in a beam have an speed of $2.2 \times 10^8\,\text{ms}^{-1}$. The beam passes perpendicularly across a magnetic field of strength 6.5 mT in a vacuum. Using an equation from Topic D.3, determine the radius of the arc of their circular path.

18 From Topic A.3, we know that kinetic energy, $E_k = \frac{p^2}{2m}$.

a Use the law of conservation of momentum to show that, after a stationary nucleus X emits an alpha particle, the kinetic energy of the alpha particle = $\frac{m_X}{m_\alpha}$ × kinetic energy of X.

b Earlier in this section it was stated that the alpha particle emitted in the decay of radium-226 had a (kinetic) energy of 4.7 MeV.
Show that the total energy released in the decay is about 4.8 MeV.

Chart of the nuclides and decay series

◆ **Chart of nuclides** A chart which displays every possible nuclide on axes of proton number and neutron number.

Every nuclide can be placed on a **chart of nuclides**, which has a square for every possible combination of proton number and neutron number.

This is a large chart, as can be seen in Figure E3.7, but the start of it can be seen in Figure E3.21 (it is discussed in greater detail later in this topic for HL students).

Figure E3.7 A full chart of the nuclides contains a lot of data and requires a large wall!

When a nucleus emits an alpha particle or beta particle, we know that it transmutes to a different nuclide. These changes can be tracked on a chart of the nuclides, as shown in Figure E3.8.

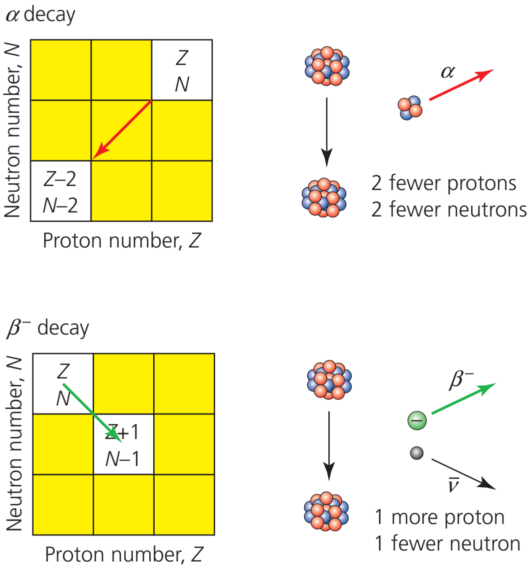
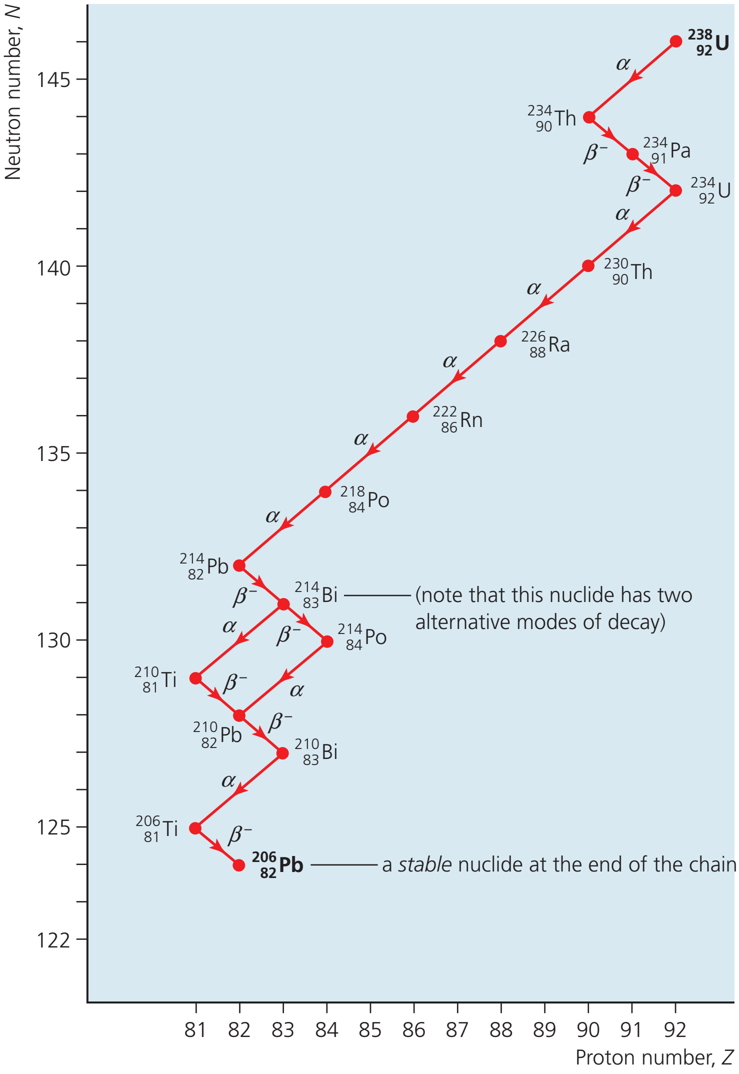


Figure E3.8 Transmutation on a chart of the nuclides

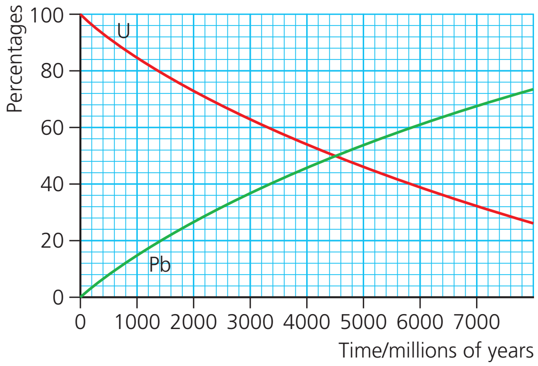
◆ **Decay series** A series of nuclides linked in a chain by radioactive decay. Each nuclide in the chain decays to the next until a stable nuclide is reached.

Heavy radioactive nuclides, such as radium-226 and uranium-238, cannot become stable by emitting just one particle. They undergo a radioactive **decay series**, producing either an alpha or a beta particle and maybe gamma radiation during each step, until a stable nuclide is formed. For example (not to be remembered!), uranium-238 undergoes the decay series seen in Figure E3.9 to eventually form the stable nuclide lead-206. Each decay will have its own particular half-life. (For the sake of clarity, the individual squares have not been included in the figure.)



■ **Figure E3.9** An example of a decay series (uranium-238) on a chart of nuclides

If it was possible to have a source of pure uranium-238, for example, it would immediately start decaying into other nuclides and, after some time, all the nuclides in the decay series would be present in the sample. The relative proportions of different nuclides depend on their half-lives. After a *very* long time most of the source will have turned into lead. Figure E3.10 is a rough indication of how the proportions of U-238 and Pb-206 change over billions of years.



■ **Figure E3.10** Uranium transmuting into lead

We should expect that most radioactive sources (of the heavier elements particularly) to contain a range of different nuclides.

19 Make a sketch, similar to those seen in Figure E3.8, to represent the transmutation that occurs as a result of beta-plus emission.

20 $^{222}_{86}$Rn, decays to Po-218. This radionuclide then emits an alpha particle to create an isotope of lead. Next in the decay series is Bi-214. Write out the full decay equations for these three nuclear reactions.

21 Use the internet to find out how the nuclide carbon-14 decays and then represent the process in a similar way to that shown in Figure E3.8.

Patterns of radioactive decay

SYLLABUS CONTENT

- Random and spontaneous nature of radioactive decay.
- Activity, count rate and half-life in radioactive decay.
- Changes in activity and count rate during radioactive decay using integral values of half-life.

Radioactivity comes from unstable nuclei, but when any particular nucleus will decay and emit a particle or radiation, is completely unpredictable and uncontrollable. At some point in time an unstable nucleus will decay, but there is no way that the process can be controlled by scientists. Temperature, for example, cannot be used to control nuclear reactions (unlike chemical reactions).

Imagine that we could observe the decay of a number of unstable nuclei (another 'thought experiment'):

◆ **Random** Without pattern or predictability.

◆ **Spontaneous** (decay, for example) Without any cause, cannot be controlled.

Individual nuclei do not decay in any pattern (the decays are **random**) and each decay occurs without any obvious cause, (the decays are **spontaneous**).

Paradoxically, such randomness and unpredictability on the scale of individual nuclei, results in predictability when we consider very large numbers of nuclei.

Nature of science: Patterns and trends

Randomness

This is not the first time that the random behaviour of particles has been discussed in this course. Our understanding of the physical properties of gases developed from an appreciation of the random motions of gas molecules. Although the individual motions of gas particles are random and unpredictable, over large numbers of particles (in bulk) we can observe patterns and trends in the properties of the gas.

In everyday life, the toss of a single coin or the throw of a single dice (die) are used to make an event random and unpredictable. However, if we toss a coin enough times, we can be sure that, to a close approximation, 50% will be 'heads' and 50% will be 'tails'. Similarly, if a six-sided dice is thrown, for example 100 times, then any particular number can be expected to occur about once in every six throws (about 17 times in 100 throws). The greater the number of events, the more precisely the outcome can be predicted.

The same principle can be applied to random nuclear decays: we might say, for example, that 50% of the nuclei of a particular nuclide in a source will decay during the next year.

Activity of a radioactive source

◆ **Activity, *A* (of a radioactive source)** The number of nuclei which decay in a given time (second).

◆ **Becquerel, Bq** The SI unit for (radio) activity. 1 Bq = one nuclear decay every second.

The **activity**, *A*, of a radioactive source is the total number of nuclei decaying every second.

Activity may also be described as the rate of decay. We usually assume that the activity of a source is equal to the number of particles emitted every second. The SI unit of (radio) activity is the **becquerel**, Bq.

1 Bq is equivalent to one decay every second and is considered to be a very low activity.

The activity of a source is proportional to the number of undecayed atoms it contains.

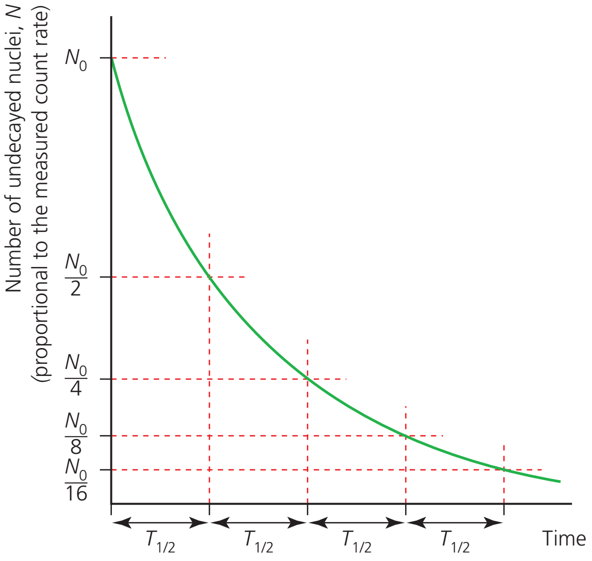


Figure E3.11 Radioactive decay curve

It should be noted that a *count rate* from a source (as being measured in Figure E3.1, for example) is *not* the same as its activity. This is because the GM tube is certainly not detecting all the radiation emitted. However, it is often assumed that a count rate is proportional to the activity.

The activity of all radioactive sources decreases with time. This is because the number of nuclei decaying every second (the activity) depends on the number of nuclei in the source which have not yet decayed. As more nuclei decay, the number remaining undecayed decreases, so the activity decreases. This reducing activity and count rate (adjusted for background count) are represented in Figure E3.11.

As explained above, half of the nuclei of any particular nuclide in a source will decay during a well-defined period of time. This is called the **half-life**, $T_{\frac{1}{2}}$, of the nuclide.

The half-life, $T_{\frac{1}{2}}$, of a radionuclide is the time it takes for half of its undecayed nuclei to decay. It is also the time taken for the activity (or count rate) to halve.

Half-lives of different radionuclides can be as short as fractions of a second, or as long as millions of years, or anything in between. See Table E3.2 for some diverse examples.

Table E3.2 Half-life examples

Radionuclide	Half-life
uranium-238	4.5×10^9 years
radium-226	1.6×10^3 years
radon-222	3.8 days
francium-221	4.8 minutes
astatine-217	0.03 seconds

The graph seen in Figure E3.11 represents an **exponential decrease**: in equal intervals of time (shown clearly on the time axis) the count rate falls by the same fraction (one half): starting at N_0, then $N_0/2$, then $N_0/4$ and so on. In theory, for an exponential decrease, the count rate will never reduce to zero.

Tool 3: Mathematics

Carry out calculations involving logarithmic and exponential functions

Any exponential decrease can be recognized by the fact that a quantity decreases to the same fraction in equal intervals of time. We usually refer to a quantity falling to *half* of its value at the end of each equal time interval, but the same behaviour also falls by *any* other chosen fraction in different, but equal, time intervals.

LINKING QUESTION

- Which areas of physics involve exponential change? (NOS)

Common mistake

Many people wrongly believe that the term ‘exponential’ is only used to describe rapid increases. However, exponential changes can also be decreases and they are just as likely to be slow: consider, for example, that uranium-238 has a half-life of about 4.5 billion years.

Figure E3.12 shows a visualization that may help understanding. The same information is displayed in Table E3.3. The radionuclide americium-242 has a half-life of 16 hours.

◆ **Half-life (radioactive)** The time taken for the activity, or count rate, from a pure source to be reduced to half. Also, equals the time taken for the number of radioactive atoms in a pure source to be reduced to half.

◆ **Exponential change** A change which occurs when the rate of change of a quantity at any time is proportional to the actual quantity at that moment. Can be an *increase* or a *decrease*.

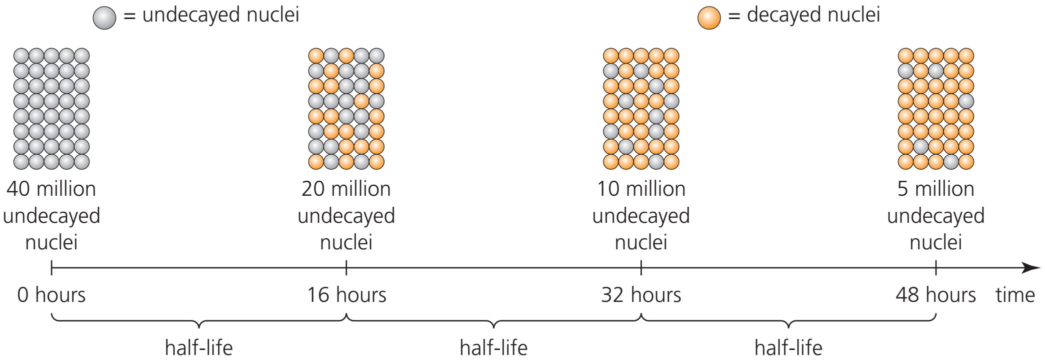


Figure E3.12 The radioactive decay of a sample of americium-242

Table E3.3

Number of undecayed nuclei	Fraction of original undecayed nuclei remaining	Number of decayed nuclei	Number of half-lives elapsed	Number of hours elapsed
40×10^6	1	0	0	0
20×10^6	$\frac{1}{2}$	20×10^6	1	16
10×10^6	$\frac{1}{4}$	30×10^6	2	32
5.0×10^6	$\frac{1}{8}$	35×10^6	3	48
2.5×10^6	$\frac{1}{16}$	37.5×10^6	4	64

WORKED EXAMPLE E3.2

Radium-226 has a half-life of 1620 years. A source which has a total mass of 0.010 g contains 30% of Ra-226 and no other radionuclides.

a Calculate the mass of Ra-226 that will remain in the source after 3240 years.

b Determine how many Ra-226 nuclei will have decayed in this time.

Answer

a After two half-lives, $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ of the unstable nuclei will remain ($\frac{3}{4}$ has decayed).

Mass of Ra-226 remaining $= \frac{1}{4} \times 0.30 \times 0.010 = 7.5 \times 10^{-4}$ g

b Mass of Ra-226 decayed $= \frac{3}{4} \times 0.30 \times 0.010 = 22.5 \times 10^{-4}$ g

226 g of radium-226 contain 6.02×10^{23} atoms (Avogadro constant)

$$\frac{22.5 \times 10^{-4}}{226} \times 6.02 \times 10^{23} = 5.99 \times 10^{18} \text{ nuclei.}$$

Experimental determination of half-life

In principle, the half-life of any radionuclide can be determined from a graph of count rate against time (as in Figure E3.11).

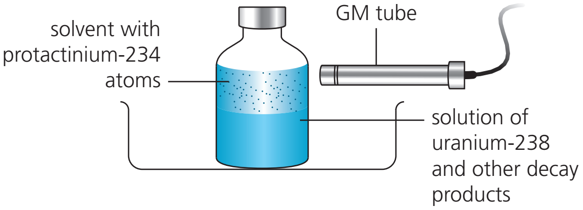


Figure E3.13 Measuring the half-life of protactinium-234

However, this can be difficult to obtain, for two reasons:

- Half-lives will usually be too short or too long for convenient measurement. For example, for a school experiment a half-life between a few minutes and few hours may be considered ideal, but there are not many obtainable radioisotopes that fit that description.
- When a nuclide decays it is probable that its daughter product will also be radioactive. This means that there will often be two (or more) radioisotopes with different half-lives in the same source.

The decay of protactinium-234 is widely used in schools as a demonstration of a half-life determination. (Details need not be remembered.) A compound of uranium-238 dissolved in water is contained in a very securely sealed plastic bottle. A separate layer contains a chemical which reacts with protactinium. U-238 decays to thorium-234 by emitting an alpha particle. The thorium then decays to protactinium-234 by beta-negative emission. Protactinium then decays to uranium-234 when beta-negative particles are emitted. These decays are shown below. This decay series should be understood, but not remembered.

$$^{238}_{92}\mathrm{U} \xrightarrow{\alpha} {}^{234}_{90}\mathrm{Th} \xrightarrow{\beta^-} {}^{234}_{91}\mathrm{Pa} \xrightarrow{\beta^-} {}^{234}_{92}\mathrm{Pa}$$

Of all the radionuclides present in the bottle, only Pa-234 has a suitable half-life for measurement. The protactinium compound can be separated chemically when the contents of the bottle are shaken up. The protactinium moves into the upper layer. See Figure E3.13.

Inquiry 2: Collecting and Processing data

Processing data

Table E3.4 shows the variation with time, *t*, of the count rate of a sample of a radioactive nuclide X. The average background count during the experiment was 36 min^{-1}.

Table E3.4 Variation with time of the count rate of a sample of radioactive nuclide *X*

***t*/hour**	0	1	2	3	4	5	6	7	8	9	10
Count rate/min^{-1}	854	752	688	576	544	486	448	396	362	334	284

Plot a graph to show the variation with time of the corrected count rate and use the graph to determine the half-life of nuclide *X*.

22 One hundred dice were thrown at the same time and all the dice that showed 6 were then removed. The remaining dice were thrown again and, again, all the 6s were removed. The process was repeated another five times.

a Draw a bar chart to represent the results you would expect.

b Explain why the shape of your chart should be similar to Figure E3.11.

23 Count rates detected every five minutes (s^{-1}) were as follows: 75, 60, 48, 38, 31, 25. Assuming that these readings were adjusted for background count, do they represent an exponential decrease? Justify your answer.

24 The initial count rate from a sample of a radioactive nuclide is 560 s^{-1} (adjusted for background count). The half-life of the nuclide is 5 minutes. Sketch a graph to show how the activity of the sample changes over a time interval of 25 minutes.

25 Explain why it would be difficult for a laboratory to provide a school with a pure radioisotope with a half-life of about of about one hour.

26 A radioactive isotope has a half-life of eight days and the initial count rate is 996 min^{-1}. If the average background count was 20 min^{-1}, predict the count rate after 32 days.

27 a The half-life of francium-221 is 4.8 minutes. Calculate the fraction of a sample of francium-221 remaining undecayed after a time of 24.0 minutes.

b The half-life radon-222 is 3.8 days. Calculate the fraction of a sample of radon-222 that has decayed after 22.8 days.

c Cobalt-60 is used in many applications in which gamma radiation is required. The half-life of cobalt-60 is 5.26 years. A cobalt-60 source has an initial activity of 2.00×10^{15} Bq. Calculate its activity after 26.30 years.

d A radioactive element has a half-life of 80 minutes. Determine how long will it take for the count rate to decrease to 250 per minute if the initial count rate is 1000 per minute.

e The half-life of radium-226 is 1620 years. For an initial sample:

i calculate what fraction has decayed after 4860 years

ii what fraction remains undecayed after 6480 years?

28 Technetium-99 is a radioactive waste product from nuclear power stations. It has a half-life of about 212 000 years.

a *Estimate* the percentage of technetium that is still radioactive after one million years.

b *Approximately* how many years are needed for the activity from the technetium to fall to 1% of its original value?

Practical uses of radionuclides

Radioactive substances have a wide range of uses including:

- diagnosis of illness
- treatment of cancer
- food preservation and sterilization of medical equipment (see Question 35)
- determining the age of rocks (see Question 34)
- locating faults in metal structures, such as pipes (see Question 32)
- **carbon dating** (See Question 30)
- determining thicknesses (see Question 33)
- smoke detectors (see Question 31).

◆ **Carbon dating** Using the radioactive decay of carbon-14 to estimate the age of once-living material.

◆ **Tracer (radioactive)** Radioisotope introduced into a system (for example, a human body) to track where it goes by detecting the radiation that it emits

The choice of a suitable radionuclide for each of these applications requires careful consideration of the health risks involved, a suitable half-life and the penetrating power of the emitted radiation.

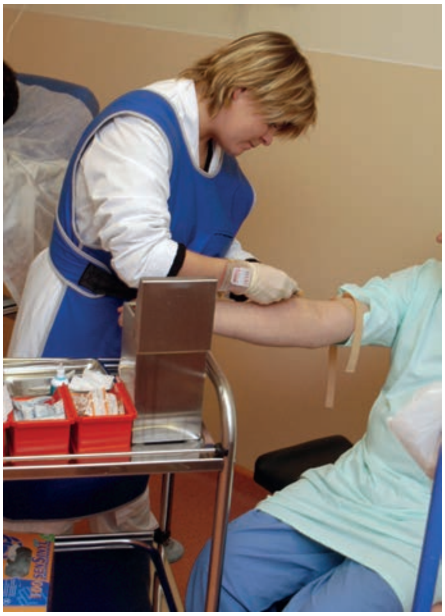


Figure E3.14 Injecting a radioactive tracer

We will look at one application in detail: medical tracers.

Medical tracers

Substances introduced into the body for the purpose of checking the functioning of particular organs are called medical **tracers**. They may be injected (see Figure E3.14) or swallowed. The radioactive substance most commonly used is technetium-99m. This is an excited nuclide produced from molybdenium-99 by beta decay. The decay of the Tc-99m has a half-life of six hours and the gamma ray photons emitted have an energy of 0.14 MeV. Both of these properties make technetium-99m a good choice for tracer studies. The amount of energy carried by the gamma photons makes them easy to detect outside of the body using a gamma ‘camera’. The doctor will be able to use the pattern of gamma rays detected to compare the patient’s organ to the pattern received from a fully healthy organ.

The half-life of 6.0 hours is a good compromise between activity lasting long enough to be useful and the need to expose the patient to ionizing radiation for as short a time period as possible. Atoms in a chemical compound which is preferentially taken up by the organ are replaced by radioactive substitutes. This process is often called ‘labelling’.

29 **a** Determine the percentage of technetium-99m that will remain in a patient 24 hours after they have been injected with a tracer.

b Discuss why gamma ray sources are needed for medical tests similar to that seen in Figure E3.14.

30 Neutrons are continually created in the Earth's atmosphere by cosmic rays. The following nuclear reaction can then occur:

$$^{14}_{7}\mathrm{N} + ^{1}_{0}n \rightarrow ^{14}_{6}\mathrm{C} + ^{1}_{1}\mathrm{H}$$

a Describe what this equation represents.

b Carbon-14 is radioactive and decays by beta-negative emission.
Write an equation for this decay.

c All living plants and animals contain many carbon atoms. A very small fraction of those atoms are C-14. This fraction remains constant while the plant or animal is alive (12 atoms of C-14 in every 1×10^{13} atoms of C-12).
Explain why this percentage will decrease after the plant or animal dies.

d If 12% of the atoms in a human body are carbon, and a body has about 7×10^{27} atoms, estimate how many radioactive carbon-14 atoms are in the body.

e C-14 has a half-life of approximately 5700 years. Predict how many years after death will the fraction of C-14 have fallen to an average of 0.15 atoms of C-14 in every 10^{13} atoms of C-12.

f **i** Explain how scientists can 'date' the age of once-living material using the presence of C-14.

ii Suggest one reason why the results of this process may have a large uncertainty.

31 Many smoke detectors (see Figure E3.15) contain a tiny amount (about one quarter of a microgram) of the radionuclide Am-241. An alarm automatically sounds if smoke comes between the Am-241 and a small radiation detector inside the unit.

■ **Figure E3.15** Smoke detector

a Suggest what kind of radiation is being used, and why.

b Discuss what would be an ideal half-life for the source.

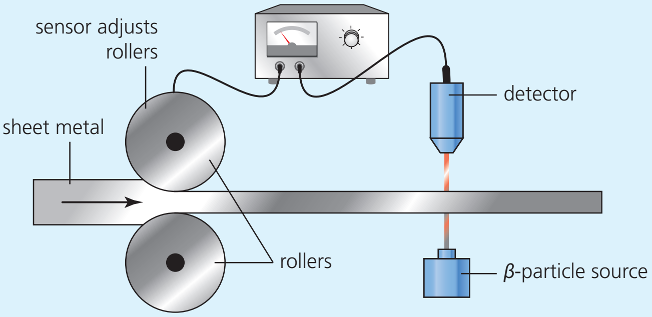
c Use the internet to determine the half-life of Am-241.

32 A leak from a pipe can be traced by using a radionuclide. Suggest how this can be done, including your choice for the type of radiation to be used and its half-life.

33 Figure E3.16 shows how the relative absorption of nuclear radiation in a metal sheet can be used to control its thickness during manufacture.

a Explain why beta particles are being used.

b Discuss whether the same source would be suitable for monitoring the production of rolls of paper or thin plastic.



■ **Figure E3.16** Using a radionuclide to control the manufacture of sheet metal.

34 Uranium-238 is a naturally occurring radionuclide with a half-life comparable to the age of the Earth.
It is widespread in the rocks of the Earth, but in relatively small quantities.

a Discuss whether U-238 in the rocks around us is a significant health hazard.

b Determine the approximate percentage of U-238 that remains since the creation of planet Earth.

c U-238 is the start of a long decay series. What stable nuclide is at the end of that series?

d Explain how the decay of U-238 can be used to obtain a value for the age of the Earth.

35 Explain why gamma rays can be used in hospitals to treat cancerous growths.

The energy inside a nucleus

SYLLABUS CONTENT

- Existence of the strong nuclear force, a short-range attractive force between nucleons.
- Nuclear binding energy and mass defect.
- Variation of the binding energy per nucleon with nucleon number.
- The mass–energy equivalence as given by $E = mc^2$ in nuclear reactions.

◆ **Binding energy** The energy released when a nucleus is formed from its constituent nucleons. Alternatively, it is equal to the work required to completely separate the nucleons.

We have already seen (Topic E.1) that it is the *strong nuclear force* that holds nucleons together in a nucleus. Because of these forces, we consider that all nuclei have *nuclear potential energy*. Importantly, the magnitude of that energy is enormous, considering the small size of nuclei.

If we wanted to completely separate the nucleons of a nucleus (another thought experiment which is impossible in practice), we would need to supply energy. After the nucleons were separated, they would then effectively have zero nuclear potential energy. As with gravitational or electric forces, the zero of nuclear potential energy is defined as occurring when the particles concerned are an infinite distance apart.

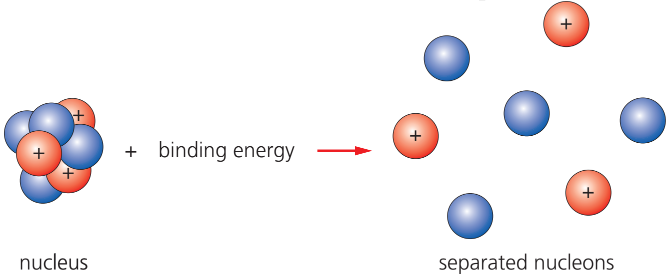


Figure E3.17 Binding energy is needed to separate nucleons; this example is lithium-7

The energy that would be needed to completely separate all the nucleons of a nucleus is called its nuclear **binding energy**.

Figure E3.17 shows an example.

Alternatively, we can consider that binding energy of a nucleus is the energy that would be *released* when the nucleus was formed from separate nucleons.

Nature of science: Models

Energy in bound systems

Consider any system in which there are attractive forces between the particles it contains. For example, between masses, between opposite charges, or between nucleons.

When the particles move, work will be done (provided the movement is not perpendicular to the force) and energy is transferred. We describe this as a change in the potential energy of the system.

If we wish to compare different systems, we need to agree on a common zero for potential energy: for this we choose the situation when the particles are a long way apart from each other (infinity), where the forces are zero.

Stationary separated particles which are free to move will be attracted closer together and gain kinetic energy and lose potential energy, so that their total energy remains the same. This implies that all potential energies in systems like these must be negative and that, if possible, any such systems will change to lower potential energy, when they could then be described as being more stable. In effect, this means that the potential energy of the system will change to a larger negative value as it becomes more stable.

If we wish to separate particles which are attracted together, for example nucleons in a nucleus, we need to do work (supply energy) so as to increase the potential energy of the system which, in effect, means to decrease the magnitude of negative potential energy. This can be confusing!

Nuclear binding energy is equivalent to nuclear potential energy in magnitude, but the concept of nuclear binding energy looks at the same situation in a different, perhaps less confusing, way. Binding energy is the energy that an *external agent* would need to supply to separate the nucleons (rather than being a property of the nucleus). Binding energy is always positive and using positive numbers is more intuitive, also larger positive binding energies corresponds well with greater stability.

A nucleus which had a potential energy of −50 MeV would be more stable than the same nucleus if it had a potential energy of −40 MeV. In terms of the binding energy for the same situations, we would say that a nucleus which had a binding energy of +50 MeV would be more stable than the same nucleus if it had a binding energy of +40 MeV.

The more nucleons in a nucleus (greater nucleon number, *A*), the greater the total binding energy will be.

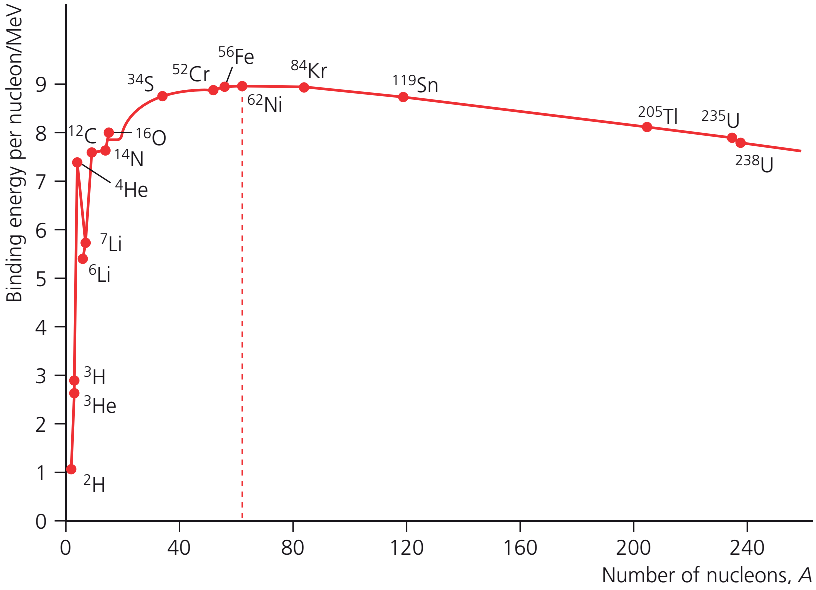
Binding energy is central to understanding nuclear events because when a nucleus changes in any way, the nuclear potential energy (binding energy) will also change. So, changes in binding energy can be more important than total binding energy, and in particular, the concept of average **binding energy per nucleon** is widely used.

◆ **Binding energy per nucleon (average)** Binding energy of a nucleus divided by the number of nucleons it contains. It is a measure of the stability of a nucleus.

$$\text{binding energy per nucleon} = \frac{\text{total binding energy}}{\text{number of nucleons in the nucleus}}$$

Binding energy per nucleon is a guide to a nucleus's stability.

Figure E3.18 shows how average binding energy per nucleon varies with nucleon number. A typical value is about 8 MeV per nucleon, but the variations seen in Figure E3.18 have important consequences, as we will explain later.



■ **Figure E3.18** A plot of binding energy per nucleon against number of nucleons

WORKED EXAMPLE E3.3

a Use Figure E3.18 to estimate the average binding energy per nucleon of the nuclide carbon-12.

b Calculate the total binding energy of the carbon-12 nuclide in:

i MeV **ii** joules.

c One mole of carbon atoms has a mass of 12 g.

Determine the total binding energy of that 12 g.

Answer

a 7.7 MeV/nucleon

b **i** 12 × 7.7 = 92 MeV (92.4 seen on calculator display)

ii $(92.4 \times 10^{6}) \times (1.60 \times 10^{-19}) = 1.5 \times 10^{-11}$ J (1.4784×10^{-11} J seen on calculator display)

c $(1.4784 \times 10^{-11}) \times (6.02 \times 10^{23}) = 8.9 \times 10^{12}$ J

This answer illustrates the truly enormous amount of nuclear energy associated with even a relatively small amount of matter. However, this energy is not usually accessible to us.

36 a Use Figure E3.18 to determine which nuclide is the most stable.

b Estimate the binding energy per nucleon of that nuclide.

37 a Use Figure E3.18 to estimate the binding energy per nucleon of oxygen-16.

b What is the total binding energy of the same nuclide?

38 If a nucleus of uranium-238 could be split in half to make two nuclei each of nucleon number 119 (this fission does not occur), estimate the change in overall binding energy.

39 If two hydrogen-2 nuclei could be combined to make one helium-4 nucleus, estimate the change in overall binding energy.

Nuclear fission and fusion

The last two questions illustrate two important types of nuclear reaction. Both of which can, under certain circumstances, release large amounts of energy from within a nucleus.

◆ **Nuclear fission** A nuclear reaction in which a massive nucleus splits into more stable smaller nuclei whose total binding energy is greater than the binding energy of the initial nucleus, with the release of energy.

◆ **Nuclear fusion** Nuclear reaction in which two low mass nuclei combine to form a more stable and more massive nucleus whose binding energy is greater than the combined binding energies of the initial nuclei, with the release of energy.

◆ **Energy–mass equivalence** Any mass is equivalent to a certain amount of energy, according to the equation $E = mc^2$.

Nuclear fission is the splitting of a massive nucleus into two smaller nuclei.

Topic E.4 is about nuclear fission.

Nuclear fusion is the combination of two small nuclei to produce a more massive single nucleus.

Topic E.5 discusses the process of nuclear fusion in stars.

Mass–energy equivalence

One of the consequences of Einstein's theory of special relativity was that particles (at rest) have intrinsic energy and that the amount of that energy, E, could be calculated from his famous equation:

$$E = mc^2$$

DB

Where m is the mass of the particle in kilogrammes and c is the speed of light in m s^{-1}.

c^2 is a constant, and mass and energy are **equivalent** to each other. As Einstein said (in a film he made in 1948 called 'Atomic Physics'):

> *'Mass and energy are both but different manifestations of the same thing.'*

If energy is added to, or removed from, a system by any means (examples: in chemical reactions, by heating or cooling, by changing speed, by moving up or down, and so on) then there will be a corresponding change in mass. Mass should not be considered to be an absolute, unchanging property of particles.

Top tip!

The equation $E = mc^2$ is being quoted here in its famous form. However, it may be more informative to express it as: $\Delta E = \Delta mc^2$, in order to stress that a *change* of energy, ΔE, is equivalent to a *change* in mass, Δm.

WORKED EXAMPLE E3.4

1.0 kg of water was raised in temperature by 10 °C. Calculate the corresponding change in mass. (Specific heat capacity of water is 4180 J °C^{-1} kg^{-1})

Answer

$Q = mc\Delta T = 1.0 \times 4180 \times 10 = 4.18 \times 10^4$ J. This amount of energy has increased the kinetic energy of the water molecules.

Then, $Q = E = mc^2$

$4.18 \times 10^4 = m \times (3.00 \times 10^8)^2$

Increase in mass = 4.64×10^{-13} kg

This is effectively unmeasurable and shows us that increases in masses involved in everyday activities are negligible. However, as we shall see, the changes in mass during nuclear reactions *are* significant.

Units of measurement for masses of atomic particles

◆ **Rest mass** Mass of an isolated particle that is at rest relative to the observer.

The **rest mass** of a particle is its mass as would be measured by an observer who is moving with the same velocity as the particle. That is, the particle would seem to be at rest as seen by the observer. This is the same as our usual understanding of mass.

The SI unit for mass is the kilogramme, but this may be considered to be an inconveniently large unit when quoting the values of the masses of atomic particles, as the following examples show:

- The rest mass of an isolated electron is **9.110×10^{-31} kg**.
- The rest mass of an isolated proton is **1.673×10^{-27} kg**.
- The rest mass of an isolated neutron is **1.675×10^{-27} kg**.

◆ **Atomic mass unit (amu), u** Unit of mass widely used in atomic physics. Approximately equal to the mass of a single nucleon. Defined to be exactly $1.660\,539\,066\,60 \times 10^{-27}$ kg.

As an alternative to the kilogramme, the masses of nuclides and subatomic particles are more usually quoted in terms of the equivalent number of nucleons. The (unified) **atomic mass unit**, u, (amu) is intended to represent the mass of a proton or a neutron (which are very similar, as can be seen above), so that the oxygen-16 nuclide, which has 16 nucleons, would have a mass of 16 u. However, this is not accurate enough for most nuclear physics calculations, so that a more precise definition is needed, and the carbon–12 atom was chosen as the standard as follows. (The mass of oxygen-16 is then 15.994914 u, rather than 16.)

The atomic mass unit, u, is defined to be exactly one twelfth of the mass of an isolated carbon-12 atom, that is:

1 u = 1.661×10^{-27} kg

We can now restate the masses of electrons, protons and neutrons (using a large number of significant figures):

- The rest mass of an isolated electron is **0.000 549 u**.
- The rest mass of an isolated proton is **1.007 276 u**.
- The rest mass of an isolated neutron is **1.008 665 u**.

It is useful to know how much energy corresponds to a mass of 1 u:

$E = mc^2 = 1\text{ u} \times c^2 = (1.6605 \times 10^{-27}) \times (2.9979 \times 10^8)^2 = 1.4924 \times 10^{-10}$ J (using 5 significant figures)

Converting to eV: $\dfrac{1.4924 \times 10^{-10}}{1.6022 \times 10^{-19}} = 9.315 \times 10^8$ eV, or 931.5 MeV

Summarizing, 1 u × c^2 = 931.5 MeV, or:

1 u = 931.5 MeV c^{-2}

We now have a third way of expressing the masses of electrons, protons and neutrons:

- The rest mass of an isolated electron is **0.511 MeV c^{-2}** (0.000549 u × 931.5 MeV c^{-2}).
- The rest mass of an isolated proton is **938 MeV c^{-2}**.
- The rest mass of an isolated neutron is **940 MeV c^{-2}**

WORKED EXAMPLE E3.5

The rest mass of an alpha particle is 6.644657 × 10^{-27} kg. Express this in:

a atomic mass units

b MeV c^{-2}.

Answer

a $\frac{6.644657 \times 10^{-27}}{1.661 \times 10^{-27}} = 4.000$ u

b 4.000 × 931.5 = 3726 MeV c^{-2}

Nuclear reactions involve changes of mass

We have seen that any change of energy of a system is accompanied by an equivalent change in mass, but such changes of mass are immeasurably small in the events of everyday life. However, because the strong nuclear forces between nucleons are relatively large and involve small masses, all nuclear reactions will involve a significant change of mass.

If a nucleon system becomes more stable (which is the usual course of events), energy is released, total mass decreases and the total binding energy increases (because the nucleons have become harder to separate).

WORKED EXAMPLE E3.6

Consider the spontaneous decay of radium-226.

$^{226}_{88}\text{Ra} \longrightarrow \, ^{222}_{86}\text{Rn} + \, ^{4}_{2}\alpha$

After the decay, the nucleus has become more stable, so that the total binding energy has increased, and the two particles have gained kinetic energy.

Consider the masses on both sides of this equation:

- Rest mass of radium = 226.0254 u
- Rest mass of radon = 222.0176 u
- Rest mass of alpha particle = 4.0026 u

Calculate:

a the change of mass that occurs during this decay

b the total kinetic energy of the resulting particles (MeV).

Answer

a 222.0176 + 4.0026 − 226.0254 = − 0.0052 u

b 0.0052 × 931.5 = 4.84 MeV

Most of this energy is carried by the smaller mass, the alpha particle.

Mass defect

In our thought experiment, we have seen that energy (binding energy) would have to be supplied to completely separate all the nucleons in a nucleus. In total, the separated nucleons have more potential energy than when they were together in the nucleus. The equivalence of mass and energy informs us that the total mass of the separated nucleons must be more than their total mass when they were combined in the nucleus. In other words:

The **mass defect** of a nucleus is the reduction in mass that occurs when separated nucleons combine together to form a nucleus. The mass defect is equivalent to the binding energy.

◆ **Mass defect** The difference in mass between a nucleus and the total mass of its nucleons if they were separated. Equal to nuclear binding energy.

Common mistake

The term *mass defect* should only be used for the change in mass when *all* the nucleons are separated (equivalent to binding energy). For example, the decrease in mass which occurs during radioactive decay (see Worked example E3.6) should *not* be described as a mass defect.

WORKED EXAMPLE E3.7

Calculate the mass defect (in electronvolts) and binding energy of a helium-4 atom (mass 4.00260 u). It consists of two protons (each of mass 1.007276 u), two neutrons (each of mass 1.008665u) and two electrons (each of mass 0.000549 u).

Answer

The total mass of the separate particles =

$(2 \times 1.007\,276) + (2 \times 1.008\,665) + (2 \times 0.000\,549) = 4.032\,98\,\text{u}$

Mass defect = $4.032\,98 - 4.002\,60 = 0.030\,38\,\text{u}$

$\Delta E = 0.030\,38 \times 931.5 = 28.30\,\text{MeV}$

WORKED EXAMPLE E3.8

A particular nucleus has a mass defect of 0.369 u.

a Calculate its binding energy in MeV.

b If it contains 40 nucleons, determine the average binding energy per nucleon.

Answer

a $0.369 \times 931.5 = 344\,\text{MeV}$ (343.72... seen on calculator display)

b $\dfrac{343.72}{40} = 8.59\,\text{MeV}$

40 The mass of a lithium-7 nucleus is 7.01600 u. Express this in:

a kilogrammes **b** MeV c^{-2}.

41 An aluminium-27 nucleus has a mass of 26.9815 u. Determine:

a its total binding energy

b its binding energy per nucleon.

42 A nuclide of $^{197}_{79}\text{Au}$ has a mass of 196.9665 u. Determine its mass defect.

43 Thorium-232 decays to form radium-228.

a What particle is emitted?

b Thorium-232 has a nuclear mass of 232.0381 u and radium-228 has a nuclear mass of 228.03107 u. Determine the energy released in this decay. (Alpha particle mass = 6.6447×10^{-27} kg)

44 A plutonium-239 nucleus can be split into smaller nuclei by interaction with a neutron as follows:

$^{239}_{94}\text{Pu} + ^{1}_{0}n \rightarrow ^{134}_{54}\text{Xe} + ^{103}_{40}\text{Zr} + 3^{1}_{0}n$

a Determine the energy released in this reaction in MeV (mass of Pu-239 nucleus = 239.0522 u, mass of Xe-134 nucleus = 133.9054 u, mass of Zr-103 nucleus = 102.9266 u).

b State the form of this released energy.

45 The following equation represents one kind of nuclear reaction that occurs in stars.

$4^{1}_{1}\text{H} \rightarrow ^{4}_{2}\text{He} + 2^{0}_{1}e + 2\nu + ^{0}_{0}\gamma$

Describe what is happening in this reaction.

The strong nuclear force and nuclear stability

SYLLABUS CONTENT

- Evidence for the strong nuclear force.
- Role of the ratio of neutrons to protons for the stability of nuclides.
- Approximate constancy of binding energy curve above a nucleon number of 60.

◆ **Mesons** Unstable subatomic particles involved with the strong nuclear force.

■ **Figure E3.19** Japanese physicist Hideki Yukawa in 1951

We have seen that the existence of an attractive *strong nuclear force* is needed to explain why the protons in a nucleus are not forced apart by the repulsion that occurs between similar charges.

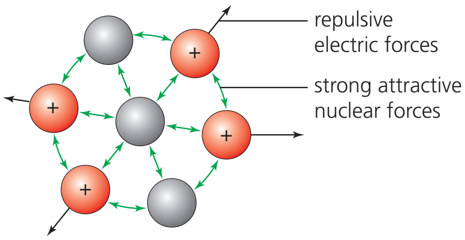
The Japanese physicist Hideki Yukawa (Figure E3.19) in 1935 proposed that the exchange of (as yet undiscovered) subatomic particles (called **mesons**) between nucleons was the cause of a *strong nuclear force* holding the nucleons together in a nucleus. His hypothesis was effectively proven when mesons were discovered in 1947.

In a stable nucleus, we can consider that the very short-range attractive strong nuclear forces are balanced by the longer range repulsive electric forces.

This is illustrated in Figure E3.20 with a nucleus which includes, as an example, four protons and three neutrons randomly arranged.

As can be seen in Figure E3.21, $^{7}_{4}$Li is a stable nuclide.

The stability of a nuclide depends on the ratio of neutrons to protons (*N/Z*) in its nucleus.

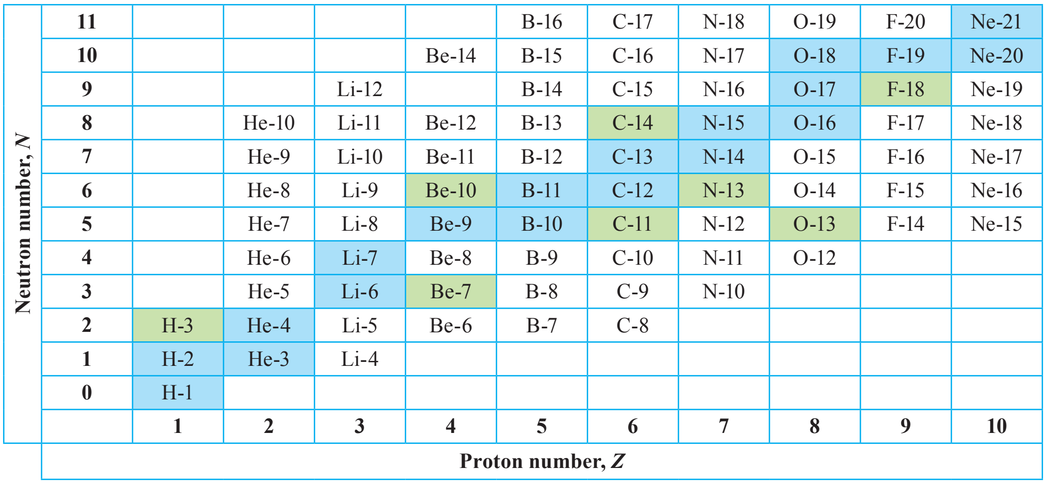


■ **Figure E3.20** Attractive and repulsive forces balance in a stable nucleus

Figure E3.21 shows a small part (the beginning) of a *chart of nuclides*. Each box on the chart corresponds to a particular nuclide. The boxes highlighted in blue show stable nuclides. The others have been produced artificially. The boxes highlighted in green show well-known radionuclides. A full chart of nuclides may show the relative abundances of the nuclides and the way in which unstable nuclides decay. (These charts are sometimes drawn with the axes reversed from that shown in Figure E3.21.)

Common mistake

Radioactivity is usually associated with the more massive nuclides, but Figure E3.21 illustrates the fact that *all* elements have radioisotopes.

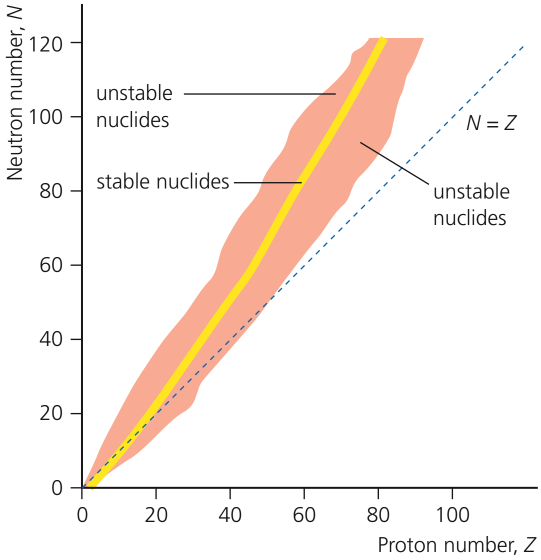


11					B-16	C-17	N-18	O-19	F-20	Ne-21
10				Be-14	B-15	C-16	N-17	O-18	F-19	Ne-20
9			Li-12		B-14	C-15	N-16	O-17	F-18	Ne-19
8		He-10	Li-11	Be-12	B-13	C-14	N-15	O-16	F-17	Ne-18
7		He-9	Li-10	Be-11	B-12	C-13	N-14	O-15	F-16	Ne-17
6		He-8	Li-9	Be-10	B-11	C-12	N-13	O-14	F-15	Ne-16
5		He-7	Li-8	Be-9	B-10	C-11	N-12	O-13	F-14	Ne-15
4		He-6	Li-7	Be-8	B-9	C-10	N-11	O-12		
3		He-5	Li-6	Be-7	B-8	C-9	N-10			
2	H-3	He-4	Li-5	Be-6	B-7	C-8				
1	H-2	He-3	Li-4							
0	H-1									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Neutron number, *N*

Proton number, *Z*

■ **Figure E3.21** A chart of the nuclides for $Z \leq 10$



■ **Figure E3.22** Stable nuclei shown on a chart of the nuclides

Figure E3.22 shows the overall pattern seen on a full chart of nuclides.

It is possible to identify several trends within the full chart of nuclides. Most importantly, from Figure E3.22, we can see that the neutron / proton (N/Z) ratio of a nuclide is a rough guide to its possible stability. $^{208}_{82}Pb$ is the nuclide with the largest nucleon number which is stable.

For nuclides with $Z < 20$, stable nuclides have $N/Z \approx 1$; For larger nuclides N/Z gradually increases to a maximum of about 1.5.

The reason for the increasing N/Z ratio of stable nuclei is as follows. In smaller nuclei with fewer nucleons, the *short-range* attractive nuclear force from any particular nucleon will have some effect on all the other surrounding nucleons. However, in larger nuclei some nucleons will be far enough apart from each other that the nuclear forces between them become less significant. The *longer range* repulsive coulomb force between protons could then make the nucleus unstable. The addition of extra neutrons (affected by the attractive nuclear force, but not the repulsive coulomb force) results in stability.

Variations in binding energy per nucleon

Adding more nucleons to a nucleus clearly increases its *total* binding energy, but we need to have some understanding of why the *average* binding energy per nucleon (and nuclear stability) varies as seen in Figure E3.22.

Imagine adding a nucleon to a nucleus *which only has a few nucleons*: the strong nuclear force will act between the additional nucleon and its closest 'neighbour', but also the surrounding nucleons. The total binding energy will increase. Then, imagine another nucleon is added: the total binding energy will increase by more than the previous amount because the strong nuclear force is affecting more nucleons close to the new nucleon. In this way, the average binding energy per nucleon increases.

However, because of the short range of the strong nuclear force, its effect on nucleons that are *not* relatively close together is insignificant. This explains why:

adding neutrons to larger nuclei with $A >$ about 60 increases the total binding energy, but has little effect on the binding energy per nucleon.

46 Calculate the N/Z ratios for the following stable nuclides:

a $^{12}_{6}C$ **b** $^{107}_{47}Ag$ **c** $^{208}_{82}Pb$.

47 Estimate the number of nucleons in stable isotopes of the following elements:

- **a** boron ($Z = 5$)
- **b** bromine ($Z = 35$)
- **c** mercury ($Z = 80$).

48 Refer to Figure E3.21.

Suggest why the following nuclides are unstable:

- **a** carbon-14
- **b** nitrogen-12.

49 Explain why a more massive nucleus needs more neutrons per proton (than a less massive nucleus) in order to be stable.

50 Describe the variation of binding energy per nucleon shown in Figure E3.18.

51 **a** Use Figure E3.18 to estimate the binding energy per nucleon for nuclides with nucleon numbers 100, 150 and 200.

b Do you agree that your answers are 'approximately constant'?

c Suggest why binding energy per nucleon shows much greater variation for smaller nuclei.

What can we learn from the spectra of alpha, beta and gamma radiations?

SYLLABUS CONTENT

- The spectrum of alpha and gamma radiations provides evidence for discrete nuclear energy levels.
- The continuous spectrum of beta decay as evidence for the neutrino.

Alpha particle spectrum

Many radionuclides which emit alpha particles, emit them with only *one* precise energy. However, some radionuclides can emit alpha particles with different energies as displayed in an **alpha particle spectrum**. This provides physicists with important information about energy levels within nuclei.

The *spectrum* of nuclear radiations emitted from an unstable nucleus describes the relative numbers of particles emitted with different energies.

◆ **Alpha particle spectrum** The range of discrete energies possessed by alpha particles emitted from the same radionuclide.

◆ **Excited state** When a particle is at a higher energy level than its ground state.

For example, nuclei of americium-241 emit alpha particles in the process of decaying to nuclei of neptunium-237.

$$^{241}_{95}\text{Am} \rightarrow {}^{237}_{93}\text{Np} + {}^{4}_{2}\alpha$$

An alpha particle emitted from americium-241 will have one of the following energies:

- 5.389 MeV (1.0%)
- 5.443 MeV (12.5%)
- 5.486 MeV (86.0%)
- 5.512 MeV (0.2%)
- 5.545 MeV (0.3%).

The spectrum is shown in Figure E3.23.

It is important to realize that the energy of any emitted alpha particle can only have one of these energies. The energies are *discrete* and the spectrum is not continuous. All alpha particles have discrete energies and all alpha particles from a particular radionuclide will have the same energies.

Alpha particles with different energies are possible because the nucleus of the daughter product (neptunium-237) can be left in its ground state, or in one of several discrete **excited states** (an energy state above the ground state). Figure E3.24 shows the emission of alpha particles of five different energies from americium-241 to various nuclear energy levels of neptunium-237. The total energy released from the Am-241 nucleus is always 5.545 MeV.

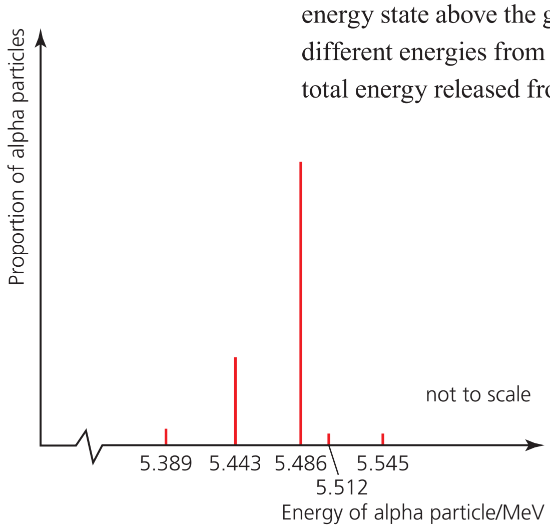


Figure E3.23 Spectrum of alpha particles

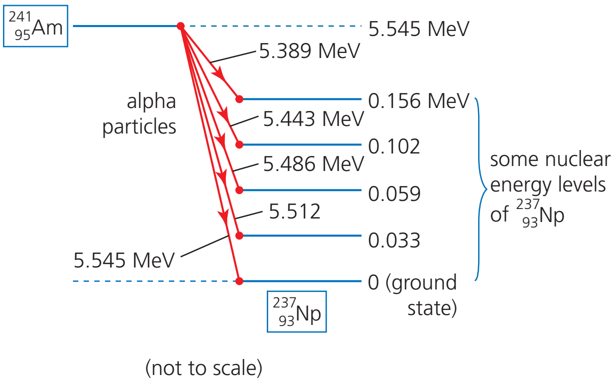


Figure E3.24 Energies of alpha particles emitted from americium-241

◆ **Gamma ray spectrum**
Range of discrete photon energies that may be emitted from a single radionuclide.

Gamma ray spectrum

Consider Figure E3.24 again. After the alpha decays, four excited states of the neptunium-237 nuclide can be seen. Afterwards, when a nucleus changes from an excited state to a lower energy level, a gamma ray photon will be emitted. The nucleus cannot emit a continuous range of gamma rays. The discrete energies of the photons (as seen in a **gamma ray spectrum**) again provide evidence for discrete energy levels within nuclei.

Figure E3.25 shows another example, nuclear energy levels within the magnesium-24 nuclide after beta-negative decay from sodium 24 left the nucleus in an excited state. Five prominent transitions to lower energy levels have been shown.

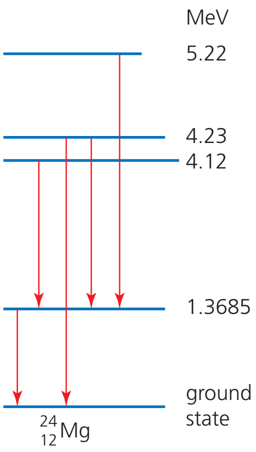


Figure E3.25 Excited energy levels within the magnesium-24 nuclide

The discrete energy levels of nuclei are the reason why alpha particles and gamma rays are emitted with discrete energies.

LINKING QUESTION

- How did conservation lead to experimental evidence of the neutrino? (NOS)

This question links to understandings in Topics A.2 and A.3.

Beta particle spectrum

The **spectrum of beta particles** emitted from a particular radionuclide is very different from alpha particles. See Figure E3.26 and compare it to Figure E3.23.

This range of beta particle energies was puzzling for physicists: why were beta particles different from alpha particles, which all have the same energy(s) from the same source? Without further information, it appeared that beta particle emission could break the laws of conservation of energy and momentum.

We now know that when scientists first detected the emission of beta-negative particles (1899), they were unaware of the undetected antineutrinos which were also emitted. It was more than thirty years, in 1930, before the Austrian physicist Wolfgang Pauli hypothesized the involvement of unknown particles in beta decay. See Figure E3.27, which shows possible relative motions of the three particles involved after the beta-negative decay of cobalt-60. The conservation of energy and momentum can only be explained by the emissions of *two* particles from the nucleus at varying angles. The energies of the beta particles will vary with the angles involved.

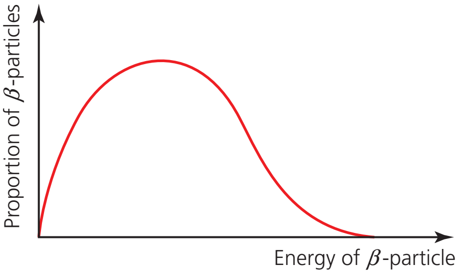


Figure E3.26 Typical energy spectrum for beta decay

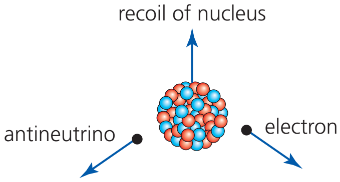


Figure E3.27 Beta-negative decay

The still undiscovered particles were named neutrinos (and antineutrinos) because it was believed that they must be uncharged and have a very small mass – the reasons why they are very hard to detect. Their existence was not confirmed until 1956.

◆ **Beta particle spectra**
The continuous range of different energies possessed by beta particles emitted from the same radionuclide.

The emission of a beta particle involves another particle (an undetected neutrino or antineutrino). The particles may travel in random directions, so that a continuous range of beta particle energies is possible.

ATL E3C: Research skills

Using search engines and libraries effectively

It is believed that neutrinos are the second most common type of particle in the Universe (after photons). It is estimated that over 10^{12} neutrinos pass through a fingernail every second. They usually pass through the entire Earth without being affected or detected.

Use the internet to learn about the Long Base-line Neutrino Experiment / DUNE, which is yet to be completed in the USA.

■ **Figure E3.28** The building at the top of the 'ice cube' neutrino detector, which is deep underground near the South Pole

TOK

Knowledge and the knower

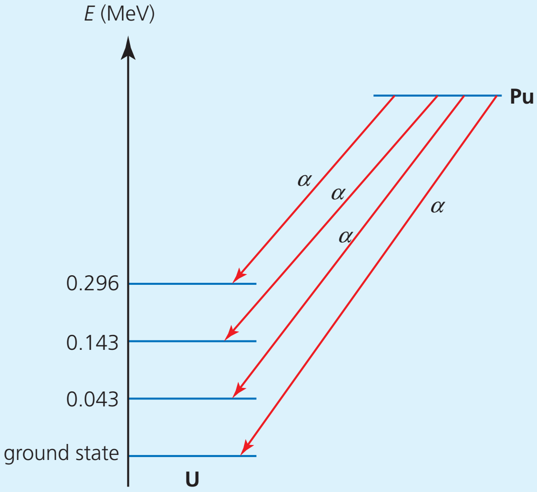
Of course, scientists involved in research (for example, into subatomic particles such as the neutrino) will claim to be open-minded and receptive to new ideas. Undoubtedly, this is a primary aim of all scientists, but to some extent we are all inevitably influenced and restricted by our previous experiences, and the culture of the society in which we live.

Expensive projects need to have specific aims (or governments would not provide funding), and these will focus and direct the thinking of the scientists involved. However, it is clear from the development of science over the centuries that it is in the nature of science and scientists to use their imagination to formulate new and original ideas. It is possible that no 'single truth' about neutrinos (in the sense of certain knowledge for all time) will emerge from the latest research, although it is to be hoped that our understanding of these elusive fundamental particles will expand.

52 Figure E3.29 shows four possible alpha particle decays for a plutonium-238 nuclide.

a If the most energetic alpha particle has an energy of 5.510 MeV, determine the energies of the other three.

b After an alpha decay, the daughter nuclide (uranium-234) is in an excited state.
Calculate the highest frequency of gamma ray that could then be emitted.



■ **Figure E3.29** Four possible alpha particle decays for a plutonium-238 nuclide to uranium-234

53 a If an isolated neutron at rest decayed into a proton and an electron (an antineutrino is also emitted), estimate the energy (MeV) that would be released.

b Explain why most of this energy will be transferred to the kinetic energy of the electron (beta-negative particle).

54 Compare the magnitude of typical nuclear energy levels with typical electron energy levels.

55 The radionuclide ${}^{15}_{6}C$ undergoes beta-negative decay. The excited state of the daughter product then emits a gamma ray as it moves down to its ground state. Sketch an energy level diagram to represent these changes.

56 Suppose that it is known that a particular nuclide emits gamma ray photons of energies 0.58 MeV, 0.39 MeV and 0.31 MeV when transitioning to the ground state. Determine the energies of another three different photons that might be emitted from the same nucleus.

Radioactive decay in more mathematical detail

SYLLABUS CONTENT

- The decay constant λ and the radioactive decay law as given by: $N = N_0 e^{-\lambda t}$.
- The decay constant approximates the probability of decay in unit time only in the limit of sufficiently small λt.
- Activity as the rate of decay as given by: $A = \lambda N = \lambda N_0 e^{-\lambda t}$.
- The relationship between half-life and the decay constant as given by $T_{\frac{1}{2}} = \dfrac{\ln 2}{\lambda}$.

Earlier in this topic the concept of *half-life* was introduced as a way of representing the pattern of decreasing activity, or count rate, from a radioactive source. We now want to develop that idea so that we can determine values of activity, or count rate, at *any time,* not just for times which are whole number multiples of the half-life. For example, we may wish to know what the activity of a source, which has a half-life of 5.3 years, will be in a year's time.

If a quantity, N, changes by an amount ΔN in a time Δt, the rate of change is $\Delta N/\Delta t$. There are many examples in science, and in everyday life, where a rate of change at any time depends on the quantity at that time. If a rate of change is always proportional to the quantity, it is described as an *exponential change*:

$$\frac{\Delta N}{\Delta t} \propto N$$

Putting in a constant, we get:

$$\frac{\Delta N}{\Delta t} = \lambda N$$

For radioactive decay (an exponential decrease), ΔN will be negative, so that:

◆ **Decay constant, λ,** The probability of decay of an unstable nucleus per unit time: $\lambda = (-\Delta N/N)/\Delta t$ (SI unit: s^{-1}).

$$\frac{\Delta N}{\Delta t} = -\lambda N$$

where λ (which is positive) is called the **decay constant**.

$$\text{Decay constant, } \lambda = -\frac{\Delta N/N}{\Delta t} = \text{probability of a nucleus decaying in unit time}$$

SI unit: s^{-1}

For example, if $N = 200$ and in time $\Delta t = 1$ s, N decreases by 5, $\Delta N = -5$, then:

$$\lambda = -\frac{\Delta N/N}{\Delta t} = -\frac{-5/200}{1} = 0.025\,\text{s}^{-1}$$

However, we need to be careful when making calculations like this, because if $\dfrac{\Delta N}{N}$ $(= -\lambda\Delta t)$ is too large, it will vary significantly during time Δt.

Larger values of a decay constant correspond to quicker decreases and shorter half-lives.

We saw earlier in this topic that the activity, A, of a radioactive source is the number of nuclei decaying every second (unit: Bq). We can write this as:

$$A = \frac{\Delta N}{\Delta t}$$

so that:

activity, $\boldsymbol{A = \lambda N}$

WORKED EXAMPLE E3.9

The activity of a radioactive sample is 2.5×10^5 Bq. The sample has a decay constant of $1.8 \times 10^{-6}\ \mathrm{s}^{-1}$. Determine the number of undecayed nuclei remaining in the sample at that time.

Answer

$$A = \lambda N$$

$$2.5 \times 10^5 = (1.8 \times 10^{-6}) \times N$$

$$N = \frac{2.5 \times 10^5}{1.8 \times 10^{-6}}$$

$= 1.4 \times 10^{11}$ undecayed nuclei.

WORKED EXAMPLE E3.10

The radionuclide Tc-99 has a decay constant of $0.115\ \mathrm{h}^{-1}$.

Calculate the percentage of the nuclei that decay every:

a hour

b minute (assume the activity is constant over the hour).

Answer

a $$\lambda = -\frac{\Delta N/N}{\Delta t}$$

$$0.115 = -\frac{\Delta N/N}{1.0}$$

$$-\left(\frac{\Delta N}{N}\right) = 0.115$$

which is equivalent to 11.5 %

b $$\frac{11.5}{60} = 0.192\ \%$$

◆ **Law of radioactive decay** The rate of decay is proportional to the number of undecayed nuclei, $\Delta N/\Delta t = -\lambda N$.

Exponential decay equations $N = N_0 e^{-\lambda t}$. N_0 is the number of undecayed nuclei at the start of time t and N is the number remaining at the end of time t. Alternatively, equations of the same form can be used with activity, A, or the count rate. Activity is linked to the initial number of atoms by the equation $A = \lambda N_0 e^{-\lambda t}$.

Solutions to the decay equation

Although the equation $\frac{\Delta N}{\Delta t} = -\lambda N$ (the **law of radioactive decay**) defines the mathematics of radioactive decay, it does not directly provide us with what we are most likely to want to know – the value of N, A, or a count rate at *any* time, t. For that we can use the following equations, which may be described as solutions to the previous equation. You are *not* expected to understand the origin of these **exponential decay equations**.

number of undecayed nuclei, $\boldsymbol{N = N_0 e^{-\lambda t}}$

In this equation, N_0 represents the number of undecayed nuclei in a source at the beginning of a time t, and N represents the number of undecayed nuclei at the end of time t. Similarly:

activity, $A = A_0 e^{-\lambda t}$

where A_0 represents the activity from a source at the beginning of a time t, and A represents the activity at the end of time t.

Since $A = \lambda N$, the equation can also be expressed as:

activity, $\boldsymbol{A = \lambda N_0 e^{-\lambda t}}$

The activity of a source is not easily determined, as is explained below. The count rate, C, *(a non-standard symbol)* measured by a radiation detector in a laboratory is not directly measuring the activity of the source, but it will usually be proportional to the activity, so that we can also write:

$$C = C_0 e^{-\lambda t}$$

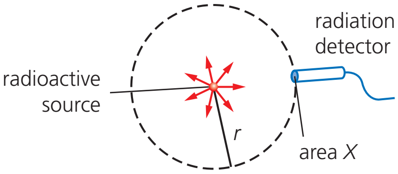


Figure E3.30 A radiation detector receives only a fraction of the radiation emitted from a source

Figure E3.30 shows that a detector can only receive some of the total radiation from a source.

We could use the ratio $X/4\pi r^2$ to determine a value for the fraction of the radiation emitted that arrives at the detector, but that would assume that:

- No radiation was absorbed between the source and the detector.
- The radiation was emitted equally in all directions.

Determining the activity of the source would also be difficult because not all of the radiation passing into the detector will be counted.

Tool 3: Mathematics

Carry out calculations involving logarithmic and exponential functions

$N = N_0 e^{-\lambda t}$

Suppose we are given N_0 and λ and wish to find a value for N at a known time, t, later:

Take natural logarithms of both sides:

$\ln N = \ln N_0 - \lambda t$

Suppose we are given N_0 and N at a known time, t, later, and wish to find a value for λ.

Divide both sides by N_0:

$$\frac{N}{N_0} = e^{-\lambda t}$$

Take natural logarithms of both sides:

$$\ln\left(\frac{N}{N_0}\right) = -\lambda t \quad \text{or,} \quad \ln\left(\frac{N_0}{N}\right) = +\lambda t$$

WORKED EXAMPLE E3.11

The number of radioactive atoms in a source decays by 7/8 in 12 days. Predict the fraction of radioactive atoms remaining after 24 days.

Answer

This can be done in a straightforward way without using exponentials: the source passes through three half-lives to reduce to 1/8, so the half-life is four days. In 24 days, there are six half-lives, so the fraction reduces to 1/64 ($1/2^6$).

We could get the same answer using the exponential equation $N = N_0 e^{-\lambda t}$.

After 12 days: $N/N_0 = 1/8 = e^{-\lambda(12)}$

$8 = e^{12\lambda}$

$\ln 8 = 12\lambda$

$\lambda = \ln 8/12 = 0.173$

After 24 days: $N/N_0 = e^{-0.173(24)} = 0.0157 = 1/64$

WORKED EXAMPLE E3.12

The decay constant for a radioisotope is $0.054\,\text{y}^{-1}$. If the activity at the start of year 2022 was 470 Bq, calculate the activity at the start of year 2026.

Answer

$A = A_0 e^{-\lambda t}$

$A_{2026} = 470 \times e^{-0.054 \times 4}$

$\ln A_{2026} = \ln 470 - (0.054 \times 4)$

$A_{2026} = 379\,\text{Bq}$

57 A sample of radium-226 contains 6.64×10^{23} radioactive atoms. It emits alpha particles and has a decay constant of $1.38 \times 10^{-11}\,s^{-1}$. Determine how many atoms of radium-226 are left after 1000 years.

58 A radioactive nuclide has a decay constant of $0.0126\,s^{-1}$. Initially a sample of the nuclide contains 1.0×10^{10} nuclei.

a Calculate the initial activity of the sample.

b Predict how many nuclei remain undecayed after 200 s.

59 a The count rate from a radioactive source is measured to be $673\,s^{-1}$ (adjusted for background count). If exactly three hours later the count rate has reduced to $668\,s^{-1}$, determine a value for **i** the decay constant and **ii** the half-life.

b Explain why the value may be unreliable.

60 A source has an activity of 4.7×10^4 Bq. Assuming that the background count is negligible and the experiment is carried out in a vacuum, predict the count rate that could be recorded by a detector that has an effective receiving area of $0.85\,cm^2$ if it was placed:

a 50 cm from the source

b 5 cm from the source.

61 The activity from a radioactive source is 8.7×10^5 Bq. If its decay constant is $6.3 \times 10^{-6}\,s^{-1}$, calculate how many days will pass before the activity falls to 1.0×10^4 Bq.

62 A radioactive source of gamma rays, cobalt-60, is commonly used in school demonstrations. If the maximum allowable activity is 200 kBq, calculate the maximum mass of cobalt-60 in a school source. (Decay constant for cobalt-60 is $0.131\,y^{-1}$. Molar mass of Co-60 is 59.93 g)

Decay constant and half-life

The concept of the half-life of a radioactive nuclide was introduced earlier in this topic. Using the equation $N = N_0 e^{-\lambda t}$, it is straightforward to derive an equation that relates the half-life, $T_{½}$, to the decay constant, λ.

For any radioactive nuclide, the number of undecayed nuclei after one half-life is, by the definition of half-life, equal to $N_0/2$, where N_0 represents the original number of undecayed nuclei. Substituting this value for N in the radioactive decay equation at time $t = T_{½}$ we get:

$$\frac{N_0}{2} = N_0 e^{-\lambda T_{½}}$$

Dividing each side of the equation by N_0:

$$\frac{1}{2} = e^{-\lambda T_{½}} \quad \text{or,} \quad 2 = e^{\lambda T_{½}}$$

Taking natural logarithms

$$\ln 2 = \lambda T_{½}$$

So that:

Connection between half-life and decay constant:

$$\boldsymbol{T_{½} = \frac{\ln 2}{\lambda}}$$

Alternatively, inserting a value for ln 2:

$$T_{½} = \frac{0.693}{\lambda}$$

WORKED EXAMPLE E3.13

A radioactive source gives a count rate of 100 s^{-1} at a certain instant of time. After 100 s the count rate drops to 20 s^{-1}. The average background count rate is measured to be 1.5 s^{-1}.

Calculate the half-life of the source. Assume that the count rate is a measure of the activity.

Answer

Initial count rate due to source = 100 – 1.5 = 98.5 s^{-1}.

Count rate due to source after 100 s = 20 – 1.5 = 18.5 s^{-1}.

$$C = C_0 e^{-\lambda t}$$

$$18.5 = 98.5 \times e^{-100\lambda}$$

$$100\lambda = \ln\left(\frac{98.5}{18.5}\right)$$

$$\lambda = 0.016\,72$$

$$T_{\frac{1}{2}} = 0.693/0.016\,72 = 41 \text{ s}$$

63 Radioactive carbon-14 in a leather sample decays with a half-life of 5730 years.

- **a** Determine the decay constant of C-14.
- **b** Calculate the percentage of radioactive carbon remaining after 10 000 years.

64 At a certain time, a pure source contained 3.8×10^{15} radioactive atoms. Exactly one week later the number of radioactive atoms had reduced to 2.8×10^{14}.
Calculate the half-life of this source.

65 The activity of an americium-241 source used in a school laboratory was 1.6×10^5 Bq. Am-241 has a half-life of 432 years.

- **a** Show that the source contained about 3×10^{15} americium-241 atoms.
- **b** Determine the mass of this number of americium atoms.

66 The half-life of strontium-90 is 28.8 years.
Calculate how long it will take for the count rate from a sample to fall by:

- **a** 1%
- **b** 99%.

67 A cobalt-60 source used in radiotherapy in a hospital needs to be replaced when its activity has reduced to about 40% of its value when it was purchased.
If it was purchased December 2015 and replaced in November 2022, estimate its half-life.

68 The (adjusted) count rate from a radioactive source is 472 s^{-1}.
If its half-life is 13.71 minutes, predict the count rate exactly 1 hour later.

69 Uranium-238 is the most common isotope of uranium. It has a half-life of 4.47 billion years. The age of the Earth is 4.54 billion years.
Calculate the percentage (to 1 decimal place) of the original uranium-238 still present in the Earth's crust.

70 The radionuclide $^{40}_{19}K$ has a half-life of 1.3×10^9 y. It decays into the stable nuclide $^{40}_{18}Ar$. Rocks from the Moon were found to contain a ratio of potassium to argon atoms of approximately 1:7.

- **a** Estimate the age of these rocks.
- **b** Discuss whether the answer to part **a** is also a reasonable approximation for the age of the Moon. Explain your answer.

Experimental determination of half-life

If the count rate from a radioactive source is measured for more than one half-life, it is straightforward to determine a value for that half-life (as discussed earlier in this topic). We will now explain:

- how the calculation can be made more accurate
- how to determine the half-life of a radionuclide which is so long that no change in count rate can be detected.

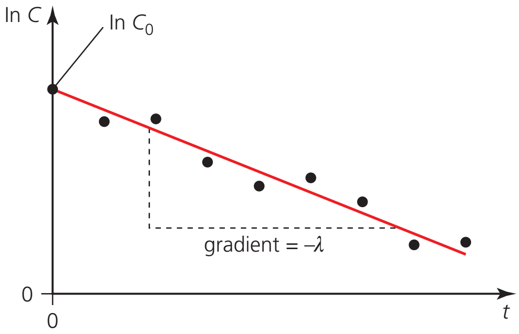


Figure E3.31 A plot of the natural logarithm of the count-rate against time

Radioisotopes with relatively short half-lives

This method is suitable if the count rate from a radionuclide decreases measurably over the time available for the experiment, but it is not necessary that the half-life is less than duration of the experiment.

The decay constant can be determined directly from the equation $C = C_0e^{-\lambda t}$, but a more accurate method involves plotting a graph of the natural logarithm of the count rate, C, against time, t. This should give a straight best-fit line (Figure E3.31).

Taking natural logarithms of $C = C_0e^{-\lambda t}$: $\ln C = \ln C_0 - \lambda t$. This equation can be compared to the equation for a straight line ($y = mx + c$), showing that the gradient is equal to $-\lambda$, from which the half-life can be calculated.

Inquiry 2: Collecting and Processing data

Processing data

Determination of half-life

Table E3.4 shows the results from a 2.5 h experiment to determine the half-life of radionuclide. The average background count was 0.20 every second.

Draw a graph of the natural logarithm of the adjusted count rate against time. Use the graph to determine a value for the half-life of the radionuclide. Discuss the accuracy of this experiment.

Table E3.4 Results of half-life experiment

Time	Count rate/min^{-1}
9.30 am	96
10.00 am	88
10.30 am	81
11.00 am	75
11.30 am	69
12.00 pm	64

Radioisotopes with relatively long half-lives

If the count rate does not change during the course of an experiment (other than random variations), we cannot use the equation $C = C_0e^{-\lambda t}$ to determine a decay constant (and half-life). Instead,

To determine the decay constant of radionuclides with long half-lives, we use the equation $A = \lambda N$

We would need to measure the activity, A, (not the count rate) and the number of undecayed nuclei, N, in a sample. However, this is difficult and it will not be possible in a school laboratory.

The activity can be determined from the count rate and the geometry of the apparatus (see Figure E3.30). The number of undecayed atoms in a *pure* sample, N, can be determined from its mass, m.

$$N = \frac{mN_A}{\text{relative atomic mass}} \text{ (see Topic B.3)}$$

WORKED EXAMPLE E3.14

A radioactive source of cobalt-60 (only) is known to produce an activity of 4.37×10^8 Bq. Estimate (i) the decay constant for this radionuclide if its mass is 10.5 µg, and (ii) the half-life of cobalt-60.

Answer

(i) First, we need to determine the number of atoms of cobalt-60:

$$N = \frac{mN_A}{\text{relative atomic mass}} = \frac{(10.5 \times 10^{-6}) \times (6.02 \times 10^{23})}{60} = 1.05 \times 10^{17}$$

Then use $A = \lambda N$

$4.37 \times 10^8 = \lambda \times (1.05 \times 10^{17})$

$\lambda = 4.1 \times 10^{-9}\,\text{s}^{-1}$ ($4.1619... \times 10^{-9}$ seen on calculator display)

(ii) $T_{\frac{1}{2}} = \ln 2/\lambda = 1.67 \times 10^8$ s or 5.3 years

71 The cadmium-109 radioisotope has a half-life of 463 days.
Predict the activity you would expect from a mass of 4.3×10^{-4} g of pure cadmium-109.

72 A pure radioisotope has a half-life of 49 hours and is producing a count rate of 81 every minute. The background count averages at 18 every minute.
Discuss, using appropriate calculations, whether it would be possible to accurately confirm the half-life experimentally over the course of a morning in the laboratory.

73 A small rock of mass 214 g is known to contain 8.2% uranium-238 (by mass). U-238 has a half-life of 4.5 billion years. There are no significant amounts of any other radioisotope.
What level of radioactivity will be emitted in this rock?

E.4 Fission

Guiding questions

- In which form is energy stored within the nucleus of the atom?
- How can the energy released from the nucleus be harnessed?

Nuclear fission

SYLLABUS CONTENT

▶ Energy is released in spontaneous and neutron-induced fission.

Nuclear fission was introduced briefly in Topic E.4, we will now look at it in more detail. Fission is the splitting of a massive nucleus into two smaller nuclei.

◆ **Fission fragments** The nuclei produced in a fission reaction.

After nuclear fission occurs, the resulting nuclei are called **fission fragments**. Typically, neutrons and gamma rays are also released during the fission.

Nuclear fission can sometimes be a *spontaneous* (random, without cause) form of radioactive decay of very massive nuclei. The following fission / decay of californium-252 (a nuclide which does not occur naturally) is an example.

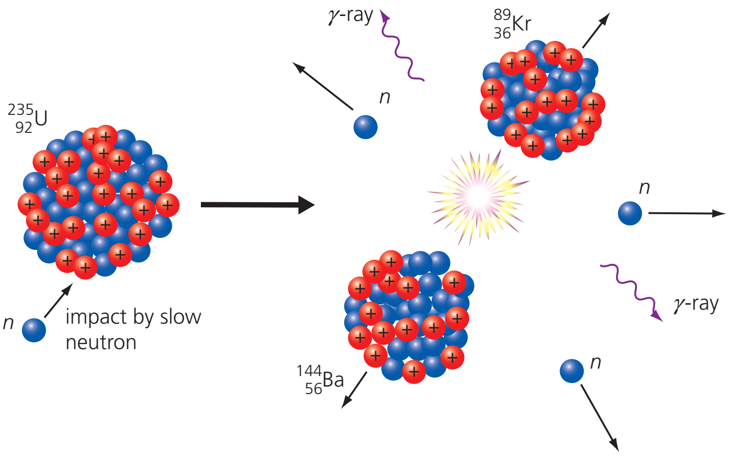
$$^{252}_{98}\text{Cf} \longrightarrow {}^{142}_{56}\text{Ba} + {}^{106}_{42}\text{Mo} + 4^{1}_{0}n$$

As we saw in Topic E.3, a detailed analysis of the masses involved in nuclear reactions can lead to a determination of the energy released.

More important than spontaneous fission, fission can be deliberately *induced* in nuclei which would not otherwise readily split into fragments. Most commonly, nuclear fission is induced by using *slow-moving* neutrons.

See Figure E4.1, which shows the well-known and important induced fission of uranium-235.

Figure E4.1 Fission of uranium-235



◆ **Fissile** Capable of sustaining a nuclear fission chain reaction.

Uranium-235 is the world's only *naturally occurring* nuclide that is capable of sustained nuclear fission (it is said to be **fissile**). There are other fissile nuclides (plutonium-239, for example) but they are not found naturally: they need to be produced in nuclear reactors.

$$^{1}_{0}n + ^{235}_{92}\mathrm{U} \rightarrow \left[^{236}_{92}\mathrm{U}\right] \rightarrow ^{144}_{56}\mathrm{Ba} + ^{89}_{36}\mathrm{Kr} + 3^{1}_{0}n \text{ (+ gamma ray photons and antineutrinos)}$$
$$+ \text{ about } 180\,\text{MeV of energy}$$

You need to understand this equation, but you do *not* need to remember it.

In this example, the fission fragments are barium and krypton, but these fragments can vary. The number of neutrons released can also vary.

The neutron is absorbed by the uranium-235 nucleus, becoming uranium-236 which very quickly fissions as shown.

Three further neutrons are produced, which means that further fission may be possible (see later).

The amount of energy released is considerable (≈ 180 MeV for each fission reaction).

LINKING QUESTION

- In which form is energy released as a result of nuclear fission?

The majority of the energy released is the form of kinetic energy of the fragments. Neutrons and gamma rays also transfer significant energy.

More energy is released later as the fragments decay radioactively and the neutrons and gamma rays are absorbed. A total of over 200 MeV will be transferred to internal energy in the material containing the uranium-235. This is the energy that is utilized in nuclear reactors which generate electrical energy.

Changes of binding energy and mass during nuclear fission

For nuclear fission to occur, the fission fragments must be more stable than the original nucleus. That is, the fission fragments must have higher values of binding energy per nucleon than the original nucleus undergoing fission.

Looking again at Figure E3.18 we can see that fission will only be theoretically possible for nuclides which can split into two nuclides which have a greater nucleon number than (approximately) nickel-62. This is also represented in Figure E4.2.

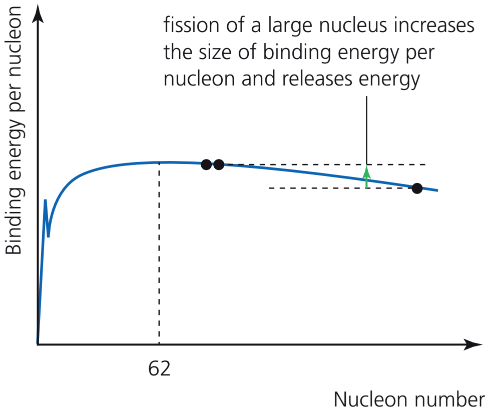


Figure E4.2 Fission increases binding energy per nucleon

WORKED EXAMPLE E4.1

Returning to the fission of uranium-235, The average binding energies (MeV) per nucleon involved are uranium 235: 7.59; krypton 89: 8.72; barium 144: 8.27.

Determine the change in total binding energy in this fission process.

Answer

$(89 \times 8.72) + (144 \times 8.27) - (235 \times 7.59) \approx 183\,\text{MeV}$ (increase)

The increase in binding energy calculated in the Worked example E4.1 is accompanied by the release of the same amount of energy in the fission process, as shown in the equation above.

The release of energy from the nuclei must result in a decrease in mass (see mass–energy equivalence in Topic E.3), as follows for the fission of uranium-235:

- Rest mass of each neutron = 1.008 665 u.
- Rest mass of uranium-235 = 235.043 930 u.
- Rest mass of barium-144 = 143.922 955 u.
- Rest mass of krypton-89 = 88.917 836 u.
- Rest mass on left-hand side of the equation = 236.052 595 u.
- Rest mass on right-hand side of the equation = 235.866 786 u.
- Mass difference = 0.185 809 u.

This mass difference is equivalent to 173 MeV (0.185809 × 931.5), which is in reasonable agreement with the 180 MeV quoted above.

1 Consider the spontaneous fission:
$^{252}_{98}\mathrm{Cf} \rightarrow {}^{142}_{56}\mathrm{Ba} + {}^{106}_{42}\mathrm{Mo} + 4{}^{1}_{0}n$
Use the internet to determine the relevant masses, and hence calculate a value for the total energy released in this fission.

2 Explain why a nuclide such as bromine-79 cannot undergo nuclear fission.

3 One possible induced fission of $^{235}_{92}\mathrm{U}$ produces Xe-140 (proton number = 54) and Sr-94, after the uranium nucleus captures a neutron.
  a Write a nuclear equation for this process.
  b Determine the energy released. (mass of uranium-235 = 235.044 u, mass of xenon-140 = 139.922 u, mass of strontium-94 = 93.915 u).
  c The mass defect of the xenon-140 nucleus is 1.2461 u. What is its binding energy per nucleon (MeV)?
  d Compare you answer to part **c** with the binding energy per nucleon of uranium-235 (use the internet for data).

4 The following is a possible fission of a plutonium nuclide.
$^{1}_{0}n + {}^{239}_{94}\mathrm{Pu} \rightarrow {}^{?}_{?}? \rightarrow {}^{134}_{54}\mathrm{Xe} + {}^{?}_{?}\mathrm{Zr} + 3{}^{1}_{0}n$
Identify the 5 question marks.

Controlled release of nuclear energy in chain reactions

SYLLABUS CONTENT

- ▶ The role of chain reactions in nuclear fission reactions.
- ▶ The role of control rods, moderators, heat exchangers and shielding in a nuclear power plant.

We have seen that a neutron is usually needed to induce nuclear fission, but then that fission produces further neutrons. These neutrons may be able to induce further fission in other nuclei, so that there is the possibility of continued fissions and the continual release of energy, as is needed in a nuclear power plant (station). (A relatively small number of neutrons are emitted randomly in any sample of uranium-235.)

◆ **Chain reaction (nuclear)** Self-sustaining nuclear fission because each fission causes further fission.

If, on average, the neutrons produced in each nuclear fission produce at least one further fission, the process continues, and it is known as a **chain reaction**.

After each U-235 fission, two or three neutrons are created and then they may cause further fission, but, as we will explain, sustained fission is impossible unless we arrange for suitable circumstances, as described below. Figure E4.3 shows an example of a chain reaction involving uranium-235, in which a variety of different fission fragments can be seen.

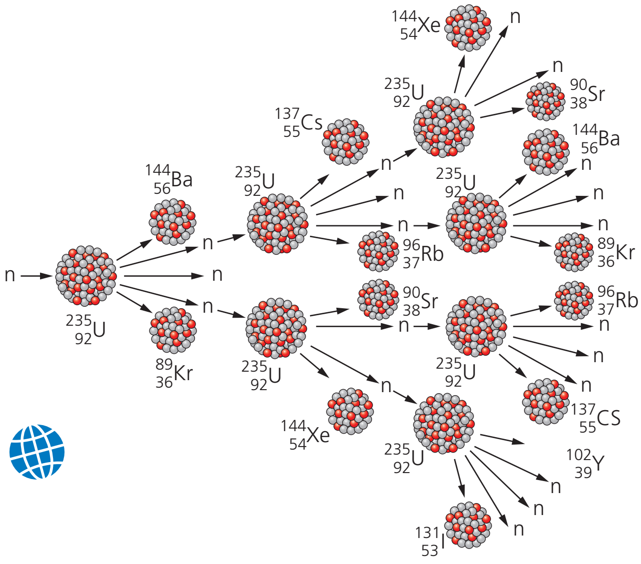


Figure E4.3 The principle of a chain reaction

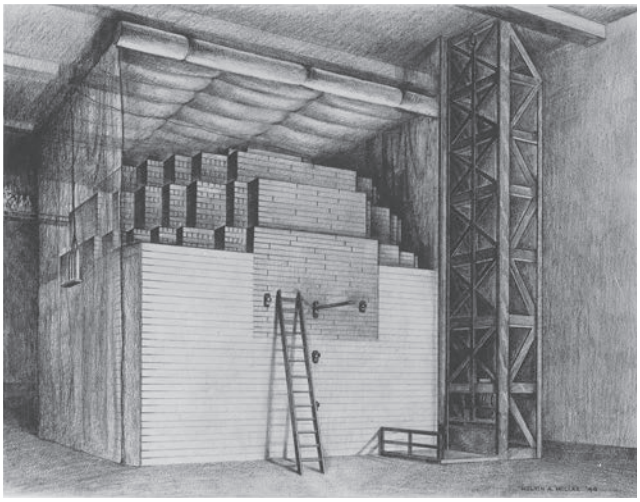


Figure E4.4 First self-sustaining nuclear chain reaction

Figure E4.5 McClean Lake uranium mine in Saskatchewan, Canada

Chain reactions will *not* normally occur. To understand why not, we need to consider the following inter-related factors.

- Uranium-235 atoms will only be a relatively small percentage of all the uranium atoms in the material being used.
- Neutrons are penetrating particles and many will simply pass out of the material without interacting with any nuclei.
- In order to cause fission, the neutrons need to be travelling relatively slowly.
- Because neutrons are uncharged, some of them will be absorbed into uranium-235 and (especially) uranium-238 nuclei without causing fission. We say they are ‘captured’. (The newly formed nuclides will probably emit gamma rays.)

It is technically *very* difficult to produce the necessary conditions for sustained and controlled chain reactions, and the continual release of energy. The following sections explain how it is done. It was famously first achieved in 1942 at the University of Chicago. See Figure E4.4.

Fuel enrichment

The metal uranium occurs throughout the Earth’s crust and is said to be a lot more common than gold (for example). The rocks from a uranium mine (see Figure E4.5) typically contain less than 0.1% uranium. The isotopes in uranium ore that is extracted from the ground are approximately in the ratio 99.3% uranium-238 and 0.7% uranium-235 (with traces of other isotopes). All the isotopes of uranium are radioactive, but the half-life of uranium-238 is very long (4.5×10^9 years), similar to the age of the Earth, whereas uranium-235 has a half-life of 7.0×10^8 years.

After the ore has been purified its appearance is as shown in Figure E4.6. It is often known as ‘yellowcake’. The isotope proportions are the same as before.

Figure E4.6 Yellowcake

For a chain reaction and power generation, the percentage of uranium-235 has to be increased to around 3% to 5%, although higher percentages are needed for specialized reactors. (Nuclear weapons require a *much* higher percentage.) This process is called **fuel enrichment**. Figure E4.7 shows a photograph of unused enriched uranium in the form of *fuel rods*. Remember that naturally occurring radionuclides of solid uranium have very long half-lives and are not usually a significant health hazard if handled correctly.

◆ **Fuel enrichment** Increasing the percentage of 235-U in uranium fuel in order to make it of use in a nuclear power station or for a nuclear weapon.

Figure E4.7 Uranium fuel rods

Uranium-238 nuclei can absorb / capture neutrons without causing fission, so too much uranium-238 will also discourage a chain reaction. Enrichment cannot be done chemically because isotopes of the same element have identical chemical properties, so physical processes need to be involved (for example, using the diffusion of gaseous uranium hexafluoride) but these are difficult and expensive technologies. The remaining uranium is called *depleted* uranium; it has physical properties, especially its high density, which have made it useful in military engineering but this has been controversial (because it is slightly radioactive).

Critical mass

The ratio of volume to surface area of a solid increases as it gets larger (consider solid cubes of different sizes). This means that the more massive a material is, the smaller the percentage of neutrons that will reach the surface and escape. That is, a higher percentage may cause fission. The **critical mass** of a material is the minimum mass needed for a self-sustaining chain reaction. (Uranium that contains 20% of uranium-235 has a critical mass of over 400 kg, which is equivalent to a sphere of radius 17 cm.)

The critical mass can be reduced by surrounding the material with neutron reflectors.

◆ **Critical mass** The minimum mass needed for a self-sustaining nuclear chain reaction.

◆ **Moderator** Material used in a nuclear reactor to slow down neutrons to low energies and enable nuclear fission.

Moderator

The neutrons released in nuclear fission have typical energies of more than 1 MeV, which means that they travel very fast. This is usually too fast to initiate another fission reaction. The slower a neutron travels, the higher the probability it has of causing fission. Therefore, before a chain reaction can occur the neutrons need to be slowed down to energies of less than 1 eV. They are often then described as *thermal neutrons*, meaning that they have average kinetic energies similar to that of the surrounding particles at the same temperature.

Reducing the speed of neutrons is called *moderating* and the material used is called a **moderator**.

In order for the fast neutrons to lose so much of their kinetic energy, they need to collide many times with the nuclei of atoms. In general, when particles collide there is a higher rate of transfer of kinetic energy between them if they have approximately the same mass (see Topic A.2). The mass of a neutron is always lighter than the mass of a whole nucleus, but the difference is less for nuclei with low mass. This is why atoms with nuclei of small mass are preferable for this process of moderation, but it is also important that the nuclei do not absorb neutrons. Commonly, the hydrogen atoms in water molecules, or graphite (carbon) is used as a moderator.

Essential features of a nuclear reactor

Figure E4.8 shows the essential features of one type of nuclear reactor. In this type of reactor, water is the moderator, but the same water is used to transfer thermal energy from the reactor vessel to another, separate, water system, using a **heat exchanger**. The water that cools the fuel rods is in a tightly sealed system and none of it leaves the concrete containment.

◆ **Heat exchanger** Apparatus designed to efficiently transfer energy from one system to another.

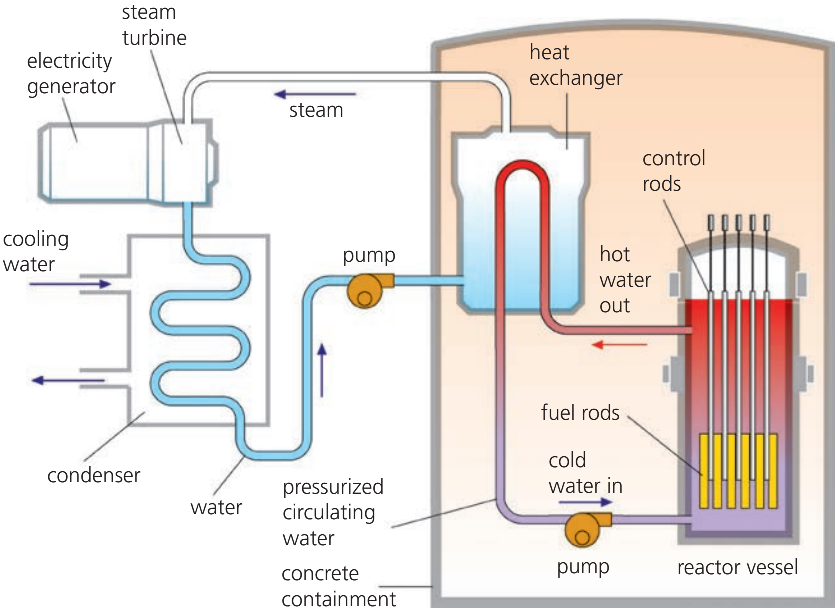


Figure E4.8 A pressurized water-cooled reactor

Steam is generated by the thermal energy and this is used to drive a turbine and generate electricity, as shown in the energy flow diagram of Figure E4.9.

A heat exchanger transfers thermal energy from the reactor to the water and steam used in the separate turbine system.

Controlling the rate of fission

In a controlled, sustained chain reaction, on average each fission will result in one further fission.

Control rods are used for adjusting the rate of the fission reactions by absorbing neutrons.

This is done by moving the rods up and down, into or out of the system as necessary. The control rods are made of a material, boron for example, which is excellent at absorbing / capturing neutrons.

As a safety measure (in the case of an electricity supply failure, for example) the control rods will fall under gravity all the way into the core and quickly shut down the reactor.

Shielding

The contents of the nuclear reactor emit large quantities of alpha, beta and gamma rays, as well as neutrons. These could be significant health hazards and people in the surroundings need to be protected:

- during normal operation
- in the event of an accident.

Shielding is used to protect people from the dangers of nuclear radiation.

The air-tight containment building is typically made of very thick concrete (with steel reinforcements) to absorb nuclear radiations and prevent any high-pressure leaks if a fault occurred in the system inside. The building should also be strong enough to withstand impact from aircraft, bombs or earthquakes.

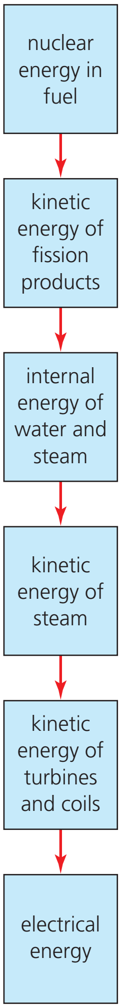


Figure E4.9 Energy flow in a nuclear power station

LINKING QUESTION

- How is binding energy used to determine the rate of energy production in a nuclear power plant?

This question links to understandings in Topic E.3.

◆ **Control rods** Used for adjusting the rate of fission reactions in nuclear reactors by absorbing more, or fewer, neutrons.

◆ **Shielding** Protective barrier around a nuclear reactor designed to absorb and reflect dangerous radiations.

Figure E4.10 Fukushima nuclear reactor after explosions in the wake of the tsunami (2011)

Safety issues

It is fair to say that the risks of nuclear power are very well understood and, as a result, safety standards are very high. But, for many people, this is not reassuring enough, because no matter how careful nuclear engineers are, accidents and natural disasters can happen. Safety standards do vary from country to country, but the consequences of a nuclear accident anywhere could be really disastrous.

The world's worst nuclear accidents, at Chernobyl in Ukraine in 1986 and Fukushima (see Figure E4.10) in Japan in March 2011, remain as vivid warnings about the possible risks of nuclear power.

Since the Chenobyl disaster, modern designs and safety procedures have been vastly improved but may still be insufficient in the face of an extreme natural disaster. The radiation leaks and explosions at Fukushima followed the damage caused by a tsunami. Hundreds of thousands of people had to be evacuated because of these incidents, and the number of long-term illnesses and deaths caused by them will take many years to be confirmed. A thermal *meltdown* is probably the most serious possible consequence of a nuclear accident. If, for some reason (for example, the loss of coolant or the control rods failing to work properly), the core of the reactor gets too hot or even melts, the reactor vessel may be badly damaged. Fires and explosions may happen as extremely hot materials are suddenly exposed to the air. Highly concentrated and dangerous radioactive materials may then be released into the ground, water or air so that they are spread over large distances by geographic and weather conditions.

◆ **Nuclear waste** Radioactive materials associated with the production of nuclear power that are no longer useful, and which may have to be stored safely for a long period of time.

When considering the dangers of nuclear power, it should always be remembered that the uses of other kinds of energy resources also have their various risks. In particular, coal mining has been responsible for an extremely high number of serious injuries, long-term health problems and deaths over the past 200 years.

Nature of science: Science as a shared endeavour

Nuclear power plants are operating in 32 different countries, producing about 10% of the world's electricity. It is believed that nine countries have nuclear weapons. The aim of the International Atomic Energy Agency (IAEA), based in Vienna, Austria, is 'to promote the safe, secure and peaceful use of nuclear technologies'. IAEA is reported to have more than 2000 multi disciplinary professional and support staff from over 100 different countries.

Waste materials from nuclear reactors

SYLLABUS CONTENT

- Properties of the products of nuclear fission and their management.

After fission has occurred, the various fission fragments are radioactive, and maybe the nuclides in any decay series from them are too. These 'waste' radionuclides can have a wide range of different concentrations and half-lives, some very long. This makes the safe storage of **nuclear waste** for the foreseeable future very important. The first step is usually storage on site in water ponds to allow time for the initial decrease in radioactivity and the dissipation of thermal energy. See Figure E4.11.

Figure E4.11 Storage of nuclear waste

High level nuclear waste is often secured underground in strong and secure containers.

Some countries prefer to store hazardous waste securely in shielded containers on the same site as the reactor.

It is also possible to recycle nuclear fuel after the percentage of fissile material decreases, but these processes are not included in this course.

5 Calculate the volume to surface area ratios for solid cubes with sides of 10 cm, 20 cm and *x* cm.

6 The critical mass of a pure uranium-235 sphere is reported to be about 50 kg.
  a Explain what this means.
  b Why is the critical mass of the uranium used in a reactor larger than 50 kg?

7 Explain why isotopes of uranium are difficult to separate from each other.

8 Calculate the speed of a neutron which has kinetic energy of 1.0 MeV.

9 **a** Write down the equation (from Topic B.1) which relates the average kinetic energy of particles to the temperature.
  b Calculate the average kinetic energy of particles (J) at 300 °C (the approximate temperature inside a nuclear reactor).
  c Determine a typical value for the speed of thermal neutrons at this temperature.

10 Explain why a heat exchanger is needed in a nuclear reactor.

11 Represent the energy transfers seen in Figure E4.9 in a Sankey diagram by making rough estimates for the efficiency of each transfer.

12 Outline the differences between the purposes of control rods and moderators in a nuclear reactor.

13 Discuss whether it is reasonable to claim that the longer the half-life of a radionuclide, the less dangerous it is.

14 **a** Use the internet to determine which countries of the world have the greatest percentage of their electricity generated by nuclear fission.
  b Suggest possible reasons for the popularity of nuclear power in those countries.

15 Radon, $^{222}_{86}\mathrm{Rn}$, is gas produced naturally from $^{226}_{88}\mathrm{Ra}$ in the uranium decay series.
  a Write an equation for this decay.
  b Explain why radon is a health hazard of particular concern, including for workers in uranium mines.

16 Discuss the advantages and disadvantages of storing nuclear waste **i** deep underground and **ii** on the site of the nuclear power plant.

Energy density of nuclear fuels

In Topic A.3 we met the concept of *energy density* of fuel sources (energy from unit volume, $\mathrm{J\,m^{-3}}$).

It is one of the major advantages of nuclear power that the fuels have exceptionally high-energy density, as in shown by Worked example E4.2.

Common mistake

Energy density (energy transferred/volume) is often confused with *specific energy* (energy transferred/mass). Both concepts are in common use.

WORKED EXAMPLE E4.2

We have seen that a typical fission of one uranium-235 nucleus releases about 200 MeV of energy.

a Determine the total amount of energy (J) that could theoretically be released from 1.00 kg of pure uranium-235.

b The density of uranium-235 is very high: 19.1 g cm^{-3}. Show that your answer to part **a** corresponds to an energy density of about 10^{18} J m^{-3}.

c The energy available from natural gas is 54.0 MJ kg^{-1}. Compare this to your answer to part **a**.

Answer

a 235 g of uranium-235 contains 6.02×10^{23} atoms (one mole)

1000 g of uranium-235 contains $\left(\frac{1000}{235}\right) \times (6.02 \times 10^{23}) = 2.5617 \times 10^{24}$ atoms

2.5617×10^{24} atoms × 200 MeV per atom $= 5.1234 \times 10^{26}$ MeV

$(5.1234 \times 10^{26}) \times (10^{6} \times 1.60 \times 10^{-19}) = 8.20 \times 10^{13}$ J

($8.1974... \times 10^{13}$ seen on calculator display)

b $(8.1974 \times 10^{13}) \times (19.1 \times 10^{3}) = 1.57 \times 10^{18}$ J m^{-3}

c $\frac{8.1974 \times 10^{13}}{54.0 \times 10^{6}} = 1.52 \times 10^{6}$

The energy per kilogramme available from uranium-235 is more than a million times greater than from natural gas.

17 Using data from Worked example E4.2:

a Calculate the mass of uranium-235 that would be used every year in a nuclear power station which has an output power of 0.85 GW and operates at an overall efficiency of 33%.

b Compare your answer to part **a** with a natural gas power station operating at the same output power, but with an overall efficiency of 52%.

18 Using data from Question 17a, determine how many fissions of U-235 are occurring in every kg of U-235 every second.

19 a If an individual uses electrical energy at an average rate of 1 kW, predict their annual energy consumption.

b Calculate the mass of uranium-235 atoms that has to undergo fission to provide the energy needed for a year. (Assume 35% efficiency.)

Advantages and disadvantages of nuclear power

Table E4.1 Some advantages and disadvantages of nuclear power

Advantages	Disadvantages
extremely high-energy density	dangerous and very long-lasting radioactive waste products
no greenhouse gases emitted during routine operation (some scientists think that nuclear power may be the only realistic solution to global warming)	expensive
no chemical pollution during operation	efficiency is not high when the whole process is taken into account
reasonably large amount of nuclear fuels are still available	threat of serious accidents
despite a few serious incidents, statistically over the last 50 years, nuclear power has overall proven to be a reasonably safe energy technology	possible target for terrorists
	linked with nuclear weapons
	not a renewable source

LINKING QUESTION

- To what extent is there a role for fission in addressing climate change? (NOS)

This question links to understandings in Topic B.2.

ATL E4A: Communication skills

Practice active listening skills

Debating

In groups of three or four, prepare arguments in advance either in favour of, or against, the following statement.

'Nuclear power has an important role to play in providing energy for electricity generation for the foreseeable future of planet Earth.'

Then have a thirty-minute debate, with students from other subjects invited to attend.
At the end take a vote.

TOK

The natural sciences

- Should scientific research be subject to ethical constraints or is the pursuit of all scientific knowledge intrinsically worthwhile?

Many (most?) people probably think that the world would be a better place without nuclear power plants and nuclear weapons. However, even if they were all dismantled, the knowledge needed to construct them would still exist. Fission cannot be 'undiscovered'.

Are there any scientific discoveries, or technological advances, that you wish had never happened?

Can it ever be possible to stop the ever-expanding scientific knowledge, especially into areas which may (later) be considered undesirable? Would we want to?

E.5

Fusion and stars

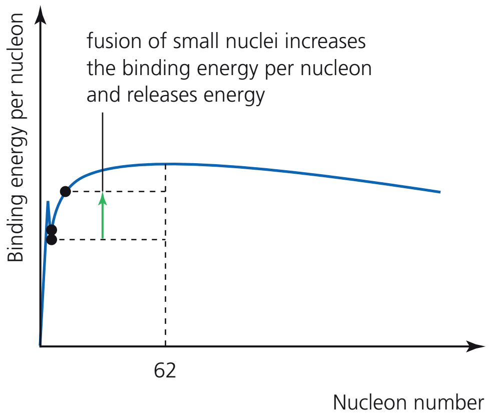


Figure E5.1 Fusion increases average binding energy per nucleon

Guiding questions

- How are elements created?
- What physical processes lead to the evolution of stars?
- Can observations of the present state of the Universe predict the future outcome of the Universe?

Nuclear fusion

Nuclear fusion was introduced briefly in Topic E.3, we will now look at it in more detail. Nuclear fusion is the combination of two small nuclei to produce a more massive single nucleus.

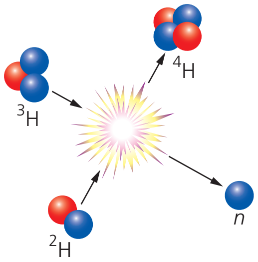


Figure E5.2 Fusion of two hydrogen nuclei

Nuclear fusion involves changing to a more stable system, with an increase in binding energy and the release of energy, mostly in the form of kinetic energy.

Consider Figure E5.1, which is similar to E3.18. In principle, the creation of a new nucleus by fusion may be possible for nuclei that have neutron numbers less than approximately 62.

An example of fusion is between a hydrogen-2 nucleus and a hydrogen-3 nucleus. (Hydrogen-2 is known as deuterium, hydrogen-3 is known as tritium.) See Figure E5.2.

$$^2_1\mathrm{H} + ^3_1\mathrm{H} \Rightarrow ^4_2\mathrm{He} + ^1_0n + \text{energy}$$

We can use changes in binding energy to determine a value for the energy released in this nuclear reaction, as shown in Worked example E5.1.

WORKED EXAMPLE E5.1

The binding energy per nucleon of hydrogen-2 is 1.11 MeV. The binding energy per nucleon of hydrogen-3 is 2.83 MeV. The binding energy per nucleon of helium-4 is much higher: 7.07 MeV. Calculate:

a the energy released in the fusion reaction shown in the equation above

b the change in total mass of the nucleons.

Answer

a Binding energy of hydrogen-2 nucleus = $2 \times 1.11 = 2.22$ MeV

Binding energy of hydrogen-3 nucleus = $3 \times 2.83 = 8.49$ MeV

Binding energy of one helium-4 nucleus = $4 \times 7.07 = 28.28$ MeV

The neutron has zero binding energy because it is a single particle.

Difference in binding energies = energy released = $28.28 - 2.22 - 8.49 = 17.57$ MeV.

b The mass of the system will reduce by as much as the energy that was released:

$$\frac{17.57}{931.5} = 0.019\ \mathrm{u}$$

ATL E5A : Research skills

Evaluating information sources for accuracy, bias, credibility and relevance

Use the internet to find out the latest developments in nuclear fusion research. Be sure to evaluate the sources you refer to – do they have a particular standpoint on nuclear energy research?

For nuclear fusion to occur the nuclei must initially have sufficient kinetic energy to overcome the repulsive forces between the positive charges. This requires extremely high temperatures, more than 10^7 K.

If the nuclei can overcome the electric repulsion between positive charges, they can get very close to each other, and then the attractive nuclear forces pull them together and fusion may occur.

The principles are well understood and *if* nuclear fusion could be sustained, it would release enormous amounts of energy for electricity generation from plentiful raw materials, without significantly contributing to climate change, and without producing large quantities of radioactive waste (as occurs with nuclear fission). The prospect of plentiful energy from nuclear fusion has pre-occupied and excited scientists for much more than fifty years, but the technical problems seem as large as ever.

LINKING QUESTION

- How is fusion like – and unlike – fission?

This question links to understandings in Topic E.4.

1 Calculate the repulsive electric force between two protons that are 1.0×10^{-15} m apart.

2 Use data from Worked example E5.1.
   - **a** Determine how much energy (J) would be released from the fusion of one kilogramme of helium-4.
   - **b** How would you describe the energy density of this process?

3 Calculate the kinetic energy (eV) of a proton (hydrogen nucleus) at a temperature of 1.0×10^7 K.

4 The following is the simplest possible example of fusion. It occurs in stars.

$${}^{1}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{2}_{1}\mathrm{H} + {}^{0}_{1}e^{+} + {}^{0}_{0}\nu$$

   - **a** Using data (with a suitable number of significant figures) from the internet, determine:
      - **i** the change in mass that occurs in this fusion
      - **ii** the energy released (MeV) in this reaction.
   - **b** Explain why the fusion of each helium-4 nucleus releases much more energy than the fusion of hydrogen-2.

5 Explain why nuclear fusion is only possible with two nuclei which have relatively small numbers of nucleons.

Although nuclear fusion is a rare event here on Earth, we will now turn our attention to where it dominates: the rest of the Universe.

◆ **Interstellar matter** Matter that exists in the space between stars (usually at very low density).

◆ **Nebula** (plural: **nebulae**) Identifiable, diffuse 'cloud' of interstellar matter; mainly gases (mostly hydrogen and helium) and dust.

Formation of stars

SYLLABUS CONTENT

- ▶ The stability of stars relies on an equilibrium between outward radiation pressure and inward gravitational forces.
- ▶ Fusion is a source of energy in stars.
- ▶ The conditions leading to fusion in stars in terms of density and temperature.

■ **Figure E5.3** The Orion nebula

The space between stars has evolved over billions of years to contain very low concentrations of particles, which are collectively known as **interstellar matter**, and commonly described as 'dust and gas'. Hydrogen is, on average, about 70% of all interstellar matter (by mass), helium has 28% and the remaining 2% is other elements (mainly remnants of exploded older stars). Depending on the circumstances, the gas particles may be molecules, atoms or ions. A **nebula** (Figure E5.3) is the name given to a distinct and identifiable giant 'cloud' of dust and gas in space.

In places where there is an increased density of material (for whatever reason), gravitational forces will (very slowly) pull the particles closer together. This results in increasing kinetic energies of particles, which is equivalent to increasing temperature.

◆ **Main sequence stars** Stable stars which are fusing hydrogen into helium in their cores.

Eventually, the temperature and density of hydrogen will be great enough for nuclear fusion of hydrogen into helium to occur. Temperatures of at least 10^7 K are needed. At this temperature the hydrogen nuclei (protons) have enough kinetic energy to overcome the Coulomb repulsion between them. This is the dominant energy transfer occurring in all **main sequence stars**. ('Main sequence' stars are explained below. Most stars are main sequence stars.) There are two principal ways in which fusion can happen in main sequence stars, as explained below. At these temperatures, hydrogen exists simply as protons.

Only a very small fraction of interactions between protons results in fusion. The closer the protons are to each other (on average), the greater the frequency of interactions and the rate of fusion. In other words, a high density is needed to sustain nuclear fusion in stars.

Nuclear fusion at high temperatures and densities in main sequence stars combines four protons to form helium-4.

Proton–proton cycle

◆ **Proton–proton cycle** The simplest nuclear fusion process which converts hydrogen into helium, releasing large amounts of energy in medium and smaller-sized main sequence stars.

The **proton–proton cycle** is also called the proton–proton chain.

This is the principal nuclear fusion process in main sequence stars which have a *mass similar to the Sun, or less.*

This is a 3-step process, which can be summarized as follows but the details do not need to be remembered:

$$4\,{}^{1}_{1}\mathrm{H} \rightarrow {}^{4}_{2}\mathrm{He} + 2\,{}^{0}_{1}e + \text{neutrinos} + \text{photons}$$

The three separate reactions are:

1 Two protons fuse to make a ${}^{2}_{1}\mathrm{H}$ (deuterium) nucleus. In this process, one of the protons converts into a neutron in a beta-plus decay, also forming a positron and a neutrino.

$$ {}^{1}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{He} \rightarrow {}^{2}_{1}\mathrm{H} + {}^{0}_{1}e^{+} + \nu$$

2 The deuterium nucleus fuses with another proton to make helium-3. In this process, a gamma ray photon is also emitted.

$$ {}^{2}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{He} \rightarrow {}^{3}_{2}\mathrm{He} + \gamma$$

3 Two helium-3 nuclei combine to make helium-4. Two protons are released in this reaction.

$$ {}^{3}_{2}\mathrm{He} + {}^{3}_{2}\mathrm{He} \rightarrow {}^{4}_{2}\mathrm{He} + 2\,{}^{1}_{1}\mathrm{H}$$

These three stages are illustrated in Figure E5.4.

The energy released in each 3-stage cycle is 27 MeV. This is transferred to the kinetic energy and electromagnetic energy of the products.

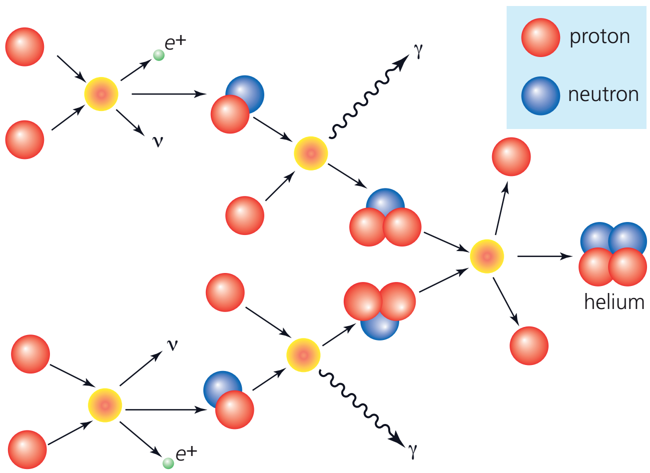


Figure E5.4 The proton–proton cycle

CNO cycle

This is the principal nuclear fusion process in main sequence stars *which have a mass significantly greater than the Sun.* This is because of the higher temperatures in the cores of the more massive stars.

CNO represents carbon, nitrogen, oxygen. These elements must be present in the core, but they are not changed by the process. (They act as *catalysts*.)

The **CNO cycle** is a complicated 6-step process (no details are needed), but in effect it still involves the fusion of four protons to one helium nucleus (similar to the proton–proton cycle) and releases about the same amount of energy (27 MeV).

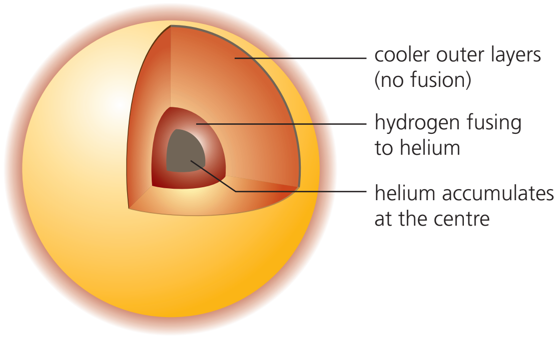


Figure E5.5 Internal structure of a main sequence star

Stellar equilibrium

Energy is transferred from the *core* of the star, where it is hottest and where the fusion is happening, by radiation, conduction and convection to the surface. See Figure E5.5. Helium will accumulate at the centre of the core because of its greater density. From the surface, enormous amounts of energy are radiated into space in all directions. These fusion processes will continue as long as there is enough hydrogen in the core, at a high enough temperature.

Main sequence stars can remain in equilibrium for millions, or billions of years, under the following condition (see also Figure E5.6):

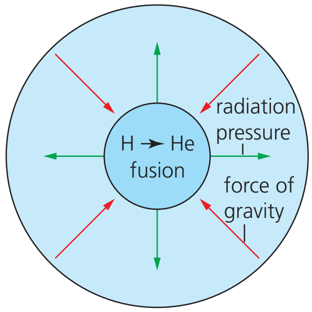


Figure E5.6 Equilibrium in a main sequence star

In **stellar equilibrium**, the outward radiation (thermal) pressure is balanced by inwards gravitational forces.

This is a self-correcting (feedback) process: if the fusion rate was to increase slightly, the temperature would increase so that the outwards pressure rises. This would result in an expansion of the star, so that density and temperature decrease, and the fusion rate falls.

Nature of science: Patterns and trends

Feedback processes

Feedback is the name used to describe any response within an on-going system which affects the future behaviour of the same system. For example, if you are told that you are doing well in your physics studies, you may be encouraged, and then work harder. This would be called *positive* feedback.

Negative feedback can often help to stabilize a system, as in the example of stellar equilibrium.

Tool 3: Mathematics

Use of units whenever appropriate

In calculations related to the properties of stars, it becomes convenient to compare a star to our Sun: $M_\odot$ = mass of Sun, $L_\odot$ = luminosity of Sun, $R_\odot$ = radius of Sun.

◆ **CNO cycle** Nuclear fusion process in larger main sequence stars which forms helium from hydrogen.

◆ **Stellar equilibrium** Main sequence stars are in equilibrium under the balanced effects of radiation pressure acting outwards against gravitational forces acting inwards.

◆ **Feedback** Occurs when a response within a system influences the continuing behaviour of the same system.

Lifetimes of main sequence stars

◆ **Lifetime (of main sequence star)** The duration for which a star is fusing hydrogen into helium, emitting radiation and maintaining stellar equilibrium.

There will come a time when most of the hydrogen in the *core* of a star becomes depleted (already used, in fusion), so that the rate of fusion decreases to such an extent that it is no longer possible to maintain the equilibrium described above. This will be the end of the **lifetime** of the main sequence star and it will then change in ways which are described later in this topic. There will still be a significant amount of hydrogen *outside* of the core.

Main sequence stars will come to the end of their lifetimes when most of the hydrogen in their cores has been converted to helium.

It might be expected that more massive stars would have longer lifetimes because they contain more hydrogen, but the opposite is true.

The rate of fusion in more massive (hotter) stars is so much greater, that they have significantly *shorter* lifetimes.

This can be seen in Table E5.1 in the next section.

We can make a *rough estimate* for the future lifetime of a main sequence star as follows (using data for the Sun as an example). *This lengthy and detailed calculation need not be remembered.*

mass, $M_{\odot} = 1.99 \times 10^{30}\,\text{kg}$

luminosity, $L_{\odot} = 3.85 \times 10^{26}\,\text{W}$

Each proton–proton cycle releases 27 MeV (= 4.3×10^{-12} J) of energy.

We will assume that, when they are first formed, main sequence stars consist of *approximately* 70% hydrogen, and that the Sun will end its main sequence lifetime when about 15% of its total hydrogen remains (mostly in the outer layers).

We can calculate an approximate value for its main sequence lifetime as follows:

amount of hydrogen that will be fused ('burned') during the main sequence lifetime = (100 − 15)% of 70 % of $1.99 \times 10^{30}\,\text{kg} = 1.18 \times 10^{30}\,\text{kg}$

mass of hydrogen involved with each proton–proton cycle = $4 \times (1.673 \times 10^{-27}) = 6.688 \times 10^{-27}\,\text{kg}$

number of proton–proton cycles during the main sequence lifetime $= \dfrac{1.18 \times 10^{30}}{6.688 \times 10^{-27}} = 1.76 \times 10^{56}$

rate of proton–proton cycles $= \dfrac{3.85 \times 10^{26}}{4.3 \times 10^{-12}} = 8.95 \times 10^{37}\,\text{s}^{-1}$ (average for the main sequence lifetime)

main sequence lifetime of our Sun $= \dfrac{1.77 \times 10^{36}}{8.95 \times 10^{37}} = 2.0 \times 10^{18}\,\text{s}$ ($\approx 6 \times 10^{10}$ years)

This is just an approximation, but it is in broad agreement with the accepted value for the future lifetime of the Sun ($\approx 10^{10}$ years).

We can also estimate the decrease in the mass of the Sun due to nuclear fusion reactions from $E = mc^2$ (Topic E.3):

energy transferred during main sequence lifetime,
$E = \text{power} \times \text{time} = (3.85 \times 10^{26}) \times (2.0 \times 10^{18}) = 7.7 \times 10^{44}\,\text{J}$

$E = mc^2$

$7.7 \times 10^{44} = m \times (3.0 \times 10^{8})^2$

$m = 8.6 \times 10^{27}\,\text{kg}$

This is equivalent to about four billion kilogrammes every second!

TOK

Knowledge and the knower

Too big or too small to comprehend?

More than any other area of knowledge, physics involves an appreciation of numerical values over enormous ranges: from subatomic particles to the size of the observable Universe.

However, the larger a number, the worse we are at really understanding what it represents. Similarly, very small numbers are difficult to comprehend, for example 10^{-15} m: the approximate radius of a nucleus. Clearly, 'four billion kilogrammes' (from last section), is a large number, but to begin to appreciate just how large, comparisons are usually helpful. We might say that 4×10^9 kg is about equal to the total mass carried by every person on Earth if they each had 0.5 kg. Or 4×10^9 kg is approximately equal to the mass of water in a reservoir of dimensions 1 km × 1 km × 4 m. However, such comparisons are less useful when much larger numbers are involved.

Suggest a comparison you could use to make 10^{-10} m (an approximate size of an atom) understandable for a 10 year-old child.

6 Discuss whether you would expect that most hydrogen in interstellar matter was in the form of ions, atoms or molecules.

7 Calculate the acceleration due to gravity of two protons separated by one metre.

8 Determine a value for the energy released (MeV) in the third stage of the proton–proton cycle, as described above. (Use the internet to determine relevant data.)

9 The CNO cycle needs carbon, nitrogen and oxygen in the core. Suggest where these elements have come from.

10 Explain what will happen if, for some reason, the fusion rate in a main sequence star was to decrease slightly. Assume that there is plenty of hydrogen available in the core.

11 a Using the same method and assumptions as shown in the calculation of the lifetime of the Sun (above), determine the lifetime of a main sequence star which has a mass ten times greater than the Sun's and a luminosity three thousand times greater than that of the Sun.

b The relationship between the luminosity and mass of a main sequence star is:

$$L \approx L_{\odot}\left(\frac{M}{M_{\odot}}\right)^{3.5}$$

(You do not need to remember this.)

Show that this relationship confirms the approximate luminosity of the more massive star given in part **a**.

Comparing main sequence stars

SYLLABUS CONTENT

- The main regions of the Hertzsprung–Russell (HR) diagram and how to describe the main properties of stars in that region.
- How to determine stellar radii.

When we observe stars from Earth, they all appear as point sources of light with different *apparent brightnesses* and slightly different colours. Direct observation provides no more information, except their positions relative to each other, which enables star maps to be drawn. Although stars move at high speeds, the distances between them are so great that no changes in these positions are apparent to observers on Earth, even over hundreds of years. (But note that some stars, which are relatively close to the Earth, show *small* repeated movements (on a star map) during the course of every year. This is called *stellar parallax*, and it is explained later.)

The same nuclear fusion processes are occurring in all main sequence stars, but astronomers have calculated that these stars have different masses, radii and temperatures (core and surface). As a rough guide, our Sun may be considered to be an 'average' main sequence star, other stars have radii which are up to 1000 × greater or smaller than the Sun's, with an even wider range of luminosities: from 10 000 × less to 1 000 000 × greater.

The differences between main sequence stars are not random. There is a very clear pattern (see Table E5.1), because the rate of fusion depends on the masses of the stars.

Main sequence stars of greater mass have greater radii, greater rates of fusion, higher temperatures, greater luminosities and shorter lifetimes.

This pattern is apparent in the data shown in Table E5.1.

Table E5.1 Properties of main sequence stars (figures are approximate)

Mass/$M_\odot$	**Luminosity/$L_\odot$**	**Effective temperature/K**	**Radius/$R_\odot$**	**Lifetime/y**
0.10	3×10^3	2900	0.16	2×10^{12}
0.50	0.03	3800	0.6	2×10^{11}
1.0	1	5800	1.0	1×10^{10}
3	60	11 000	2.5	2×10^8
5	600	17 000	3.8	7×10^7
10	10 000	22 000	5.6	2×10^7
25	80 000	35 000	8.7	7×10^6
60	790 000	44 500	15	3×10^6

The 'effective temperature' is the name given to the surface temperature calculated assuming the star behaves as a perfect black body.

Clearly, there is no need to remember any of the data seen in Table E5.1, but the trends are important and are probably easier to understand as shown graphically in Figure E5.7, Figure E5.8 and Figure E5.11.

Mass has a considerable effect on luminosity. For example, if star A has twice the mass of star B it will have *approximately* ten times the luminosity. This general trend is indicated in Figure E5.7.

This pattern is shown in more detail in Figure E5.8, which is known as a **Hertzsprung–Russell (HR) diagram**.

◆ **Hertzsprung–Russell (HR) diagram** Diagram that displays order in the apparent diversity of stars by plotting the luminosity of stars against their surface temperatures.

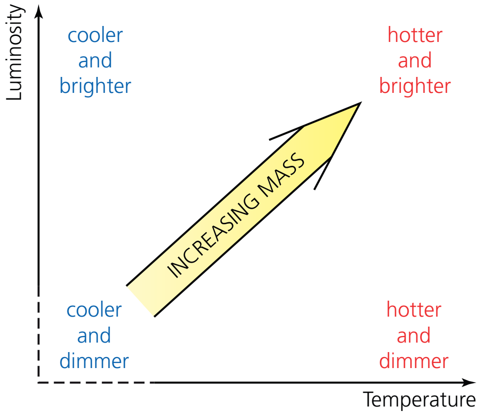


Figure E5.7 Linking mass, temperature and luminosity for main sequence stars

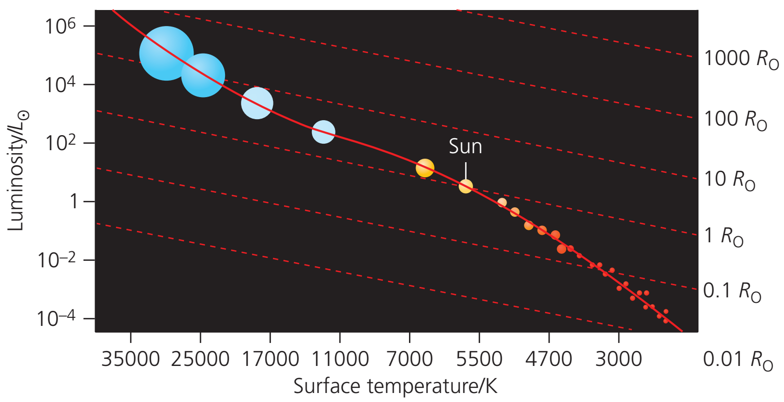


Figure E5.8 Hertzsprung–Russell diagram *(incomplete)* with colours enhanced

LINKING QUESTIONS

- How can the understanding of black-body radiation help determine the properties of stars?
- How do emission spectra provide information about observations of the cosmos?

These questions link to understandings in Topics B.1 and E.2.

Figure E5.8 is an important diagram and it should be well understood. Note, in particular:

- For historical reasons, the temperature scale is reversed, with lower temperatures on the right.
- The differences in stars' luminosities and temperatures are so great that logarithmic scales are used.
- The luminosity scale is shown relative to the Sun's luminosity, $L_\odot$ (= 3.8×10^{26} W).
- The dotted lines represent constant radius. $R_\odot$ = radius of Sun (7.0×10^{8} m).

The connection between surface temperature and colour was explained in Topic B.1.

Other types of stars (non-main sequence) will be added to the HR diagram after they have been explained.

Top tip!

Reminders from Topic B.1:

Stars can be considered to be perfect black bodies and the radius of a star can be determined from $L = \sigma A T^4$, by using surface area, $A = 4\pi r^2$. The temperature of a star's surface is sometimes called the star's effective temperature. (This is because a value for T is often calculated from knowledge of L and A, assuming that the star is acting as a perfect black body.)

The surface temperature, T, of a star can be determined from $\lambda_{\max} T = 2.9 \times 10^{-3}$ m K, where $\lambda_{\max}$ is the peak wavelength of the black-body spectrum emitted.

WORKED EXAMPLE E5.2

The large, bright star Canopus has a luminosity = 10 700 $L_\odot$ and a surface temperature of 7400 K.

Determine a value for its radius.

Answer

$L = \sigma A T^4 = \sigma(4\pi r^2)T^4$

$10\,700 \times (3.85 \times 10^{26}\ \text{W})$

$= (5.67 \times 10^{-8}) \times 4 \times \pi r^2 \times 7400^4$

$r = 4.39 \times 10^{10}$ m

WORKED EXAMPLE E5.3

The main sequence star Vega has a surface temperature of 9600 K.

a Determine the peak wavelength in its spectrum.

b Estimate the luminosity of this star.

Answer

a $\lambda_{\max} T = 2.9 \times 10^{-3}$

$$\lambda_{\max} = \frac{2.9 \times 10^{-3}}{9600} = 3.0 \times 10^{-7}\ \text{m}$$

b From Figure E5.8, $L \approx 50 L_\odot$

12 A main sequence star has a radius of approximately $10R_0$ and a luminosity of $(2 \times 10^4)L_\odot$.

a Use Figure E5.8 to estimate its surface temperature.

b State the colour you would expect this star to be.

13 Estimate the luminosity of a main sequence star which has a surface temperature of 15 000 K.

14 A main sequence star is observed to be slightly blue in colour. Suggest possible values for its luminosity, radius and surface temperature.

15 A star has a radius of $100\,R_0$ and a luminosity of $10\,000\,L_\odot$.

a Explain why you can be sure that this is not a main sequence star.

b Estimate its surface temperature.

16 The main sequence star Altair has a radius which is approximately twice the radius of the Sun and surface temperature of about 7500 K.

a Use Figure E5.8 to estimate its luminosity:

i as a multiple of $L_\odot$ **ii** in W.

b Use an equation for luminosity to determine a more accurate value for Altair's luminosity.

17 a Use Figure E5.8 to estimate how much greater the luminosity of a star of radius $100R_0$ is compared to a star of radius $10R_0$.

b Explain why the larger star has a *much* greater luminosity.

18 A main sequence star has a peak wavelength of 9.7×10^{-8} m.

a Determine its surface temperature.

b Estimate its luminosity.

Evolution of stars

SYLLABUS CONTENT

- The effect of stellar mass on the evolution of a star.
- The main regions of the Hertzsprung–Russell (HR) diagram and how to describe the main properties of stars in these regions.

◆ **Evolution (stellar)** Describes the changes that occur in a star during its 'lifetime'.

◆ **Giant (and super giant) stars** Usually relatively cool stars, so they are yellow / red in colour; their luminosity is high because of their large size. Most stars will become **red giants** (or red **super giants**) at the end of their time on the main sequence.

The term **stellar evolution** is being used here to describe what happens to stars after the depletion of hydrogen in their cores. That is, what happens to them after the end of their main sequence lifetimes. The core begins to contract because, once the rate of fusion is reduced, the inward gravitational forces are greater than the outward radiation pressure. Gravitational potential energy is then again transferred to kinetic energy of the nuclei in the core, so that the temperature rises significantly. This causes fusion of hydrogen outside of the core (in a 'shell' around the core).

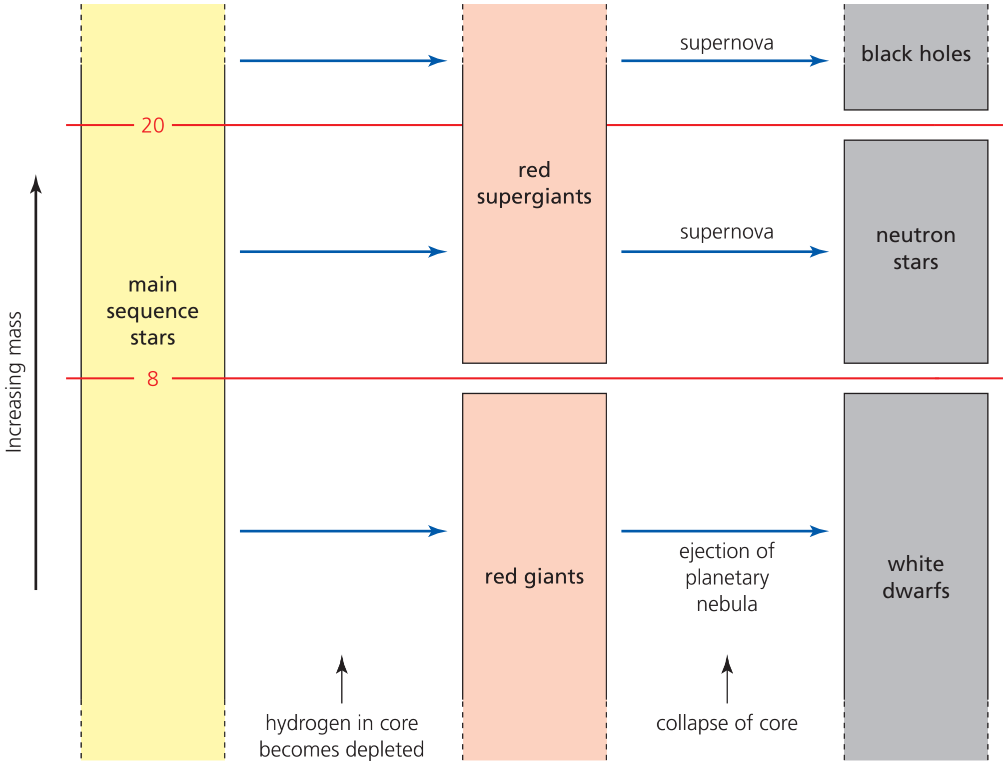
The rate of fusion in the shell is greater than in the core when it was a main sequence star. So, a star will spend more time on the main sequence than afterwards.

The resulting increased radiation pressure produces a significant expansion of the star. The radius could be as much as one thousand times greater, which would result in a one million times greater surface area. The overall effect is a *reduction* in surface temperature and therefore a change in colour to become redder. These changes are represented by the name of the type of star formed: a **red giant**. Or, if sufficiently large, a red **super giant**. Only a small percentage of main sequence stars will evolve to become red super giants.

Red giant stars (and super giants) are formed by the increased rate of nuclear fusion that occurs because of the greater temperatures created in the collapse of main sequence stars at the end of their lifetimes.

All but the very smallest main sequence stars will become red giants or super giants after the hydrogen in their core has been depleted. What happens after that depends (again) on their masses, as seen in Figure E5.9.

Figure E5.9 Evolution of stars of different masses (the numbers shown represent the approximate mass limits of the stars as multiples of the current mass of the Sun)



Common mistake

Planetary nebula is a misleading term. It has nothing to do with planets.

◆ **Planetary nebula** Material emitted from the outer layers of a red giant star at the end of its lifetime. The core then becomes a white dwarf star.

◆ **White dwarf stars** Relatively hot stars, so that they are blue / white in colour, but their luminosity is low because of their small size. They are formed after the end of the lifetime of smaller main sequence stars.

◆ **Electron degeneracy** Process occurring within white dwarf stars that keeps them stable and stops them collapsing.

◆ **Supernova** Sudden and very luminous explosion of a massive star, resulting in a neutron star or black hole.

◆ **Neutron stars** Very dense stars formed after a supernova.

◆ **Black hole** Extremely dense remnant of a giant star formed after a supernova. Gravitational forces are so great that even light cannot escape.

After all fusion has finished in a red giant, gravitational forces will cause the star to collapse inwards. The outer layers are ejected to form a relatively short-lived diffuse cloud of ionized gas called a **planetary nebula**, leaving behind a very dense core. Red super giants evolve differently, as discussed below.

The remaining core has enough internal energy to continue to emit radiation at low luminosity, for a very long time and its surface temperature is high enough that it appears white: this explains the name of this type of star: **white dwarf**. See Figure E5.10.

A white dwarf star is formed from a red giant star after all nuclear fusion has stopped.

A white dwarf star can remain stable for a long time because of a process called **electron degeneracy** (no details required).

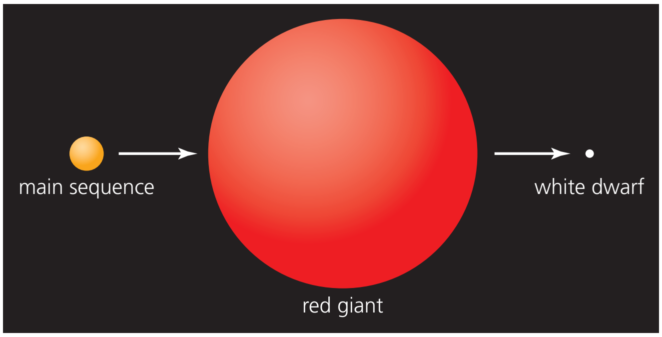


Figure E5.10 Evolution of most main sequence stars (our Sun, for example)

Red super giants do not evolve into white dwarfs. The electron degeneracy pressure is insufficient to resist the gravitational forces and the gravitational potential energy released is so high that there are dramatic changes in the core that result in an enormous explosion called a **supernova**. Here again, the result depends on the mass involved. If the original mass of the star was between 8 and 20 solar masses, the remaining core after the supernova will form a **neutron star**. If the mass was greater, a **black hole** is formed. (Further details are not required.)

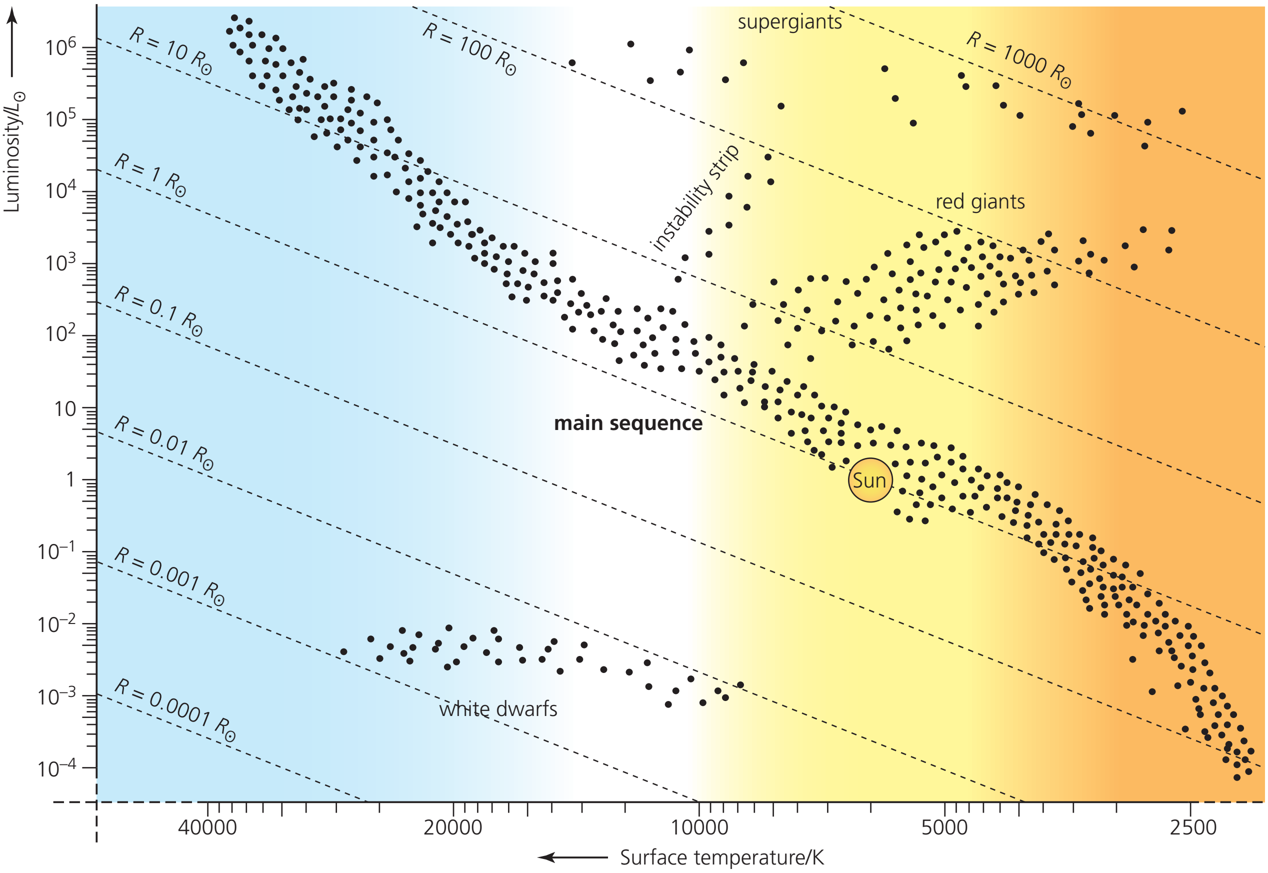
◆ **Evolutionary path** The evolution of a star as drawn on a Hertzsprung–Russell diagram.

◆ **Instability strip** A region of the Hertzsprung–Russell diagram containing pulsating, variable stars.

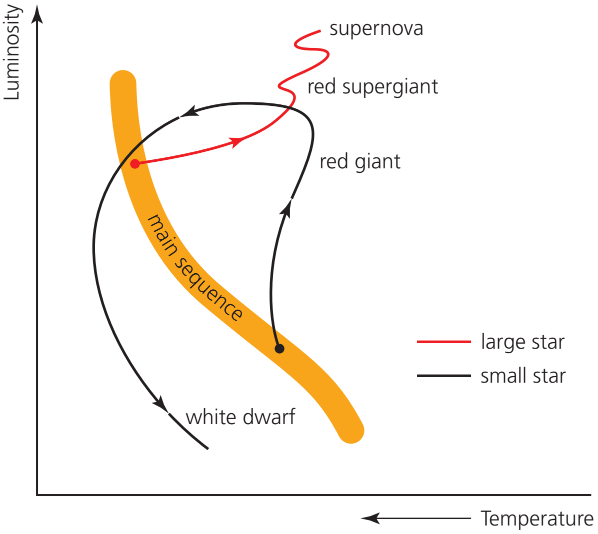
A supernova is an 'explosion' of a red super giant, creating a neutron star or a black hole.

Evolution on a HR diagram

Figure E5.11 shows a more detailed HR diagram with other types of stars included (not just main sequence stars).



■ **Figure E5.11** Detailed HR diagram



■ **Figure E5.12** Evolutionary path of stars after they leave the main sequence

When a main sequence star expands to a red giant, or a red super giant, its luminosity and surface temperature change and this, and subsequent changes, can be tracked on an HR diagram. It is known as a star's **evolutionary path**. Typical evolutionary paths of low-mass and high-mass main sequence stars are shown in Figure E5.12.

The **instability strip** on the HR diagram (seen in Figure E5.11) contains stars at an intermediate stage between being main sequence stars and being super giants or more massive red giants. These stars will be in the instability strip for relatively short times. They pulsate because their outer layers are unstable: their luminosities vary periodically.

The changes to stars which happen after the end of their main sequence lifetimes can be tracked on an HR diagram.

19 Outline the process by which a main sequence star can evolve into a red giant star.

20 Explain the difference between a red giant star and a red super giant star.

21 A red giant has a higher rate of fusion than the main sequence star from which it was formed.
Explain why it has a lower surface temperature.

22 Outline what is meant by the term ‘planetary nebula’.

23 a Explain what the colour of a white dwarf star tells us about its surface temperature.

b Explain why, despite their temperatures, white dwarfs can be difficult to observe.

24 Sketch the axes of a HR diagram and add a line to indicate main sequence stars.

a Draw the future evolutionary path of the Sun.

b Draw the evolutionary path of a massive main sequence star that eventually forms a red super giant, after spending time in the instability strip.

25 Use the internet to learn about supernovas that have been detected on Earth.

LINKING QUESTION

- HR diagrams have been helpful in the classification of stars by finding patterns in their properties. Which other areas of physics use classification to help our understanding? (NOS)

Distances from Earth to stars

SYLLABUS CONTENT

▶ The use of stellar parallax as a method to determine the distance, d, to celestial bodies as given by:

$$d(\text{parsec}) = \frac{1}{p(\text{arc-second})}$$

In Topic B.1 it was explained that, *if* we know the luminosity, L, of any star, the equation $b = \frac{L}{4\pi d^2}$ can be used to determine its distance from Earth, d, if we measure its apparent brightness b.

The problem is that, for most stars, we have no direct way of knowing their luminosities.

However, importantly, astronomers have identified a few ‘standard candles’. These are stars which have known luminosities wherever they are located (including a type of supernova and some stars in the instability strip). You are not expected to have knowledge of these methods.

The HR diagram can also be used to obtain an approximate distance to a main sequence star, as shown in the following example.

WORKED EXAMPLE E5.4

The surface temperature of a main sequence star was determined to be 17 000 K.

a Explain how this value was calculated using information from the star’s spectrum.

b Use the HR diagram to estimate the luminosity of the star (W).

c Determine an approximate distance between this star and Earth if its apparent brightness was measured on Earth to be $3.1 \times 10^{-9}\,\text{W m}^{-2}$ in:

i metres **ii** light years.

Answer

a By using Wien’s law:
$\lambda_{\text{max}} T = 2.9 \times 10^{-3}\,\text{m K}$

b $1200 L_{\odot} = 1200 \times 3.8 \times 10^{26} \approx 4.6 \times 10^{29}\,\text{W}$

c i $b = \frac{L}{4\pi d^2}$

$$3.1 \times 10^{-9} = \frac{4.6 \times 10^{29}}{4 \times \pi \times d^2}$$

$d = 3.4 \times 10^{18}\,\text{m} \approx 3 \times 10^{18}\,\text{m}$

ii In light years this is:

$$\frac{3.4 \times 10^{18}}{9.46 \times 10^{15}} \approx 4 \times 10^{2}\,\text{ly}$$

◆ **Stellar parallax** Method of determining the distance to a nearby star from measurement of its *parallax angle*.

◆ **Parallax** The displacement in the apparent position of an object (compared to its background) viewed along two different lines of sight.

Distances to 'nearby' stars

We are able to calculate the distance to 'nearby' stars using simple geometry.

Although stars are moving very quickly, we can draw *star maps* to show the positions of stars (relative to other stars) because their relative positions show no significant changes over very long periods of time. This is because of the enormous distance between the stars (and Earth). However:

if very precise measurements are made of 'nearby' stars, their positions (relative to other stars) are observed to move very slightly (forwards and backwards) during the course of a year. This is called **stellar parallax**.

This repeated apparent movement is represented in Figure 5.13, but the the movement has been exaggerated for clarity.

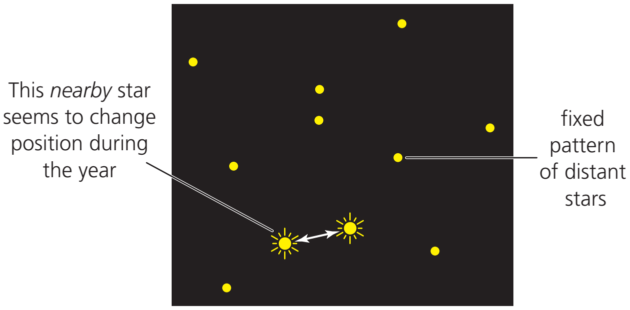


Figure E5.13 A nearby star's apparent movement due to parallax

Inquiry 3: Concluding and evaluating

Evaluating

Parallax

Parallax is the difference in the apparent position of an object (when compared to other objects in the background behind it) when an object is viewed along two different lines of sight. This is shown in Figure E5.14. The closer the object is to the observer, the greater the parallax. The simplest everyday example is seen when observing your finger held in front of your face, first with one eye, then the other.

A numerical value for parallax can be represented by half the angle between the two dotted lines in Figure E5.14.

Identify how **parallax error** can lead to misreading the scales of some measuring instruments.

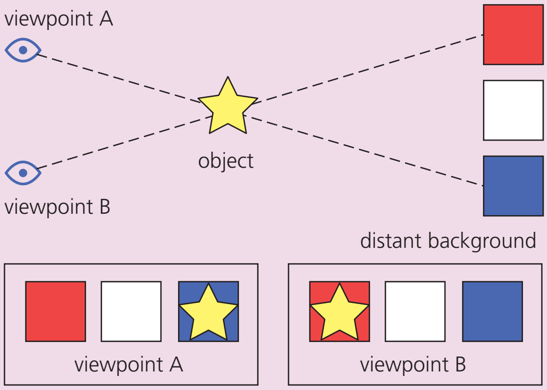


Figure E5.14 Parallax

◆ **Parallax error** Occurs when reading a scale from the wrong position.

Because of the limits to measuring very small angles accurately, stellar parallax can be only detected with stars which are less than about 300 ly from Earth. These 'nearby' stars are all within our own galaxy (The Milky Way, which has a diameter of about 10^5 ly.) and include some stars which are visible from Earth without a telescope. In Figure E5.15 and Figure E5.16 the parallax angles have been greatly exaggerated for the sake of clarity.

Using telescopes, astronomers measure the parallax angle, *p*, between, for example, observations of the star made in December and June.

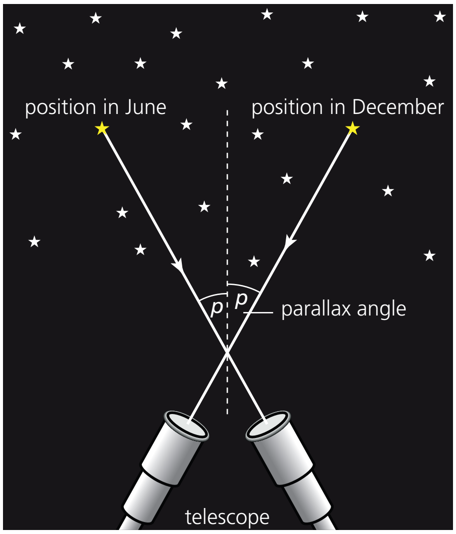


Figure E5.15 Parallax angle six months apart

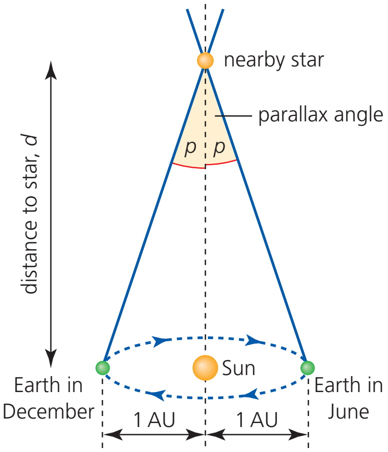


Figure E5.16 The geometry of the parallax angle

Tool 3: Mathematics

Use of units whenever appropriate

In Figure E5.16, the distance between the Earth and the Sun has been labelled as 1 AU. The Earth's orbit around the Sun is almost circular and the radius varies only by about 3%. The AU, astronomical unit is used in calculations concerning our Solar system. It is *defined* to be exactly $1.495\,978\,707\,00 \times 10^{11}$ m, but we will use 1.50×10^{11} m in calculations.

◆ **Arc-second** An angle which is 1/3600 of one degree.

The stellar parallax of even the closest stars is very small because of the long distances involved and this means that the parallax angles are so tiny that they are measured in **arc-seconds**. (There are 3600 arc-seconds in a degree.)

Once the parallax angle has been measured, simple geometry can be used to calculate the distance to the star (Figure E5.16):

$$\text{parallax angle, } p\text{ (rad)} = \frac{1.50 \times 10^{11}}{d\text{ (m)}}$$

Note that the distance from the Earth to the star and the distance from the Sun to the star can be considered to be equal for such very small angles. That is, we can assume that p (rad) = $\sin p = \tan p$.

WORKED EXAMPLE E5.5

Calculate the distance, *d,* to a star if its parallax angle is 0.240 arc-seconds. (Reminder: there are 57.3 degrees in one radian.)

Answer

$$0.240\text{ arc-seconds} = \left(\frac{0.240}{3600}\right) \times \left(\frac{1}{57.3}\right) = 1.16 \times 10^{-6}\,\text{rad}$$

$$p(\text{rad}) = \frac{1.50 \times 10^{11}}{d(\text{m})}$$

$$d = \frac{1.50 \times 10^{11}}{1.16 \times 10^{-6}} = 1.29 \times 10^{17}\,\text{m } (= 13.7\,\text{ly})$$

Tool 3: Mathematics

Use of units whenever appropriate

Calculations similar to Worked example E5.5 are common, but it is much easier to use the angle directly as a measure of distance rather than making calculations in SI units. However, this is an inverse relationship – larger parallax angles mean smaller distances.

The **parsec** (pc) – short for parallax of one arc-second – is another unit of distance used by astronomers. Its use is not restricted to stars which exhibit parallax, and it is the most widely used unit of distance in astronomy. A *parsec* is defined as the distance from the Sun (or Earth) to an object that has a parallax angle of one arc-second.

◆ **Parsec, pc** Unit of distance used by astronomers; equal to the distance to a star that has a parallax angle of one arc-second.

$$\text{distance, } \boldsymbol{d} \textbf{ (parsec)} = \frac{\mathbf{1}}{\boldsymbol{p}\textbf{(arc-second)}}$$

For example, a star with a parallax angle, *p*, of 0.25 arc-seconds will be 1/0.25 = 4 pc distant from Earth.

Table E5.2 shows the relationship between parallax angle and distance.

■ **Table E5.2**

Parallax angle/arc-seconds	Distance away/pc
0.10	10.0
0.25	4.0
0.50	2.0
1.00	1.0

Summary of the non-SI units used in astronomy

1 parsec (pc) = 3.26 ly

■ **Table E5.3** Summary of distance units commonly used in astronomy

Unit	Metres/m	Astronomical units/AU	Light years/ly
1 AU	1.50×10^{11}	–	–
1 ly	9.46×10^{15}	6.32×10^{4}	–
1 pc	3.09×10^{16}	2.06×10^{5}	3.26

WORKED EXAMPLE E5.6

The parallax angle for the star Alpha Centauri is 0.751 arc-seconds.

Calculate its distance from Earth in:

a parsec **b** metre **c** light years **d** astronomical units.

Answer

a distance in parsec = 1/*p*(arc-second) = 1/0.751 = 1.33 pc

b $1.33 \times 3.09 \times 10^{16} = 4.11 \times 10^{16}$ m

c $1.33 \times 3.26 = 4.34$ ly

d $1.33 \times 2.06 \times 10^{5} = 2.67 \times 10^{5}$ AU

26 a Use the HR diagram to estimate the luminosity of a main sequence star which has a surface temperature of 3000 K.

b Estimate the radius of this star from the same diagram.

c Compare your answer to a value of the radius determined from using $L = \sigma A T^4$.

d If the apparent brightness of this star is $4.2 \times 10^{-11}\,\mathrm{W\,m^{-2}}$, estimate its distance from Earth.

27 Explain why stars can be shown in fixed positions on star maps, even though they are moving very fast.

28 a Calculate the total distance travelled in ten years by the Earth as it orbits the Sun. Give your answer in AU.

b The distance between the Sun and Pluto is approximately six billion kilometres.
Express this in AU.

29 Convert an angle of 1 arc-second to:

a degrees

b radians.

30 The parallax angle for Barnard's star is measured to be 0.55 arc-seconds. How far away is it from Earth in:

a pc **b** m **c** ly?

31 Calculate the parallax angles for three stars at the following distances from Earth:

a 2.47×10^{15} km

b 7.90 ly

c 2.67 pc.

32 If the upper limit to parallax measurements is for stars which are 300 ly away, calculate the smallest parallax angle that can be measured accurately.

33 Star A is a distance *x* pc from Earth and has a parallax angle of *θ*.
Determine the parallax angle for a star B which is *x*/2 from Earth.

34 A star is 50 pc from Earth.

a What is this distance in light years?

b Will astronomers be able to detect stellar parallax with this star? Explain your answer.

LINKING QUESTION

- In which ways has technology helped to collect data from observations of distant stars? (NOS)

The age of an expanding Universe

We saw in Topic C.5 that the speed of stars and galaxies away from Earth (*recession speeds*) can be determined from the *Doppler shifts* of radiation received from them on Earth. Combining that information with the latest data about their distances from Earth leads us the important graph shown in Figure E5.17. (A simpler version was shown in Topic C.5.)

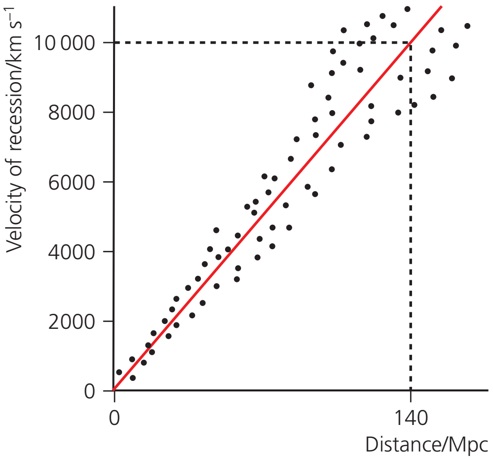


Figure E5.17 Variation of recession speeds of galaxies with their distances from Earth

Assuming that all the stars began moving at the time of the *Big Bang*, an estimate for the age of the Universe can be determined from the gradient of this graph, by a straightforward use of velocity $= \dfrac{\text{distance}}{\text{time}}$:

$$1 \times 10^4\,\mathrm{km\,s^{-1}} = \frac{140\,\mathrm{Mpc}}{\text{time}}$$

Converting to SI units:

$$10^7 = \frac{140 \times (3.09 \times 10^{22})}{t}$$

$$t = 4.3 \times 10^{17}\,\mathrm{s}\ (1.4 \times 10^{10}\,\mathrm{y})$$

The creation of different elements

We have explained the fusion of hydrogen to form helium, now we will briefly outline how other chemical elements can be created in fusion processes called **nucleosynthesis**. (Details need *not* be remembered.) Basically:

◆ **Nucleosynthesis** Creation of new nuclides (elements) from existing, less massive, nuclei.

◆ **Neutron capture** Nuclear reaction in which a neutron is absorbed by a nucleus to form a more massive nucleus.

Higher temperatures (particles with greater kinetic energy) are needed for nucleosynthesis, and these exist in the cores of red giants and super giants.

- For red giant stars formed from main sequence stars of mass less than 4 $M_\odot$, the core temperature can reach 10^8 K and this is large enough for the nucleosynthesis of carbon and oxygen. (Helium is still produced in an outer layer.)
- For larger red giant stars (formed from main sequence stars with masses between 4 $M_\odot$ and 8 $M_\odot$), the core temperature exceeds 10^9 K and this is large enough for the nucleosynthesis of neon and magnesium. (Helium, carbon and oxygen are still produced in the outer layers.)
- For red super giant stars (formed from main sequence stars with masses greater than 8 $M_\odot$), the core temperature is large enough for the nucleosynthesis of elements as heavy as silicon and iron. (The lighter elements are still produced in the outer layers.) See Figure E5.18.

From Topic E.3 we know that the nucleus of iron is one of the most stable (it has one of the highest binding energies per nucleon). This means that there would have to be an energy input to create heavier nuclides. Heavier elements are created by **neutron capture**, but that process is not included in the IB Physics course.

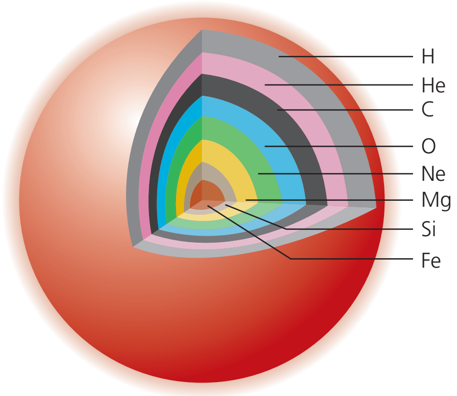


Figure E5.18 The layers of a red super giant

We saw in Topic E.1 that the elements present in a star can be identified from measurements made of line spectra.

35 Explain how an element in the outer layers of a star can be identified from the spectrum received from the star.

36 Explain:

a why very high temperatures are needed to create the more massive nuclides

b why those higher temperatures are found in more massive stars.

37 Outline why the interior of a red super giant star is composed of different layers.

TOK

Knowledge and the knower

Stardust

All the particles in our body existed for billions of years before we were born. They will continue to exist for billions of years after we die. They originated in nuclear reactions in stars and, ultimately, they will be scattered throughout space.

We are all made of stardust.

Acknowledgements

The Publishers would like to thank the following for permission to reproduce copyright material. Every effort has been made to trace all copyright holders, but if any have been inadvertently overlooked, the Publishers will be pleased to make the necessary arrangements at the first opportunity.

Photo credits

p.4 © Harry Collins/stock.adobe.com; **p.6** © Xinhua/Shutterstock.com; **p.14** © Garry Bowden/Sport in Pictures/Alamy Stock Photo; **p.18** © James Gasperotti/ZUMA Wire/Alamy Stock Photo; **p.19** © Shaunwilkinson/stock.adobe.com; **p.26** © Ktsdesign/123RF; **p.27** *l* ©Photodisc/Getty Images, *r* © Red Bull/Shutterstock.com; **p.28** © Andrey Burmakin/stock.adobe.com; **p.30** © EDWARD KINSMAN/SCIENCE PHOTO LIBRARY; **p.33** © Keystone/Zuma/Shutterstock; **p.36 ©** Aquatarkus/stock.adobe.com; **p.38** *l* © Hanan Isachar/Alamy Stock Photo, *r* © *Alaska Stock / Alamy* ; **p.45** © VLRS/stock.adobe.com; **p.46** © Digital Storm/Stock.adobe.com; **p.47** © NASA; **p.54** © Argus - Fotolia; **p.59** © Ian Walton/Getty Images; **p.61** © Vasily Smirnov - Fotolia.com; **p.66** © Robert Crum/Shutterstock.com; **p.68** © Jerry Taylor/Alamy Stock Photo; **p.71** *l* © Martin Barraud/Hoxton/Alamy Stock Photo, *r* © NOEL CELIS/AFP/Getty Images; **p.74** © JGI/Tom Grill / Blend Images / Getty Images; **p.76** *t* © Zoonar/Stephen A. Arce/ ASP Inc/Alamy Stock Photo; *b* © VanderWolf Images/stock.adobe.com; **p.79** © Nurlan Mammadzada/Alamy Stock Photo; **p.81** © Sergio Azenha/Alamy Stock Photo; **p.82** *t* © Creativ Studio Heinemann/Westend61/Getty Images, *l* © Ella/stock.adobe.com; **p.84** © Alban Egger - Fotolia; **p.86** © Master1305/stock.adobe.com; **p.88** © YOSHIKAZU TSUNO/AFP/Getty Images; **p.89** © Gopixa/stock.adobe.com; **p.90** *l* © Sinitar/stock.adobe.com, *r* © Seo byeong gon/Shutterstock.com; **p.92** © Vatcharachai/stock.adobe.com; **p.96** *l* "Courtesy Zettl Research Group, Lawrence Berkeley National Laboratory and University of California at Berkeley." All rights reserved © 2009, *r* © NASA, ESA, HEIC, and The Hubble Heritage Team (STScI/AURA); **p.100** © Creativeneko/stock.adobe.com; **p.101** © Rastislav Sedlak SK/stock.adobe.com; **p.102** © Julija Sapic – Fotolia; **p.103** © Sergey P/stock.adobe.com; **p.104** *l* © Photolink Ltd/Alamy Stock Photo, *r* © Guindilla y Pimienta/stock.adobe.com; p.**108** *l* © David Gee/Alamy Stock Photo, *r* © Andrew Jankunas/Alamy Stock Photo; **p.109** © John Allum; **p.110** © MPIX.TURE/stock.adobe.com; **p.112** © Wanchai/stock.adobe.com; **p.118** © Natatravel/stock.adobe.com; **p.121** © Chaiyasit/stock.adobe.com; **p.122** *tl* © Rattana/stock.adobe.com, *c* © Lazy_Bear/stock.adobe.com; **p.124** *cl* © Richie Chan/stock.adobe.com, *tr* © Peter Adams/stock.adobe.com; **p.127** *l* © Dja65/stock.adobe.com, *r* © Retbool/stock.adobe.com; **p.138** © MaciejBledowski/stock.adobe.com; **p.139** © Prostock-studio/stock.adobe.com; **p.145** © Vkilikov/stock.adobe.com; **p.147** © Lucien Chava/Bill Waterson/Alamy Stock Photo; **p.178** © Luk Ta Kum/Shutterstock.com; **p.179** © TONY MCCONNELL/SCIENCE PHOTO LIBRARY; **p.182** © John Locher/AP/Shutterstock.com; **p.184** © Historical Views/Agefotostock/Alamy Stock Photo; **p.187** © Pixel-Shot/stock.adobe.com; **p.188** © Buzz Pictures/Purestock/SuperStock/Alamy Stock Photo; **p.191** *cr* © Junky_jess/stock.adobe.com, *br* © Andrew_shots/stock.adobe.com; **p.192** *tr* © Igors Jefimovs/Wikimedia Commons/CC BY 3.0, *bl* © Itakdalee/stock.adobe.com; **p.193** © Elias Bitar/Shutterstock.com; **p.195** © Nischaporn/stock.adobe.com; **p.196** NASA Earth Observatory image by Joshua Stevens, using MODIS data from NASA EOSDIS/LANCE and GIBS/Worldview.; **p.197** © Hero Images/Alamy Stock Photo; **p.198** AURA, STScI, NASA, Hubble Heritage Project (STScI, AURA); **p.202** © Granger, NYC./Alamy Stock Photo; **p.203** © ESA (Image by AOES Medialab); background: Hubble Space Telescope, NASA/ ESA/ STScI; **p.210** © NorGal/stock.adobe.com; **p.212** *t* © Ashley Cooper pics/Alamy Stock Photo, *br* © Zenzeta/stock.adobe.com; **p.219** © Kevin Schafer/Alamy Stock Photo; **p.220** © Studio Harmony/stock.adobe.com; **p.221** © Belish/stock.adobe.com; **p.223** © Ugorenkov Aleksandr/Shutterstock.com; **p.230** *l* © LYDIA VEROPOULOU/stock.adobe.com, *r* © EKH-Pictures/stock.adobe.com; **p.231** © Alekss/stock.adobe.com; **p.232** © Climate Central (climatecentral.org); **p.233** © Jonathan Porter/Alamy Stock Photo; **p.235** © Science History Images/Alamy Stock Photo; **p.237** *l* © Anne Trevillion, *r* © Yurii Zushchyk/stock.adobe.com; **p.238** *bl* © Shutterbas/stock.adobe.com, *cr* © Dr Anna Clark; **p.239** *l* © WilliamJu/stock.adobe.com, *r* © Ever/stock.adobe.com; **p.244** © Tawatchai1990/stock.adobe.com; **p.247** © Yossarian6/stock.adobe.com; **p.248** © Kirill Gorshkov/stock.adobe.com; **p.252** © Brian Jackson/stock.adobe.com; **p.255** © Aleksey Telesh/stock.adobe.com;

p.256 © Dyachenko/stock.adobe.com; **p.257** *tl* © Juulijs/stock.adobe.com, *bl* © Deborah Benbrook Photography, *bc* © arquiplay77 – Fotolia, *br* © Luis Viegas – Fotolia.com; **p.271** © Stockr/stock.adobe.com; **p.272** © Ilja.stock.adobe.com; **p.275** © New Africa/stock.adobe.com; **p.278** © Drescher/Ullstein bild/Getty Images; **p.282** © JRP Studio/stock.adobe.com; **p.286** © serghi8/stock.adobe.com **p.287** © niroworld – Fotolia.com; **p.289** © niroworld – Fotolia.com; **p.291** © Olgaarkhipenko/stock.adobe.com; **p.292** © NorGal/stock.adobe.com; **p.293** © David J. Green - electrical / Alamy; **p.296** © Dizfoto1973/stock.adobe.com; **p.300** © After6pm/stock.adobe.com; **p.302** *t* © Suriyo/stock.adobe.com, *c* © Destina/stock.adobe.com, *c* © Goodpics/stock.adobe.com, *b* © Sashkin/stock.adobe.com; **p.303** © Mikel Taboada/stock.adobe.com; p.**305** *l* © Surasak/stock.adobe.com, *r* © Andrei Nekrassov/stock.adobe.com; **p.307** *l* © Cristianstorto/stock.adobe.com, *r* © Cristianstorto/stock.adobe.com; **p.308** © Nyvlt-art/Shutterstock.com; **p.313** © natureshots - Fotolia; **p.314** © Pictorial Press Ltd / Alamy Stock Photo; **p.321** © JuanSt/stock.adobe.com; **p.322** *bl* © Historic Illustrations/PhotoStock-Israel / Alamy Stock Photo, *br* © lyf2/stock.adobe.com, *cr* © JPhoto Researchers/Science History Images/Alamy Stock Photo; **p.323** © Evening_tao/stock.adobe.com; **p.335** © Konstantin Sutyagin - Fotolia; **p.336** © Laurence Cartwright – Fotolia; **p.344** © Mikhail/stock.adobe.com; **p.346** © sciencephotos / Alamy; **p.349** *c* © INTERFOTO/Alamy Stock Photo, *r* © World History Archive/Alamy Stock Photo; **p.352** © Robert Fried/Alamy Stock Photo; **p.354** © Richard Carey/stock.adobe.com; **p.355** © Christopher McGowan/Alamy Stock Photo; **p.357** © alexandre zveiger - Fotolia; **p.359** © Spass/Shutterstock.com; **p.362** © Horizon International Images Limited/Alamy Stock Photo; **p.364** © EDWARD KINSMAN/SCIENCE PHOTO LIBRARY; **p.366** © Sciencephotos/Alamy Stock Photo; **p.369** © Dietrich Zawischa, https://farbeinf.de/static_html/multibeam.html; **p.376** © Mark Williamson – Fotolia; **p.380** © Magnetix/Shutterstock.com; **p.384** © EDWARD KINSMAN/SCIENCE PHOTO LIBRARY; **p.387** *l* © Mary Evans / Grenville Collins Postcard Collection, *r* © 1997 Doug Menuez/Photodisc/Getty Images; **p.389** © IanDagnall Computing / Alamy Stock Photo; **p.392** © Sabena Jane Blackbird / Alamy Stock Photo; **p.394** *l* © Dionisvera/stock.adobe.com, *r* © Whipple Museum of the History of Science; **p.395** *cl* © lkordela – Fotolia.com, *br* © Navintar/stock.adobe.com; **p.396** © IOFOTO/Canva Pty Ltd /Alamy Stock Photo; **p.398** *tl* © REX/Sipa Press, *cl* © Pavel Svoboda – Fotolia.com, *bl* patrickw – Fotolia.com; **p.402** © dave stamboulis / Alamy; **p.403** © CHRISTOPHER PASATIERI/Reuters/Corbis; **p.404** © Avpics / Alamy Stock Photo; **p.405** *tl* © Dave Stamboulis/Alamy Stock Photo, *br* © Carol Kohen/Cultura Creative RF/Alamy Stock Photo; **p.408** © Paopano/stock.adobe.com; **p.411** © CBW/Alamy Stock Photo; **p.414** *cl* © Giorgio Morara/Alamy Stock Photo, *bl* © Laurent dambies/stock.adobe.com; **p.417** © Sciencephotos/Alamy Stock Photo; **p.424** © Vitalii/stock.adobe.com; **p.443** © Jasmina/stock.adobe.com; **p.444** *l* © TED KINSMAN / SCIENCE PHOTO LIBRARY, *r* © James Kirkikis/Alamy Stock Photo; **p.447** Courtesy of Artmechanic/ Wikimedia; **p.451** © Martin/stock.adobe.com; **p.456** © Muhamad zaidi zubir/EyeEm/stock.adobe.com; **p.457** *t* © Africa Studio/stock.adobe.com, *b* © Wittayayut/stock.adobe.com; **p.459** © New Africa/stock.adobe.com; **p.466** © Matt Writtle/PA Images/Alamy Stock Photo; **p.467** © Helene Rogers/Art Directors & TRIP/Alamy Stock Photo; **p.477** © Dr Peter Reid, University of Edinburgh; **p.478** © Joshua Strang, USAF, Wikipedia; **p.481** *l* © CERN, *r* © CERN; **p.493** © John Allum; **p.494** *t* © MaxSafaniuk/stock.adobe.com, *b* © John Allum; **p.496** *l* Courtesy of Wikimedia Commons (http://en.wikipedia.org/wiki/Michael_Faraday), *r* © Granger, NYC./Alamy Stock Photo; **p.501** © Lucky Dragon - Fotolia; **p.505** *l* © Industrieblick/stock.adobe.com, *r* © Blue Planet Studio/stock.adobe.com; **p.506** © oslobis – Fotolia.com; **p.507** © Lena_zajchikova/stock.adobe.com; **p.508** © ARCHIVIO GBB / Alamy Stock Photo; **p.510** © Cesar Reyes; **p.517** © John Allum; **p.525** © Archive Pics/Alamy Stock Photo; **p.549** © European Union, 2022; **p.555** © BSIP SA / Alamy Stock Photo; **p.556** © Nikkytok/stock.adobe.com; **p.563** © GRANGER - Historical Picture Archive / Alamy Stock Photo; **p.567** © NSF/Xinhua Press/Corbis; **p.578** *cl* © Everett Collection/Shutterstock.com, *bl* © DAVID BOILY/AFP/Getty Images, *br* https://commons.wikimedia.org/wiki/File:Yellow_Cake_Uranium_%2814492248719%29.jpg; **p.579** © Fottoo/stock.adobe.com; **p.581** *t* © BJ Warnick/Newscom/Alamy Stock Photo, *b* © MERE WORDS/SCIENCE PHOTO LIBRARY; **p.586** © pixbox77 – Fotolia.com

Index

This page intentionally left blank

FOR THE
IB DIPLOMA
PROGRAMME

THIRD EDITION

Physics

Trust experienced and best-selling authors to navigate the syllabus confidently with this co-published coursebook that implements inquiry-based and conceptually-focused teaching and learning.

- Consolidate skills and improve exam performance with short and simple knowledge-checking questions, exam-style questions and hints to help avoid common mistakes.
- Integrate Theory of Knowledge into your lessons and provide opportunities for cross-curriculum study with TOK boxes and Inquiries that provide real-world examples, case studies and questions.
- Provide opportunities for conceptual discussions and comparisons with linking questions throughout each chapter.
- Develop ATL skills with a range of engaging activities.

About the authors

John Allum taught Physics to pre-university level in international schools for more than thirty years (as a head of department). He has now retired from teaching, but lives a busy life in a mountainside village in South East Asia. He has been an IB examiner for many years.

Paul Morris has taught IB Physics and IB Theory of Knowledge for over 20 years, has led teacher workshops internationally and has examined Theory of Knowledge. As an enthusiast for the IB concept-based continuum Paul designed and developed Hodder Education's 'MYP by Concept' series and was author and co-author of the Physics and Sciences titles in the series.

HODDER EDUCATION
e: education@hachette.co.uk
w: hoddereducation.com

ISBN 978-1-3983-6991-7
9 781398 369917