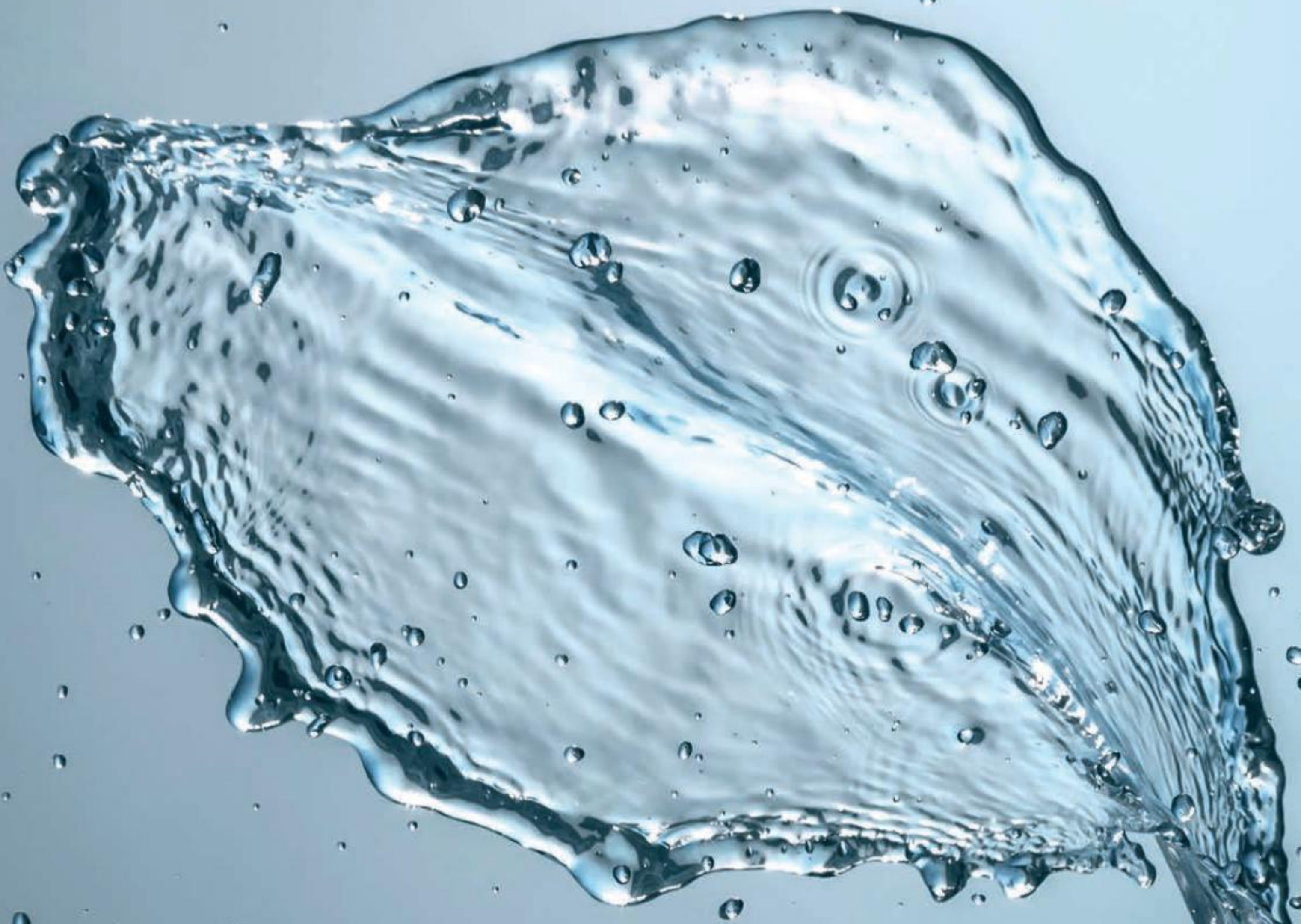
NELSON PHYSICS UNITS 1 & 2

FOR THE AUSTRALIAN CURRICULUM



NEIL **CHAMPION**
GEOFF **CODY**
ROBERT **FARR**
MEGAN **MUNDY**
KATE **WILSON**

This is a legal agreement between the you, (the "Customer") and Cengage Learning Australia Pty Limited (ABN 14 058 280 149) (the "Licensor") which provides the terms and conditions of this non-exclusive licence and the limited warranty for the Product. Use of the Product indicates an acknowledgement that the Customer has read and agreed to be bound by the terms and conditions of this Agreement. If you do not agree to these terms and conditions, return the Product to the place of purchase within 15 days of the date of purchase (with proof of purchase) for a full refund

1. **Licence Grant**
   You do not receive title to the Product. Copyright in the Product (which includes all images, photographs, video, animations, audio, music and text incorporated in the Product, including all of the accompanying printed material) is owned by the Licensor and/or its suppliers and is protected by Australian copyright laws. The Licensor grants you a non-exclusive licence to use the Product subject to the restrictions and terms set out in this Agreement.

2. **A Licence allows you to:**
   Use the Product on your computer. The Customer represents that they shall in no way place the Product in the public domain or in any way compromise our copyright in the Material. You agree to take reasonable steps to protect our copyright.

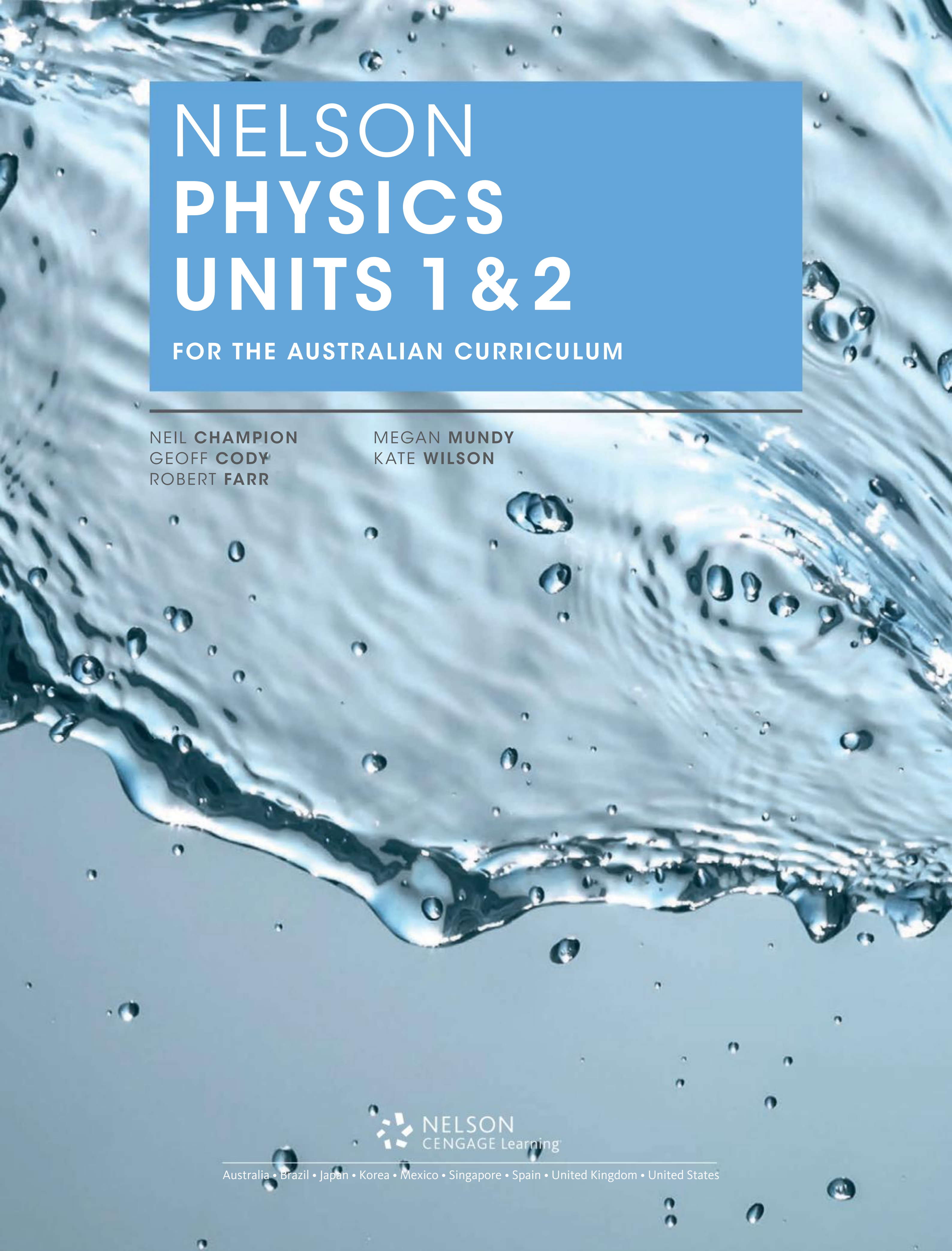
3. **You may not:**
   Alter, modify, translate, reverse engineer, decompile, or adapt the software or create derivative works based on the Product.

   Make further copies by any means technological, electronic, digital whatsoever without the written permission of the Licensor.

   Rent or transfer all or any part of your rights under this Agreement. Remove or alter any copyright or other proprietary notice or label attached to the software.

4. **Termination**
   Any failure to comply with the terms and conditions of this agreement will result in the automatic termination of this licence. Upon termination of this licence for any reason, the Customer must destroy or return to the Licensor all copies of the software and accompanying documentation.

5. **Warranties**
   To the extent permitted by law, the Licensor's liability for any breach of the warranty or any term implied by law into this licence is limited to the lowest cost of replacing the goods, acquiring equivalent goods or having the goods repaired.



NELSON PHYSICS UNITS 1 & 2

FOR THE AUSTRALIAN CURRICULUM

NEIL **CHAMPION**
GEOFF **CODY**
ROBERT **FARR**
MEGAN **MUNDY**
KATE **WILSON**

NELSON
CENGAGE Learning

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

Nelson Physics Units 1 & 2 for the Australian Curriculum
1st Edition
Neil Champion
Geoff Cody
Robert Farr
Megan Mundy
Kate Wilson

Publishing editor: Eleanor Gregory
Senior editor: Kelly Robinson
Editor: Marta Veroni
Proofreader: Jill Pope
Indexer: Rod Andrew
Cover design: Claire Atteia
Text design: Claire Atteia
Art direction: Luana Keays
Cover image: Shutterstock.com/Dmitry Laudin
Permissions researcher: Miriam Allen
Production controller: Julie McArthur
Typeset by: MPS Limited

Any URLs contained in this publication were checked for currency during the production process. Note, however, that the publisher cannot vouch for the ongoing currency of URLs.

For product information and technology assistance,
in Australia call **1300 790 853**;
in New Zealand call **0800 449 725**

For permission to use material from this text or product, please email
aust.permissions@cengage.com

National Library of Australia Cataloguing-in-Publication Data
Champion, Neil (Neil Douglas), author.
Nelson physics units 1 & 2 for the Australian curriculum /
Neil Champion, Geoff Cody, Rob Farr,
Megan Mundy, Kate Wilson.

9780170242103 (paperback)
Includes index.
For secondary school age.

Physics--Textbooks.
Physics--Study and teaching (Secondary)
Education, Secondary--Curricula--Australia.

530

Cengage Learning Australia
Level 7, 80 Dorcas Street
South Melbourne, Victoria Australia 3205

Cengage Learning New Zealand
Unit 4B Rosedale Office Park
331 Rosedale Road, Albany, North Shore 0632, NZ

For learning solutions, visit **cengage.com.au**

Printed in China by China Translation & Printing Services.
1 2 3 4 5 6 7 18 17 16 15 14

CONTENTS

PREFACE

Nelson Physics Units 1 & 2 for the Australian Curriculum has been written to meet the requirements of the ACARA Australian Senior Secondary Curriculum – Physics. The text has been written to enable students to meet the A level Achievement Standard. It also allows all students to maximise their learning and results.

Physics deals with the wonderfully interesting and sometimes strange universe. Physicists investigate space and time (and space–time), from the incredibly small to the incredibly large, from nuclear atoms to the origin of the universe. They look at important, challenging and fun puzzles and try to work out solutions.

Physicists deal with the physical world where energy is transferred and transformed, where things move, where electricity and magnetism affect each other, where light and matter interact. As a result, physics has been responsible for about 95 per cent of the world's wealth – electricity supply and distribution, heating and cooling systems, computers, diagnostic and therapeutic health machines, telecommunications and safe road transport.

But physicists are not just concerned with observing the universe. They explain these observations, using models, laws and theories. Models are central to physics. Physicists use models to describe, explain, relate and predict phenomena. Models can be expressed in a range of ways – via words, images, mathematics (numerical, algebraic, geometric, graphical) or physical constructions. Models help physicists to frame physical laws and theories, and these laws and theories are also models of the world. Models are not static. As scientific understanding of concepts or physical data or phenomena evolves, so too do the models scientists use to describe, explain, relate and predict these. Thus, the text emphasises both the observations and the quantitative data upon which physicists develop the models they use to explain the data. Central to this is the rigorous use of mathematical representations as a key element of physics explanations.

Nelson Physics Units 1 & 2 for the Australian Curriculum is written by academic and classroom teaching experts. They were chosen for their comprehensive knowledge of the physics discipline and best teaching practice in physics education at secondary and tertiary levels. They have written the text to make it accessible, readable and appealing to students. They have included numerous, current contexts to ensure students gain a wide perspective of the breadth and depth of physics. This contextual, mathematically rigorous and methodological approach is designed to ensure students can reach the highest possible standard. The intention has been to ensure all students achieve the level of depth and interest necessary to pursue tertiary studies in physics, engineering, technology and other scientific courses.

Each chapter follows a consistent pattern. Learning outcomes from the Science Understanding strand appear on the opening page. Learning outcomes from the Science as a Human Endeavour and Science Inquiry Skills strands are mapped on pages XII–XIII. The text is then broken into manageable sections under headings and subheadings. Relevant diagrams support the text. New terms are bolded and defined in a glossary at the end of the chapter. Important concepts are summarised in boxes to assist students to take notes.

Worked examples, written to connect important ideas and solution strategies, are included throughout the text. Solutions are written in full, including algebraic transformations, substitution of values with units, and a proposed marking scheme. In order to consolidate learning, students are challenged to try similar questions on their own.

Numerous Question sets appear at the end of logical sections. There is a comprehensive set of Review questions at the end of each chapter. All questions have been graded from lower to higher order thinking skills: Remembering, Understanding, Applying, Analysing and Reflecting. Numerical answers appear at the end of the book. Complete worked answers appear on the NelsonNet teacher website.

Experiments and Investigations demonstrate the high level of importance the authors attach to understanding-by-doing physics. These activities introduce, reinforce and enable students to practise Science Inquiry skills, especially experimental design, data collection, analysis and conclusions. The

Scientific Investigations chapter consolidates important investigative concepts and values. It enables students to learn and reflect on their experience as puzzle-solvers and investigators. It is an invaluable tool for students undertaking an Extended experimental investigation (EEI).

Système Internationale (SI) units and conventions, including accuracy, precision, uncertainty and error, are introduced in the Measurement chapter. This invaluable tool supports student learning through chapter questions, experiments and investigations as well as their EEI.

Case studies elucidate Examples in context, part of the Science as a Human Endeavour strand. Scientific literacy activities ensure students develop the key Science Inquiry Skills capability of comprehending and evaluating scientific claims and synthesising a response.

Nelson Physics Units 1 & 2 for the Australian Curriculum provides students with a comprehensive study of modern physics that will fully prepare them for exams and any future studies in the area.

Neil Champion
Series editor

AUTHOR AND REVIEWER TEAMS

Authors

Neil Champion

Neil Champion BSc(Hons) DipEd MScEd was directly involved in writing the Australian Senior Physics Curriculum. He is an experienced secondary Physics teacher at Buckley Park College in Victoria and has held numerous school leadership positions in state and independent schools, including Principal. Neil has taught university level Physics and Physics Teaching Method for pre-service teachers. As VCAA Science Manager, he was responsible for the development and implementation of the world's first senior secondary curriculum in photonics and synchrotron physics. For over 20 years, Neil has published school science texts for VCE Physics, IB Middle Years Program and 7–10 Australian Science Curriculum.

Geoff Cody

Geoff Cody graduated with a BSc and DipEd from the University of Western Australia and commenced a teaching career that spanned forty years. For most of that period, including a year on exchange in England, he taught Physics and Chemistry at the senior level. In addition to his teaching commitments Geoff developed a number of teaching materials and courses for the University of Western Australia's science enrichment program 'Spice', Edith Cowan University and the Gravity Discovery Centre. He was Head of Science in both government and private schools before leaving to take up a post teaching Science Communication at Curtin University. Geoff's passion for science is exemplified by the innovative science programs and materials he has written to excite and nurture students' interest in science. In 2009 Geoff won the Tower award for his outstanding contribution to The Gravity Discovery Centre.

Rob Farr

Rob Farr has been teaching Physics in NSW schools for 30 years, 18 of those as Head of Science. He is currently teaching at Brigidine College, St Ives in Sydney. He has extensive experience working with the Board of Studies in various roles reviewing syllabuses and HSC examinations. He has had 25 years experience in marking HSC examinations. Rob co-authored the successful *Physics in Focus* texts and he contributed to the *Biology in Focus* series, along with the *iScience for NSW* texts being used for the new NSW syllabus based on the ACARA Australian Curriculum. Rob's ability to bring his enthusiasm and passion for Physics into his writing is evident in his work, making Physics accessible to students around the country.

Megan Mundy

Megan Mundy is a science teacher, photographer and swing dancer based in Victoria. She taught Physics at Brunswick Secondary College where she was also the Head of Science. She has marked year 12 exams on numerous occasions for the VCAA. Megan has written previously for Nelson Physics textbooks as well as developing teacher support materials.

At present Megan runs a collaborative science outreach program from six Victorian universities. She trains university science students to work in classrooms as peer mentors to generate enthusiasm for science. She has developed hands-on resources for mentors and teachers based on the new Australian curriculum.

Dr Kate Wilson

Dr Kate Wilson is a senior lecturer in the School of Engineering and Information Technology at the Australian Defence Force Academy (UNSW Canberra). She has a PhD in physics and a Graduate Diploma in Secondary Teaching (Science). Kate is a past director of the Australian Science Olympiads Physics Program. She has been a member of the Physics Education Research Group at the University of Sydney and held an Innovative Teaching and Educational Technology Fellowship at UNSW.

Reviewers

The publisher and authors would like to thank the reviewers for their valuable assistance in developing the chapter manuscripts.

- Gary Bass – Macleod College, Vic
- Charles Coleiro – Ave Maria College, Vic
- Dr Stephen Hughes – Queensland University of Technology, Qld
- Dr David Mills – Monash University, Vic
- Dr Geoffrey Pang – Central Queensland University, Qld
- Dr Anton Rayner – University of Queensland, Qld
- Dr Geoff Swan – Edith Cowan University, WA
- Dr Stephen Zander – Christ Church Grammar School, WA

USING NELSON PHYSICS

Nelson Physics Units 1 & 2 for the Australian Curriculum has been purposely crafted to enable students to achieve maximum understanding and success in this subject. Each page has been carefully considered to provide students with all the information that they need without appearing cluttered or overwhelming. Students will find it easy to navigate through each chapter and see connections between chapters through the use of linking icons. Practical work has been integrated within the text so the students can see the importance of the interconnectedness between the theoretical and practical aspects of physics.

Each chapter begins with a **Chapter opener**. This presents the content descriptions from the Science Understanding strand of the senior Physics Australian Curriculum that will be incorporated into the chapter.

The text has been authored and reviewed by experienced physics educators, academics and researchers to enable students to achieve the maximum level of achievement of which they are capable. A number of devices have been utilised to improve literacy and understanding. One of these is the use of shorter sentences and paragraphs. This is coupled with clear and concise explanations and real-world examples. New terms are bolded as they are introduced and appear in an end-of-chapter and an end-of-book glossary.

Important concept

Throughout the text, important ideas, formulas and laws are summarised in **Important concept** boxes.

WORKED EXAMPLE

Mathematical representations and relationships are presented in context. Step-by-step instructions on how to perform mathematical calculations are shown in the **Worked examples**. The logic behind each step is explained and approximate marks allocated so that students can see that they need to show their full working out. Students are then able to practise these steps by attempting the related problems presented at the end of the worked example. Answers to these problems are given at the back of the book.

Physics is a practical subject and students need to be given the opportunity to explore and discover through practical activities. These are presented in three different types of boxes throughout the text.

ACTIVITY

The **Activities** provide the opportunity for short hands-on tasks to clarify or reinforce a concept. The activities can be performed either individually or in groups.

EXPERIMENT

The **Experiments** introduce and reinforce the Science Inquiry Skills strand of the Australian Curriculum. Experiments contain guided instruction on the materials, procedure, collection and analysis of results and discussion.

INVESTIGATION

The **Investigations** allow students to practise Science Inquiry Skills. They provide them with the opportunity to design and carry out their own scientific investigation either individually or in a group. Students are prompted to consider ideas for improvement and further investigation to illustrate that science is an ongoing and improving process. Further information on how to conduct a scientific investigation can be found in Chapter 13 Scientific Investigations on page 419.

Risk assessment

The **Risk assessment** table occurs within the experiment and investigation boxes. The table highlights the risks to the students and provides suggestions on how to minimise these risks. Teachers are able to supplement this table by adding any further risks specific to their school situation.

Many so-called scientific claims are used to promote an issue, product or idea. It is important that students are able to understand and analyse the information presented to them so they can make well-informed choices. The **Scientific literacy** box presents a scientific text or media piece that enables students to use evidence to evaluate the claims and conclusions presented. It allows them to use reasoning and knowledge beyond the information presented to construct a valid scientific argument.

Scientific literacy

Case studies provide students with the opportunity to see how science is applied using an up-to-date and real-world example in context.

Case study

Full understanding of a concept is often constructed from many pieces of information. Due to the sequential nature of a book, this information cannot always be presented together as it is best placed in other chapters. Links between concepts that occur on other pages and chapters are indicated using **Margin notes**.

Margin note

Review of student understanding is attained through the **Question sets** throughout each chapter. Questions are ordered from lower to higher order thinking skills. The addition of reflection questions gives students the opportunity to reflect upon not only what they are learning but why they are learning it, and how they are learning it.

QUESTION SET

The end of chapter review provides:

- a **Summary** of the important concepts presented within the chapter. This will be a valuable tool when students are revising for tests and exams
- a **Glossary** of all the new terms introduced within the chapter
- **Chapter review questions** that review understanding of concepts from the chapter.

Questions are ordered from lower to higher order thinking skills and also include reflection questions.

Where answers to questions are numerical, they are provided in the back of the book.

NelsonNet

NelsonNet is your protected portal to the premium digital resources for Nelson textbooks located at www.nelsonnet.com.au. Once your registration is complete you will have access to an exciting and stimulating digital suite of resources for each chapter that are designed to enhance and reinforce learning.

Each chapter will be supplemented with the following digital resources.

- **Prior learning worksheet** to revise content from Year 9 or 10 that is a prerequisite to understanding
- **Activity sheets,** including theory and practical exercises
- **Revision sheets** to complete at home to revise class work
- A **review quiz** containing 20 auto-correcting multiple-choice questions at the end of each chapter to review understanding
- **Videos** to assist students in understanding complex concepts. Pages that have videos associated with them are indicated with a blue icon
- **Tutorials** to assist students in understanding complex concepts
- **Chapter summaries** to support students when revising for tests and exams
- **Links** to websites that contain extra information. These are hotspotted within the ebook and they can also be accessed at http://pac1and2.nelsonnet.com.au.

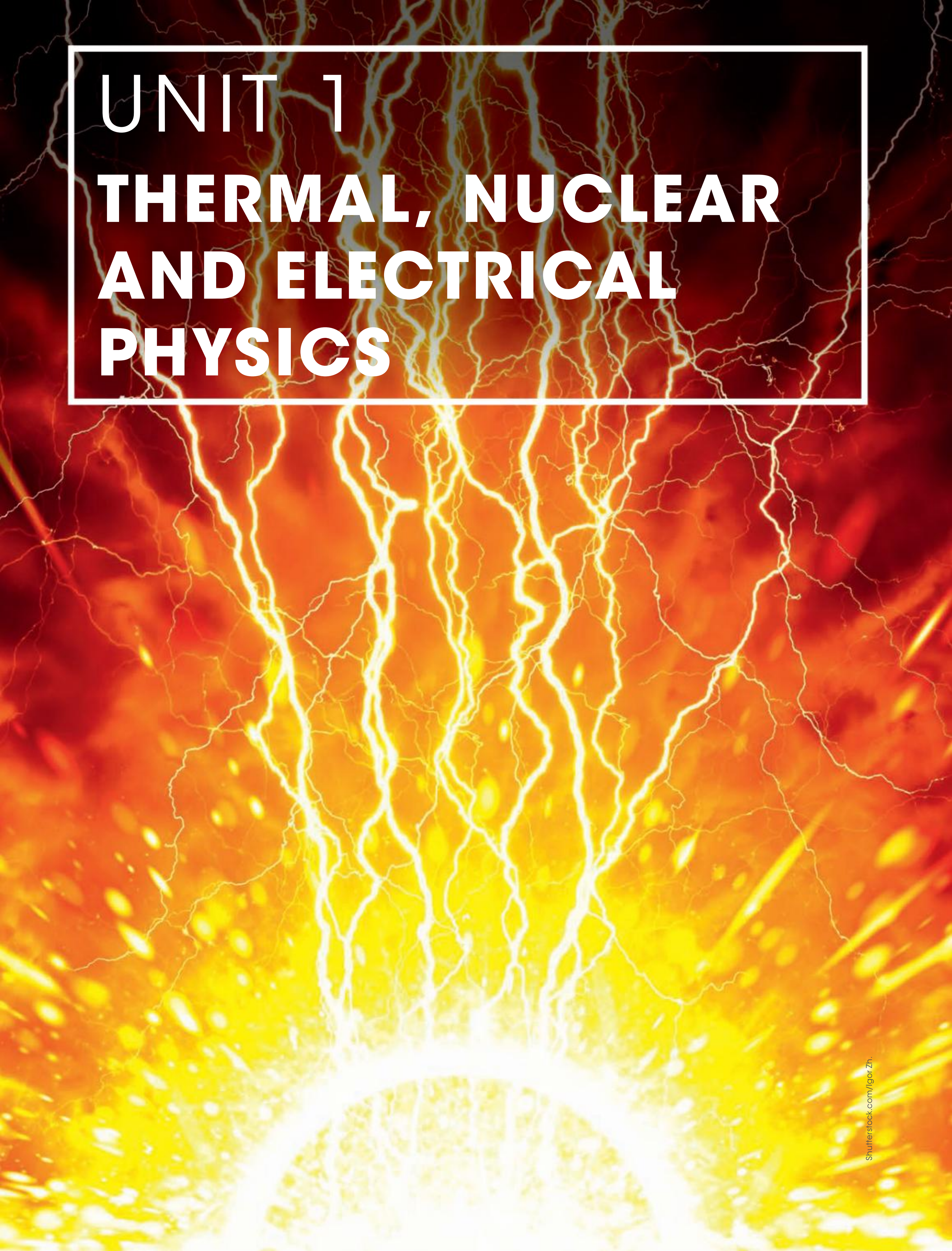
Please note that complimentary access to NelsonNet and the NelsonNetBook is only available to teachers who use the accompanying student textbook as a core educational resource in their classroom. Contact your sales representative for information about access codes and conditions.

CURRICULUM GRID

		Unit 1						Unit 2						
		Chapter												
		1	2	3	4	5	6	7	8	9	10	11	12	13
Science Inquiry Skills	Identify, research, construct and refine questions for investigation; propose hypotheses; and predict possible outcomes (ACSPH001 AND ACSPH040)	✓	✓		✓			✓		✓		✓		✓
	Design investigations, including the procedure/s to be followed, the materials required, and the type and amount of primary and/or secondary data to be collected; conduct risk assessments; and consider research ethics (ACSPH002 AND ACSPH041)	✓	✓				✓	✓		✓	✓	✓		✓
	Conduct investigations, including using temperature, current and potential difference measuring devices (ACSPH003), the manipulation of devices to measure motion and the direction of light rays, safely, competently and methodically for the collection of valid and reliable data (ACSPH042)	✓	✓		✓	✓	✓	✓		✓		✓		✓
	Represent data in meaningful and useful ways, including using appropriate Système Internationale (SI) units and symbols; organise and analyse data to identify trends, patterns and relationships; identify sources of random and systematic error and estimate their effect on measurement results; identify anomalous data and calculate the measurement discrepancy between experimental results and a currently accepted value, expressed as a percentage; and select, synthesise and use evidence to make and justify conclusions (ACSPH004 AND ACSPH043)	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
	Interpret a range of scientific and media texts, and evaluate processes, claims and conclusions by considering the quality of available evidence; and use reasoning to construct scientific arguments (ACSPH005 AND ACSPH044)	✓	✓	✓			✓	✓	✓	✓	✓			✓
	Select, construct and use appropriate representations, including text and graphic representations of empirical and theoretical relationships, flow diagrams, nuclear equations and circuit diagrams, vector diagrams, free body/force diagrams, wave diagrams and ray diagrams, to communicate conceptual understanding, solve problems and make predictions (ACSPH006 AND ACSPH045)	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓		✓

		Unit 1						Unit 2						
		Chapter												
		1	2	3	4	5	6	7	8	9	10	11	12	13
	Select, use and interpret appropriate mathematical representations, including linear and non-linear graphs and algebraic relationships representing physical systems, to solve problems and make predictions (ACSPH007 AND ACSPH046)	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓
	Communicate to specific audiences and for specific purposes using appropriate language, nomenclature, genres and modes, including scientific reports (ACSPH008 AND ACSPH047)	✓	✓					✓	✓	✓		✓	✓	✓
Science as a Human Endeavour	Science is a global enterprise that relies on clear communication, international conventions, peer review and reproducibility (ACSPH009 AND ACSPH053)	✓		✓	✓						✓	✓		
	Development of complex models and/or theories often requires a wide range of evidence from multiple individuals and across disciplines (ACSPH010 AND ACSPH054)	✓	✓	✓	✓				✓	✓	✓	✓		
	Advances in science understanding in one field can influence other areas of science, technology and engineering (ACSPH011 AND ACSPH055)	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		
	The use of scientific knowledge is influenced by social, economic, cultural and ethical considerations (ACSPH012 AND ACSPH056)	✓	✓	✓	✓	✓	✓		✓	✓		✓		
	The use of scientific knowledge may have beneficial and/or harmful and/or unintended consequences (ACSPH013 AND ACSPH057)	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		
	Scientific knowledge can enable scientists to offer valid explanations and make reliable predictions (ACSPH014 AND ACSPH058)		✓	✓	✓		✓		✓	✓	✓	✓		
	Scientific knowledge can be used to develop and evaluate projected economic, social and environmental impacts and to design action for sustainability (ACSPH015 AND ACSPH059)				✓	✓	✓		✓	✓				

UNIT 1
THERMAL, NUCLEAR AND ELECTRICAL PHYSICS



Shutterstock.com/Igor Zh.

CHAPTER 1 HEATING AND COOLING

By the end of this chapter you will have covered the following material.

Science Understanding

- The kinetic particle model describes matter as consisting of particles in constant motion, except at absolute zero (ACSPH017)
- All systems have thermal energy due to the motion of particles in the system (ACSPH018)
- Temperature is a measure of the average kinetic energy of particles in a system (ACSPH019)
- Provided a substance does not change state, its temperature change is proportional to the amount of energy added to or removed from the substance; the constant of proportionality describes the heat capacity of the substance (ACSPH020)
- Change of state involves internal energy changes to form or break bonds between atoms or molecules; latent heat is the energy required to be added to or removed from a system to change the state of the system (ACSPH021)

Introduction

According to the Big Bang theory, the universe began with energy – incredibly hot, massively dense energy. Then the energy expanded. With the expansion came space and time, intimately connected as space–time. Things began to condense out of the energy. In the first few minutes, the forces that connect things condensed out. Electrons, protons and neutrons were formed. Over the next 500 000 years, matter as we know it began to form. In one sense, we are all about 13.8 billion years old, the age of the universe. This is when the energy that eventually condensed into the matter of our bodies came into existence.

You will learn more about the Big Bang in Unit 4 Physics.

WOW

Big Bang

According to the Big Bang theory, the 'big bang' was a 'little squib' for about 10^{-39} seconds. Then rapid cosmic inflation took over. Inflation explains the extreme uniformity of the universe on the grand scale. After inflation the universe continued to expand at a much slower rate. A few billion years later, it began a new, accelerating expansion phase. This appears to be caused by dark energy, about which very little is known. Cosmic expansion continues to take place in all directions and from all points in space.

The amount of energy in the universe at the beginning is the same amount of energy that is in the universe today. It has been condensed into matter and re-distributed in an extraordinary diversity of ways. The **average temperature** of the universe is now 2.725°C above **absolute zero**, the coldest possible temperature (−273.15°C).

HOT BIG BANG THEORY

Learn more about the Hot Big Bang theory from Cambridge University cosmologists.

Human control over heat energy has advanced from simple fires for warmth and cooking to heat engines that perform work. Steam engines and turbines, external and internal combustion engines and heat exchange systems all contribute to prosperity, health and comfort.

But there has been a cost. Since the 19th century Industrial Revolution, fossil fuel use has significantly affected Earth's energy balance. The scientific consensus is that climate change is occurring, powered by these **anthropogenic** factors. Pressure by governments to ensure **energy security** may lead to greater political instability.

Living things have adapted to maintain a constant **core temperature** by the **internal process** of homeostasis. Humans use both internal and external ways of maintaining our core temperature at 37°C by internal processes (blood distribution, sweating, goosebumps, shivering, increased **metabolic activity**) and external effects (clothing, energy-efficient buildings, air conditioning, heaters, fans).

Strange things can happen when some materials are cooled to very low temperatures. Helium at almost absolute zero flows up and out of its container. Superconductors lose their electrical resistance. Low temperature physics or **cryogenics** is applied to efficient electrical energy distribution, rocketry, satellites, food preservation and organ storage.

In this chapter, we investigate **thermodynamics**, the physics of heating and cooling phenomena and their explanation.

NASA/NOAA/GSFC/Suomi NPP/VIIRS/Norman Kuring

Figure 1.1 ▲
Clouds are created and driven by the Sun's solar energy.

ACTIVITY 1.1

SENSING HOT AND COLD

A thermometer gives you an **objective measurement** (quantitative) of the temperature. Your hands give a **subjective indication** (qualitative) of the temperature. Heat-detecting nerves in your skin detect the rate at which energy (heat) is transferred to or from your skin. The rate of heat transfer depends upon temperature difference. We interpret the sensation as temperature.

Aim

To explore how we sense hot and cold

What you will need

- Three bowls of water at different temperatures: one cold, one as hot as your hands can stand, and one with an equal mixture of both hot and cold
- A thermometer

What to do

1 Record the temperature of each bowl of water.

2 Place your left hand in the cold water and your right hand in the hot water for 2 minutes.

3 Now place both your hands in the mixture.

◀ **Figure 1.2**
a) First, place your hands in the hot and cold water as shown.
b) Second, place your hands in the warm water as shown.

What did you discover?

1 Record descriptions of the sensation of 'hotness' or 'coldness' in each hand:
  a at the start.
  b in the mixture.

Conclusion

Use the idea of heat transfer to or from the water to explain your observations.

Discussion

Identify two or more advantages and disadvantages of subjective and objective measurements of heat.

Models and explanations

At the beginning of this chapter, we introduced a theory: the Big Bang theory. This is an explanation of a set of very significant observations about the way the entire universe works. Theories work through models. Models are representations of reality that help us see connections between data and make predictions that might otherwise not be noticed. They include descriptions, diagrams, images, physical models, graphs and equations. Models all represent the reality, but they are not the reality. The Big Bang theory is just that – a theory that explains and predicts phenomena.

Energy is a central concept in many theories and models in physics, including the Big Bang theory. We use the **law of conservation of energy** to explain many observations and predict what will happen in many situations. There are many forms of energy.

One of the key ideas that we will be using in this chapter is the concept of **heat** energy.

Heat is energy that is in the process of being transferred from one place to another due to a temperature difference.

We also use the verb **to heat** to mean the *process* of adding heat to something; for example, heating food. In physics, we do not use the word heat to describe a quantity of stored energy, rather it is energy that is being transferred.

This is similar to the way we define **work**. Work is energy that is being transferred due to the action of a force, but it is also not stored in an object or system.

It took a long time for the scientific community to accept that heat was a form of energy. In the 17th century, Antoine Lavoisier (1743–94) developed the caloric theory. He argued that caloric was a kind of weightless **fluid** capable of flowing from hotter to colder bodies through tiny holes in materials.

Sir Benjamin Thompson (Count Rumford, 1753–1814) undertook many experiments on heat and mechanical work. He concluded they were manifestations of the same thing – energy. The French physicist Sadi Carnot (1796–1832) developed these ideas further, especially for engines.

Figure 1.3 ▲
James Joule 1818–89

James Joule (1818–89) conducted some extremely careful experiments on mechanical energy and heat. He used a known amount of mechanical energy to raise the temperature of a known quantity of water. His data demonstrated conclusively that heat and mechanical energy are quantitatively equivalent. Joule's evidence led to the establishment of the law of conservation of energy (the **first law of thermodynamics**). This is the foundation of all studies of heat, or thermodynamics. The importance of his work has been honoured by the SI unit for energy, joule, J.

Joule worked with Lord Kelvin to develop the absolute scale of temperature, and made observations on magnetostriction in which materials change shape when magnetised. He also discovered the relationship between the flow of current through a resistance and the amount of heat generated. This is now called Joule's law.

Caloric theory and energy theory co-existed for a long time before energy became the preferred concept. In our discussions of heat transfer, there are still echoes of caloric theory in the way we discuss the flow of energy from hotter to colder objects. However, heat is not a substance.

The kinetic particle model of matter

Matter can exist in four different states: solid, liquid, gas and **plasma**. Solids have fixed shapes, fixed volumes and are mostly incompressible. Liquids have fixed shapes, fixed volumes and are more or less incompressible. Gases have no fixed shape or volume and are compressible. Plasmas are similar to gases but are made up of charged particles. How is all this to be explained?

We use the **kinetic particle model** of matter to explain the states of matter, and changes between states. According to the kinetic particle model, all matter is made up of small particles that are in constant motion.

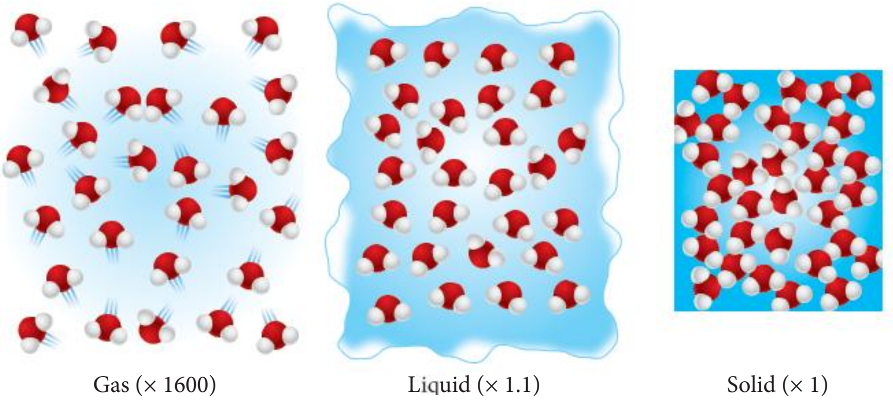


Figure 1.4 ▶
The states of matter showing approximate volume changes. (Note the change in scale: a mass of gas has a volume about 1600 times greater than the solid state.)

In a solid, the particles (atoms or molecules) are attached to each other by bonds that behave a bit like springs. There is an ideal length for any bond, but it is possible for the bond to be stretched and compressed like a spring. The solid material has **potential energy** because of these bonds. When people talk about chemical energy in food or fuels, it is this potential energy associated with the bonds between atoms that they are referring to. The atoms also have **kinetic energy**. Even in a solid, the atoms are all vibrating and moving about constantly. The material itself may not be going anywhere, but every atom is moving. It is a bit like a large assembly of students all sitting in their own chairs, but each one fidgets and leans side to side to talk to their neighbours.

In a liquid, the particles are only very loosely bound. There is still potential energy associated with interactions between the particles, but less than in a solid. However, the particles typically have very much more kinetic energy.

In a gas, the bonds between molecules or atoms have broken and the particles are free to move. There is no longer potential energy associated with bonds between particles (although there is still energy associated with bonds within particles). We model the particles of a gas as being in constant motion. When they interact they do so by colliding and undergoing elastic collisions. In an **elastic collision** kinetic energy is conserved. Kinetic energy is transferred from one particle to another, but not converted into potential energy. This model of a gas is the kinetic particle model or ideal gas model. The pressure of a gas is due to the constant collisions between the gas particles. As we shall see, temperature is a measure of the average kinetic energy of particles. Hence, pressure can be seen as an energy density (energy per unit volume). You may have met the ideal gas law in chemistry: $PV = nRT$. This is really just a statement that the energy of a gas (PV) is equal to the amount of gas (n, moles) multiplied by the temperature of the gas (T). The constant, R, is a scaling factor to convert between temperature and energy units.

Assumptions of the kinetic particle model:

- All matter is made up of small particles in constant motion; they have kinetic energy.
- Collisions between particles are perfectly elastic; the total kinetic energy before and after the collision is the same.
- Potential energy is stored in the 'springs' that connect the particles; potential energy depends on the distance between particles.

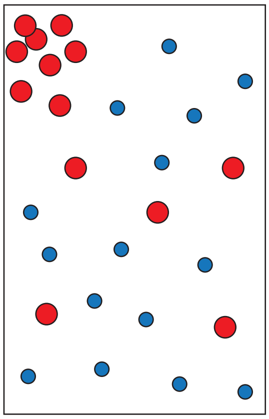
We call the sum of the kinetic energy of the atoms and the potential energy due to their interaction the **internal energy** of the material. Internal energy is energy that cannot easily be seen. It does not include any kinetic energy due to the bulk movement of the material, or potential energy due to external forces such as gravity. However, internal energy is a form of energy, and as such is important to consider when applying the principle of conservation of energy.

A system with internal energy is able to transfer heat to its environment, and may also be able to do work by applying a force to some part of its environment. Note that in Figure 1.4 the volume of an amount of gas is typically 1600 times greater than the same amount of the liquid material. This difference in volume is used in engines to do work. For example, in a steam engine water is boiled by burning coal in a firebox inside or against a boiler. The heat from the burning coal acts to change the state of the water from liquid to gas. The pressure of the steam pushing on a piston does work on rods that connect to the wheels, and thus drive the locomotive. Hence, the internal energy of the fuel, the coal, is converted into work done on a train being pulled behind the engine. This is possible because of the change of state of the water, and the different properties of the two states. We shall look in more detail at energy transfers by heat and work in the next chapter.

Diffusion

Smells waft to us from many places. This **diffusion** of gases is explained by the kinetic particle model. Particles move unseen from their source through a 'sea' of randomly moving air particles that are relatively far away from each other.

Diffusion is very rapid in gases, slower in liquids and can even occur between solids under pressure.



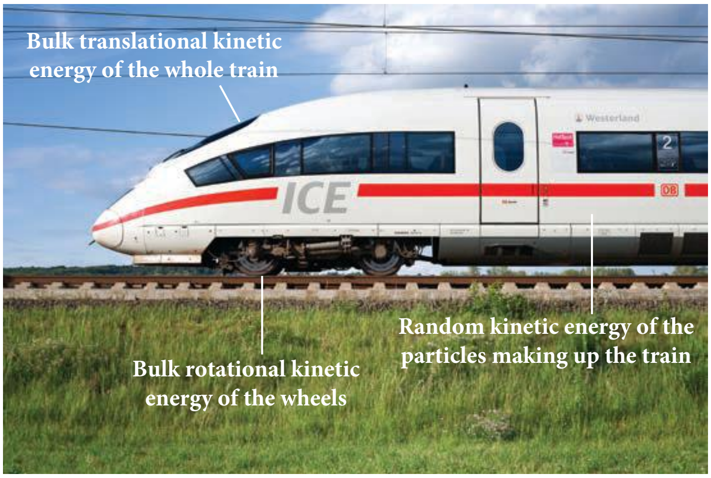
▲ **Figure 1.5**
Particles of a perfume (red) diffuse through the air (blue) away from the source in the top left-hand corner (very small scale).

The energy model

Energy exists in many forms, including heat, light, mechanical, gravitational, electrical, magnetic, sound and chemical. We will even find that mass is a form of energy. No matter the form, energy is still energy. The 'form' is often named by its origin (e.g. nuclear, solar). All forms of energy can be *transformed* from one form to another and *transferred* from one place to another. For example, when you turn on an electrical bar heater, the electrical energy is transformed to radiant heat and light energy.

The SI unit of measurement of energy is the **joule** (J). It is approximately equivalent to the effort required to lift a 100 g apple from the ground to a height of 1 m.

The two major forms of energy are kinetic energy (energy associated with movement) and potential energy (energy ready to be used). All energy sources can ultimately be reduced to these two forms.



iStockphoto/ollo

Figure 1.6 ▼ A moving train possesses different forms of kinetic energy.

Kinetic energy

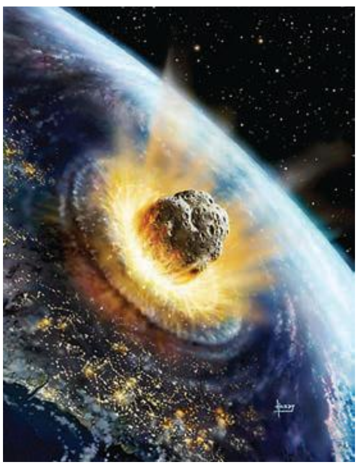
Kinetic energy is the energy a body possesses due to its motion. There are a number of forms of kinetic energy. Take a moving train, for example. It has bulk translational kinetic energy due to the straight-line motion of the whole train. It has bulk rotational kinetic energy in the rotating wheels and engine parts. It has disorganised vibration kinetic energy due to the vibrations of the atoms and molecules in the solid materials from which it is made.

WOW

Kinetic energy killed the dinosaurs

The mass extinction event that we believe occurred in Mexico's Yucatan peninsula 65 million years ago was huge. A trillion tonne, 10–15 kilometre wide asteroid hit Earth at 20000 $m s^{-1}$. The energy dissipated as heat and rock movement was about 10^{23} joules, a billion times more than the first atomic bomb.

Smoke from fires and dust thrown up by the impact darkened and dramatically reduced solar energy input. Photosynthesis ceased. Food chains were disrupted and destroyed. More than half of all species on the planet, including dinosaurs, became extinct.



Science Photo Library/ David A. Hardy, Futures: 50 years in space

Figure 1.7 ◀ An artist's impression of the asteroid collision that is thought to have killed the dinosaurs 65 million years ago

Potential energy

When you stretch an elastic band you do work on it and store energy in it. The rubber band now has the potential to do work. It has stored the energy. When the rubber band is released it transforms this stored energy to kinetic energy.

Potential energy is stored in the way the particles are connected to each other. This is energy associated with the change in position of the particles. The 'springs' (bonds) between the atoms that make up the rubber band have been stretched – a change of position of the atoms.

iStockphoto/cotesebastien

◀ Figure 1.8
Potential energy is stored in the stretched elastic bungee cord due to the change in position of the atoms from which it is made.

Internal energy

If a solid body is heated its temperature increases. The particles will gain kinetic energy and, on average, vibrate faster. The particles move apart and the 'springs' store more energy. The sum of kinetic energy and potential energy is the internal energy.

Internal energy = kinetic energy + potential energy

At melting point, there is a **phase change**. The kinetic energy of the particles does not change any more until the phase change is completed. However, the 'springs' are affected and the particles become further separated. At phase change, the energy input is used to increase the distance between particles, not their kinetic energy.

THERMO RUBBER BAND HEAT ENGINE

Find out why rubber bands warm and cool when stretching and shrinking.

QUESTION SET 1.1

Remembering

1 List the assumptions of the kinetic particle model.

2 What is the difference between energy transformation and energy transfer?

Understanding

3 Explain why diffusion occurs faster in a gas than in a liquid.

4 Why are measurements using a thermometer considered objective and the sensations in your hands subjective?

5 Explain how kinetic energy, potential energy and internal energy are related.

Applying

6 What types of kinetic energies do the bodies have in the following situations?

a A freefalling, spinning ball

b A stationary block of ice

Analysing

7 Use the kinetic particle model to explain the increase in volume of steam, compared to the same volume of liquid water.

Reflecting

8 What do you now understand about how science uses models to explain observations?

Temperature explained

A cup of water takes a much shorter time to come to a boil (100°C) than a saucepan of water. The final temperature is the same for both, but the larger mass of water requires more heat to bring it to the same temperature. However, all the particles in the cup have, on average, the same kinetic energy as the particles in the saucepan. Thus, **temperature** is, to a good first approximation, a measure of the average random kinetic energy of the particles making up a body.

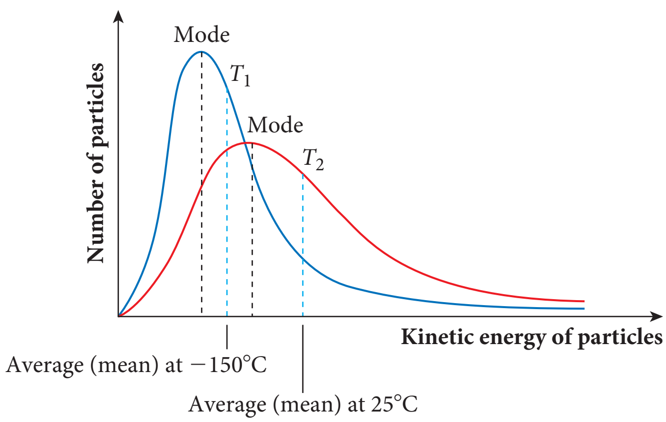
Temperature is a measure of the average kinetic energy of the particles in a body.

Temperature variation and kinetic energy distribution

When you heat a material the average kinetic energy of the particles increases. Figure 1.9 shows the wide range of kinetic energies of particles in the same mass of iron at two different temperatures. The peak of the curve is the average kinetic energy of all the particles.

When heat energy is added the proportion of atoms vibrating faster increases. The average kinetic energy of particles, therefore the temperature, increases.

Figure 1.9 ▶ The graph shows the energy distribution of the particles in a sample of iron at two different temperatures, T_1 (−150°C) and T_2 (25°C).



How temperature is measured

A thermometer has two elements, the temperature sensor system and a scale. A mercury thermometer makes use of the known expansion of mercury. For the Celsius scale, the length of the mercury column is marked 0°C and 100°C, the freezing point and the boiling point respectively of water. The Celsius scale is then divided into 100 equal divisions of 1°C.

Absolute zero, −273.15°C, or zero kelvin (0 K), occurs when the particles have no kinetic energy. The kelvin or absolute scale of temperature has one fixed point, absolute zero. An interval of one-degree Celsius is defined as 1 K. This means, for example, that the boiling point of water is 100°C on the Celsius scale and 373 K in the kelvin scale. Most countries have abandoned the Fahrenheit scale, although it is still common in the United States. It is not usually used in scientific writing.

Properties of materials can vary with temperature

Some materials change in predictable ways as temperature changes. Table 1.1 shows some thermometers that rely on changes to length, electrical resistance or colour. For example, a thermostat relies on the change of electrical resistance to measure temperature in an oven.

Table 1.1 Physical properties used by thermometers to indicate temperature change

Type of thermometer	Description
Mercury in glass	Uses different coefficients of expansion between mercury and glass
Thermocouple	Uses different temperature-dependent electrical properties of different metals that are brought into contact
Thermostat	Uses variation in electrical resistivity of a material with temperature
Thermal paint	Uses colour change with temperature
Bimetallic strip	Uses variation in coefficients of expansion between two different metals to detect temperature changes
Infrared	Uses the electromagnetic radiation radiated from a surface to measure temperature on the absolute temperature scale
Digital	Uses the variation in resistivity of a material with temperature; the greater the resistance the lower the current

Thermal expansion and contraction of materials is particularly important. It provides challenges when designing bridges, buildings, aircraft and spacecraft. Other situations in which thermal expansion has an effect include the following:

- Telephone wires are hung up slack in hot summer weather so that they do not pull the telegraph poles over when they contract in winter.
- Concrete roads are laid in sections with soft pitch between the sections.
- Glass to be used in cooking has to be a low expansion type such as Pyrex. Otherwise, it would shatter when heated.
- High-speed planes are warmed by air friction and become longer.
- Rocks shatter and eventually turn to sand.

▼ Figure 1.10 Railway tracks buckle due to expansion during a heat wave.

Newspix/Calum Robertson

Table 1.2 Interesting temperatures on the kelvin and Celsius scales

Interesting temperatures	Temperature	
	K	°C
Absolute zero	0	−273.15
The average temperature of space	2.725	−270.425
Helium liquefies	4	−269
Oxygen liquefies	190	−183
The lowest recorded surface air temperature on Earth (Soviet Vostok Station in Antarctica, on 21 July 1983)	184	−89
Average surface air temperature on Earth	288	15
Normal human body temperature	310	37
Highest recorded surface air temperature on Earth (Furnace Creek Ranch in Death Valley California, USA)	331	58
Temperature of the surface of the Sun	5778	5505
Titanium melts	1941	1668
Temperature in the core of the Sun	15.7×10^6	15.7×10^6

Trains make their characteristic 'clackety-clack' noise as the wheels go over the expansion gaps between short sections of rail. The railway line between Alice Springs and Adelaide comprises 830 km of continuous, welded rail on concrete sleepers. The rails slide along the concrete sleepers as they expand and contract. No gaps, no 'clackety-clack'.

WOW

Temperature of space

The measured temperature of space is 2.72548 ± 0.00057 K above absolute zero. That's very precise! It is this precision, and its consistency with predictions, that strengthen support for the Big Bang theory. A series of probes, COBE launched in 1989, WMAP (2001) and Planck (2009), have refined temperature data to very high levels of accuracy.

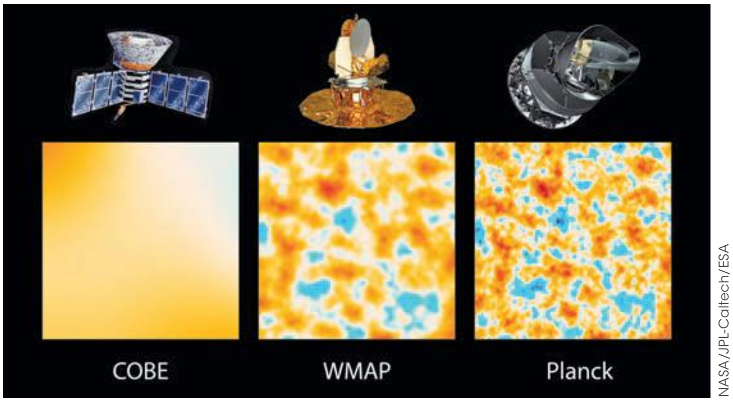


Figure 1.11 ◀ Images mapped by the probes show the increasing sensitivity of the measurements of the tiny variations in the temperature of space.

THE WILKINSON MICROWAVE ANISOTROPY PROBE (WMAP)

Learn about WMAP, a 2001 NASA Explorer mission that makes fundamental cosmological measurements.

Thermal equilibrium

If you place a hot stone in a container of cold water, heat energy is transferred from the hot stone to the cold water and container. The heat lost by the stone is equal to the heat gained by the water and the container. The stone gets cooler and the water and its container get warmer. This transfer will continue until both reach the same temperature. They are now said to be in **thermal equilibrium**. Collisions still occur between the water particles and the stone particles. The amount of heat going from the water into the stone exactly balances the amount of heat leaving the stone and going into the water. There is no longer a net transfer of energy from one to the other.

When two substances at different temperatures are mixed, the heat lost by one is equal to the heat gained by the other.

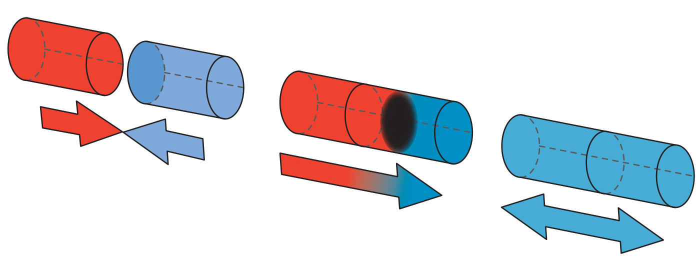
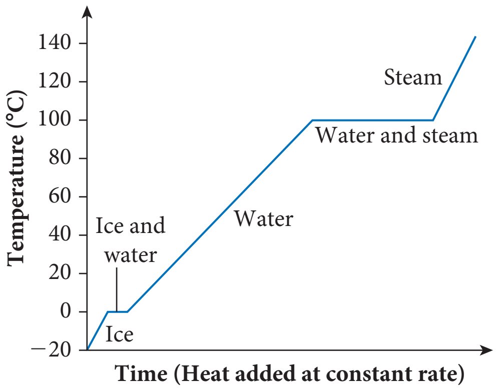


Figure 1.12 ▶ Hot and cold objects reach thermal equilibrium due to the transfer of energy by particle collisions. Eventually, the average kinetic energy of particles in both objects becomes the same.

Heating and cooling curves

When a pure substance is heated, its temperature is directly proportional to the energy transferred from a source provided that (a) it does not change phase, (b) the mass is constant and (c) the energy input rate is constant.

Heating and cooling curves always take the basic shape shown in Figure 1.13. The bottom left of the curve shows ice heating up to its melting point at 0°C. The curve becomes level because, as the ice melts, there is no change in temperature. The energy goes into the bonds between particles, not the kinetic energy of the particles.



◀ **Figure 1.13**
Heating curve for water

When all the ice has melted the water continues to heat until it reaches 100°C. At 100°C, it starts to boil. The curve remains at that temperature until all the water turns to steam. The steam temperature then rises.

See page 21 'Latent heat'.

QUESTION SET 1.2

Remembering

1 Define 'temperature'.

2 What is the lowest possible temperature on the kelvin scale? How is this temperature defined?

Understanding

3 A person with a fever has a temperature of 40.7 degrees. What temperature scale was used?

4 Describe how the internal energy of a substance could be increased.

Applying

5 a Write an equation that enables you to convert the Celsius scale into the kelvin scale.

b Sea water freezes at about 271 K. What is its freezing point on the Celsius scale?

Analysing

6 Why is the mode (peak of the graph in Figure 1.9) higher for $T_1 = -150°C$ than for $T_2 = 25°C$?

7 Why do most railway lines have gaps between each section of line?

Reflecting

8 What have you learnt about how a specific property of a material can be used to make a thermometer?

9 What have you learnt about the relationship between heat, energy and temperature?

INVESTIGATION 1.1

BIMETALLIC STRIPS AND THERMOMETERS

You are to investigate how a bimetallic strip behaves when heated. A bimetallic strip consists of two different metals that have been welded together.

◀ **Figure 1.14**
Bimetallic strip

You are then to use that information to design and build a working thermometer. Your design must be able to give quantitative temperature readings when placed in the oven. Your teacher will provide you with the equipment and materials you will need. You will then write a report to be submitted to your teacher.

What is your aim?

To build a working bimetallic strip thermometer. To do this, you will need to investigate the thermal expansion of various metals to understand how this works.

What will you need?

- metals to be investigated (your teacher may have some recommendations based on what is available at your school)
- water baths (0, 50 and 90°C)
- metal clamps and tongs
- thermometer
- protective mats
- access to an oven and/or a freezer

See page 17 for an example of how to fill in a risk assessment box.

What are the risks?

Construct a table similar to the one below. Identify specific risks involved in the investigation and ways that you will manage the risks to avoid injuries or damage to equipment.

What are the risks in doing this investigation?	How can you manage these risks to stay safe?

How will you carry out your investigation?

Write a procedure using point form for how you intend to carry out your investigation.

What results will you collect?

Think about the types of results you will be collecting. Remember that your measurements will be objective.

How will you analyse your results?

Think about how you are going to display your results. Will you use a table or a graph? What techniques will you use to analyse your results to find out what your results are telling you?

What have you found?

Your report is to contain the following.

1 A title that communicates what the investigation is about
2 An introduction describing what you did to discover the thermal properties of the bimetallic strip
3 A description of how you will use the discovered properties to make your thermometer
4 A description of how you created your temperature scale
5 A labelled scale drawing of your thermometer
6 A data chart that contains the data you obtained
7 A graph of your data
8 An analysis of the temperature range and accuracy of your thermometer
9 Answers to the questions below

For more information on how to write a scientific report see Chapter 13.

Ideas for improvement or further investigation

1 How could you improve the temperature range of your thermometer?
2 How could you improve the accuracy of your scale?
3 Is your scale linear? Explain your answer.
4 Would your thermometer be as good as a glass-mercury bulb thermometer to measure daily temperature differences? Explain your reasons.

Specific heat capacity

To boil water, a 1000 W electric kettle transfers 1000 J of energy every second for quite a long time. It takes a large amount of energy to raise the temperature of water. The same mass of cooking oil at the same starting temperature would take about half the time to reach 100°C in the same kettle.

The amount of energy required to raise the temperature of 1 kg of a substance by 1°C is the **specific heat capacity** of that substance. This is a physical property of the material and is related to its structure. Water has a high specific heat capacity. Cooking oil has a much lower specific heat capacity. Oil heats up and cools down almost twice as quickly as water.

Specific heat capacity of a substance is the amount of energy required to increase the temperature of one kilogram by one degree Celsius (or kelvin) without a change of phase.

Unit: $J\,kg^{-1}K^{-1}$ or $J\,kg^{-1}{}^{\circ}C^{-1}$

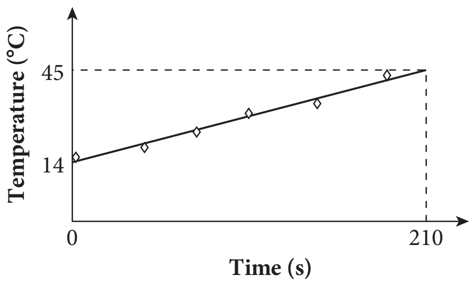
Remember: 1°C = 1 K

Investigating specific heat

A student completed an investigation by heating 1 kg of an unknown liquid X by adding heat energy to it at a steady rate of $150\,J\,s^{-1}$ for 210 s. The temperature was measured at regular intervals during the heating process and the data recorded was plotted as shown in Figure 1.15.

The graph shows that temperature T of a body (**independent variable**) goes up in direct proportion to the amount of heat energy ΔQ (**dependent variable**) put into the liquid:

$$\Delta T \propto \Delta Q$$

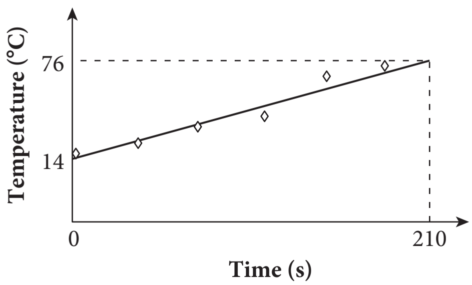


▲ Figure 1.15
The graph shows that a change in temperature of the liquid is directly proportional to the amount of energy put in.

The experiment was then repeated using only 0.5 kg of the unknown liquid. All other conditions were kept the same as the first experiment. The data were recorded and plotted as shown in Figure 1.16.

The second graph shows that for the same energy input, *half* the mass increases its temperature (independent variable) by *twice* as much. Therefore the change in the temperature of the body is inversely proportional to the mass of the body (a dependent variable):

$$\Delta T \propto \frac{1}{m}$$



▲ Figure 1.16
The graph shows that for the same energy input into half the mass the temperature increase is doubled.

Putting these two findings together gives us the relationship expressed below:

$$\Delta T \propto \frac{\Delta Q}{m}$$

There is always a constant, c, that makes a proportionality an equality, so by rearranging you get:

$$c\Delta T = \frac{\Delta Q}{m}$$

$$c = \frac{\Delta Q}{m\Delta T}$$

The constant c is the specific heat capacity of the substance that is being heated; ΔQ is the quantity of energy supplied; m is the mass of the body being heated and ΔT is the change in temperature. The units of the specific heat capacity can be found by substitution of the units into the formula:

$$\text{Units of } c = \frac{\text{J}}{\text{kg K}} = \text{J kg}^{-1}\,\text{K}^{-1}$$

The relationships are then expressed in their simplest algebraic form as shown below:

$$\Delta Q = mc\Delta T$$

This is a good example of how careful experimentation provides useful data to find meaningful relationships (formulas). These relationships can then be used to predict what will happen under a different set of given conditions.

The specific heat capacity of water

Water plays an integral part in evolution and the sustaining of life. It has the highest specific heat capacity of most commonly occurring substances: 4200 J kg^{-1}°C^{-1}. Water (a) heats up more slowly, (b) cools down more slowly and (c) stores more internal energy than the same mass of most other substances.

Many cooling and heating systems, from hot water bottles and hydronic heaters to water-cooled engines, utilise water's high specific heat capacity. Large bodies of water, such as oceans, seas and lakes, absorb large amounts of energy with only small temperature changes than water. For the same amount of energy input, landmasses have much greater temperature changes. Thus, the temperatures inland are much hotter than on islands and in coastal regions. During the warmer months when the sea temperature is less than the average air temperature, the sea acts as a **heat sink**. It stores energy. During the colder months when sea temperature is warmer than the average air temperature, it releases the stored energy. This release of energy moderates the temperature of regions close to large bodies of water.

Table 1.3 Specific heat capacity of some common substances

Substance	Specific heat capacity (J kg^{-1} K^{-1})
Water	4200
Ethylene glycol (antifreeze)	2400
Cooking oil	2800
Ice	2100
Steam	2000
Air	1000
Aluminium	900
Soil	800
Crown glass	670
Iron	450
Copper	380
Lead	130

EXPERIMENT 1.1

SPECIFIC HEAT CAPACITY OF METALS

Experiments have demonstrated that a small metal block will take about 3–5 minutes to come to thermal equilibrium with boiling water. Once it has reached this temperature it can be placed in a known mass of water (specific heat capacity = 4200 J kg^{-1}°C^{-1}) at a different temperature. Left for long enough in a calorimeter, the two will reach thermal equilibrium. From this and the mass of the metal block, the specific heat capacity of the metal can be calculated.

Aim

To find the specific heat capacity of one or more metals

Materials

- calorimeter
- thermometer or calibrated temperature probe
- heating equipment
- glass stirring rod
- electronic scales
- different metal cubes with dimensions about 2 cm × 2 cm × 2 cm
- strong cotton thread
- paper towel

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
Heating equipment can cause burns.	Avoid touching the equipment. Wait for the equipment to cool before you put it away.
It is possible to lose control of the hot block while transferring it from beaker to calorimeter or to burn yourself while doing so.	Double-check that the cotton thread is secured. Avoid touching the metal while transferring the block.
Boiling water can scald.	Wear safety glasses. Lower the block gently into the water. Avoid spilling or splashing boiling water.

Add any other risks that you can think of and ways to manage them.

Procedure

1 Set up the equipment as shown in Figure 1.17.

2 Heat the water until it is boiling.

3 Determine the mass of the cube and securely tie the cotton thread around it.

4 Gently lower the metal cube into the boiling water and leave until it is at 100°C (3–5 minutes).

5 Measure and record the mass of the calorimeter.

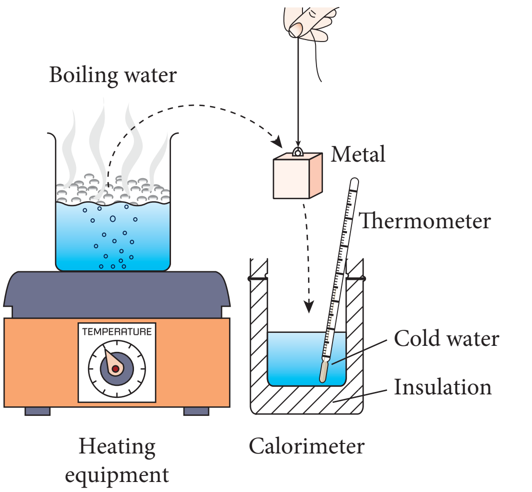
6 Add approximately 150 mL of cold water to the calorimeter and measure and record the mass of calorimeter and water.

7 Suspend the thermometer or temperature probe in the cold water.

8 Gently stir the water with the stirring rod and wait for the temperature of water and thermometer or temperature probe to come to equilibrium. Record this temperature.

9 Carefully lift the hot metal cube out of the boiling water, quickly dry it, then lower it gently into the calorimeter water. Stir the water gently and frequently.

10 Record the temperature of the mixture of the metal block and the water when it reaches its maximum.



▲ Figure 1.17
Experimental set-up for the transfer of the hot metal to the cold water

11 Repeat the experiment with a second trial.

12 If your teacher directs you to, repeat the experiment with a different metal.

Results

Record results in a table similar to the one below. Include an estimate of the uncertainty in each measurement (see page 411).

Data	Trial 1	Trial 2
Mass of metal block (g)		
Mass of calorimeter (g)		
Mass of cold water and calorimeter (g)		
Mass of cold water (g)		
Initial temperature of cold water and calorimeter (°C)		
Initial temperature of metal cube (°C)		
Final temperature of water and metal cube (°C)		

Analysis of results

1. Use the data to find the specific heat capacity of the metal for both trials.
2. Use the data to determine the measurement value and the estimate of the uncertainty.
3. Look up the accepted value of the specific heat capacity of the metal, including the uncertainty associated with this value. Decide whether the range of your measurement value overlaps the range of the accepted value.

Discussion

1. Why were you instructed to dry the metal cube before placing it in the cold water?
2. Why is it desirable to start with the water temperature below room temperature and have a final temperature above room temperature?
3. Why were you asked to do two trials? Does this improve accuracy or precision?
4. Did your best estimate of the specific heat capacity of the metal differ from the accepted value? Explain.
5. Is it meaningful to calculate the percentage error in this experiment? Explain.

Conclusion

Write a conclusion to answer your aim.

WORKED EXAMPLE 1.1

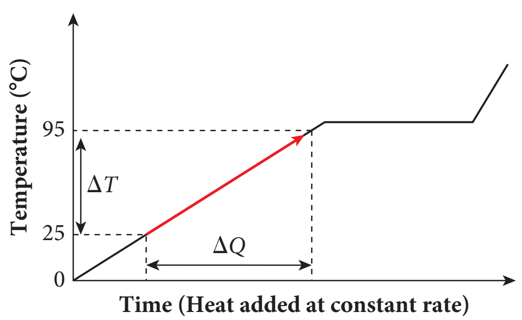
250 mL of pure water at 25°C is heated to 95°C.

a Sketch a heating curve for the water from 0°C to 100°C. Show on the graph the section relevant to this question. (2 marks)

b How much energy is needed to achieve this temperature change? (4 marks)

Answers

a



Logic

Draw correct graph and show section. 2 marks

b	$\Delta Q = mc\Delta T$	Find energy by this formula.	1 mark
	$\Delta Q = 0.250\,\text{kg} \times 4200\,\text{J kg}^{-1}\,°\text{C}^{-1} \times (95°\text{C} - 25°\text{C})$	Substitute known variables into the formula. Find *c* in list of specific heat constants.	2 marks
	$\Delta Q = 7.35 \times 10^4\,\text{J}$	Calculate the correct answer.	1 mark

Try these yourself

1 A sample of an unknown substance of mass of 505 g is heated from 21.0°C to 56.0°C. The energy required was 4.90×10^4 J. Calculate the specific heat of the substance. Use Table 1.3 to identify it. (4 marks)

2 A pure iron nail of unknown mass requires 860 J of energy to change from 23.0°C to 305.0°C What is the mass of the nail? (3 marks)

WORKED EXAMPLE 1.2

A nurse prepares a bath that needs to be at 41°C for a patient. First he adds 53 L of water at 23°C from the cold tap.

a What four assumptions must be made to find the solution? (2 marks)

b How much water did he need to add from the hot tap at 68°C for him to achieve the required of temperature of 41°C? (5 marks)

Answers	**Logic**	
a Assumptions: **1** No energy is lost to the surroundings such as the taps, the air and the bath. **2** The mixing process does not add energy to the water. **3** The water is pure. **4** 1 L of water has a mass of 1 kg.	Give the correct assumptions.	2 marks
b $\Delta Q_{\text{hot water}}\text{ lost} = \Delta Q_{\text{cold water}}\text{ gained}$	Use the equation heat lost equals heat gained.	1 mark
$-m_{\text{hot}}c\Delta T_{\text{hot}} = m_{\text{cold}}c\Delta T_{\text{cold}}$ $-m_{\text{hot}} = \dfrac{m_{\text{cold}}c\Delta T_{\text{cold}}}{c\Delta T_{\text{hot}}}$	Expand the relationship and make *m* the subject, then cancel *c*.	2 marks
$-m_{\text{hot}} = \dfrac{m_{\text{cold}}\Delta T_{\text{cold}}}{\Delta T_{\text{hot}}}$ $-m_{\text{hot}} = \dfrac{53\,\text{kg} \times (41°\text{C} - 23°\text{C})}{41°\text{C} - 68°\text{C}}\,\text{kg}$	Put the known variables into the formula.	1 mark
$m_{\text{hot}} = 35.3\,\text{kg}$ 35.3 L of hot water was added.	Calculate the answer.	1 mark

Try these yourself

A hot iron barbecue plate of mass 1.20 kg is placed in a tub that contains 22.2 L of cold water at 19.3°C. The water and the plate reach a final equilibrium temperature of 21.3°C.

a What assumptions are you going to make before you start solving the problem? (2 marks)

b What was the original temperature of the hot barbecue plate? (3 marks)

QUESTION SET 1.3

Remembering

1 What does it mean when we say the specific heat capacity of iron is 450 kJ kg^{-1} K^{-1}?

Understanding

2 Why do the oceans have a modifying effect on the temperature of the land near them?

3 Use the formula $\Delta Q = mc\Delta T$ to derive the units of specific heat capacity.

Applying

4 A 5.0 L tub contains 3.0 L of water at 23°C. The tub is then filled to the top with hot water. The final temperature of the water in the tub is 26°C. What was the temperature of the water added?

5 A 60 kg bushwalker is suffering from hypothermia, having an average body temperature of 33.5°C. When rescued, she was wrapped in a blanket and given two 310 mL cups of warm tea each at a temperature of 60°C. Calculate the maximum rise in the bushwalker's body temperature due to the heat of the tea. ($c_{\text{human body}} = 3.5 \times 10^3$ J kg^{-1} K^{-1})

6 A cook pours 0.80 kg of soup at 98.0°C into a 1.00 kg vacuum flask (specific heat capacity of 32.0 J kg^{-1} K^{-1}.) When soup and flask reach thermal equilibrium, the temperature of the flask has gone from 10.0°C to 97.0°C. Calculate the specific heat capacity of the soup.

Analysing

7 The specific heat capacity of a watermelon is very large, almost as large as that of water. Why do you think that is?

8 Equal masses of lead and water are in contact and at thermal equilibrium. Equal amounts of heat are now transferred to the water and the lead. Will they be at thermal equilibrium immediately after the transfer of energy? Explain in detail.

Reflecting

9 Why is water's high specific heat capacity important for its use in heating and cooling systems?

10 How has your understanding of the kinetic particle model helped you understand thermal equilibrium and temperature?

Conservation of energy

In physics, a closed system is one that matter cannot enter or leave, but energy can be transferred into or out of the system by work or heat. An isolated system is one that neither energy nor matter can enter or leave.

An **isolated system** is one in which no energy or matter can enter or leave. Therefore, within an isolated system, the total amount of energy remains constant. Energy can be converted from one form to another. Energy can be transferred from place to place. But the total energy within the isolated system remains constant.

Except for the universe, there is no such thing as a perfect isolated system. We can carefully insulate a system from its surroundings, but some energy is always transferred in or out of the system. Very small losses of energy occur even in the most insulated systems (think about your calorimeter experiments). As long as the losses are negligible (losses that are too small to have any have significant influence) we can model the system as isolated.

Energy transfer occurs when energy flows from one object or substance to another object or substance. Energy transformation occurs when energy changes from one form to another. This transformation can occur within a single object or when energy is being transferred from one object to another. For example, when you heat one end of a metal rod by putting it in a flame, heat is transferred from the flame. Soon the other end of the rod gets hotter.

If you turn on a torch, chemical energy is transformed into electrical energy. The electrical energy is transformed into light and heat energy. In this case, the battery converts the stored chemical energy into electrical energy. A globe or a **LED** converts the electrical energy into light and heat energy.

The **first law of thermodynamics** (the law of conservation of energy) states that in an isolated system, energy can neither be created nor destroyed. Energy can be transferred or transformed but the total energy of an isolated system remains constant. The total change of energy is zero.

Some substances can change state (phase) directly from a solid to a gas (sublimation) or a gas to a solid (deposition), without going through the liquid state. Solid carbon dioxide (dry ice) does this at −78.5°C. It is primarily used for cooling that does not leave a residue.

State changes and latent heat and power

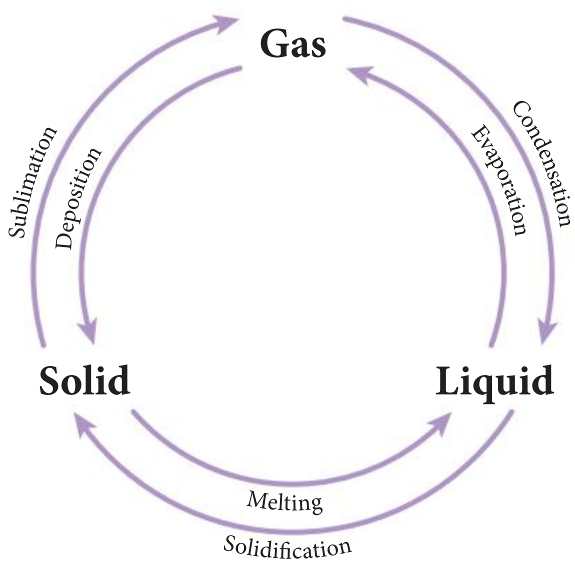
A pure solid starts to change state to a liquid at its **melting point**. A pure liquid starts to change state to a gas at its **boiling point**. Both processes, **melting** and **vaporisation** respectively, require energy input. Energy removal causes gases to undergo **condensation** and liquids, **solidification**.

Evaporation and vaporisation are often confused. Below the boiling point, evaporation from a liquid occurs at the surface. Some particles with high kinetic energy escape.

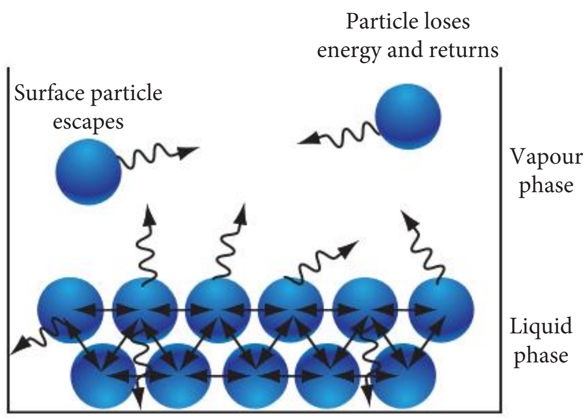
Vaporisation occurs when the liquid changes to gas. Bubbles form below the surface. No temperature change occurs during vaporisation.

CHANGING PHASE

Watch a simulation of matter changing phase.



▲ Figure 1.18
State change cycles



▲ Figure 1.19
Evaporation occurs at the surface when water molecules that are less tightly bound and have relatively higher kinetic energy than those in the body of the water escape.

Latent heat

During a change of state, energy is required to be added or removed. The energy added during a state change is called **latent heat**. The **specific latent heat of fusion** of a substance is the energy required to change the state of 1 kg of the substance from its solid state to its liquid state without any change in temperature. The specific latent heat of melting for water is 334 kJ kg^{-1}. In its solid state the water particles are held tightly together. To pull them apart requires a large amount of energy. This energy to separate the particles is supplied externally. It does not increase the kinetic energy of the particles; it is used to increase their average separation. As a result, only the internal energy increases. The change in internal energy cannot be measured by a thermometer because there is no increase in the average kinetic energy.

The **specific latent heat of vaporisation** of a substance is the heat required to change the state of 1 kg of the substance from its liquid to gaseous state. The specific latent heat of vaporisation of water is 2260 kJ kg^{-1}. You can see from the very large value that it requires a huge amount of energy to separate the particles from each other.

Figure 1.20 ▶
Distinctive anvil-shaped clouds are often seen at the tops of thunder clouds.


Alamy/PhotoStock-Israel/Ohad Shahar

Both these processes are reversible. For example, a quantity of steam loses energy to its surroundings, cooling until it reaches its boiling point. It then remains at a constant temperature while the water molecules get closer together. Energy is being released to the surroundings during the condensation process. This energy comes from a reduction of the internal energy of the molecules as they draw closer together to form liquid water. This is why steam at 100°C will cause much more severe burns than the same mass of water at 100°C.

Condensation and heat exchange occurs in cloud formation. A pocket of moist air surrounded by dry air rises because it is less dense. It ascends into a cooler region causing the water vapour to condense as clouds. The latent heat of vaporisation is released into the surrounding air, which becomes warmer. Warm air, being less dense than cooler air, continues to rise. Eventually, the moist air hits the 'roof' of the weather zone, the troposphere. Cloud formation then continues horizontally rather than vertically. Distinctive anvil-shaped clouds form at the tops of thunder clouds, especially in the tropics.

Investigating latent heat

A student completed an investigation by heating different masses of ice at 0°C that had been dried to remove any water in liquid form. The heat energy was supplied at a steady rate of 1000 J s^{-1} until the ice had just melted. The time was recorded and precautions were taken to minimise any external heat gains or losses. The data were recorded in Table 1.4. Figure 1.21 shows a graph of this data.

Table 1.4 Data from student investigation

Mass of ice at 0°C (g)	Time for melting (s)	Total energy input (kJ)
0.10	33	33
0.22	72	72
0.39	88	88
0.52	175	175
0.64	217	217
0.90	300	300

The graph indicates that there is a direct proportionality. In general:

$$\Delta Q \propto m$$

There is always a constant that makes a proportionality an equality; in this case L is used as the constant:

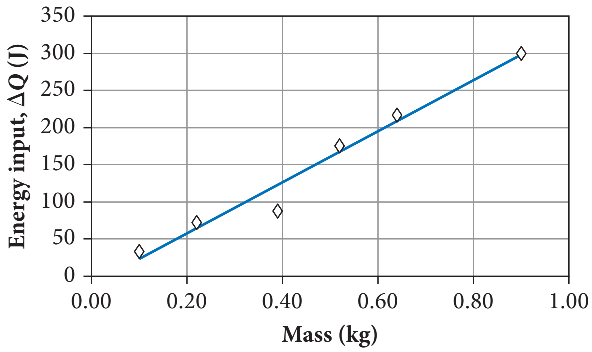
$$\Delta Q = Lm$$

Rearranging:

$$L = \frac{\Delta Q}{m}$$

L, the latent heat, is the gradient of the graph.

$$\text{Units of } L = \frac{\text{J}}{\text{kg}} = \text{J kg}^{-1}$$



▲ **Figure 1.21** Finding the relationship between energy input and mass when water melts. The equation of the line is $\Delta Q = m \times 334$.

This gives us the algebraic expression of the relationship between state changes and energy required to change the state:

$$\Delta Q = mL$$

Table 1.5 Latent heats of fusion and vaporisation for a number of common substances. Unlike specific heat, $\text{J kg}^{-1}\,{}^{\circ}\text{C}^{-1}$, latent heat is given in kJ kg^{-1}

Substance	Latent heat of fusion (kJ kg^{-1})	Latent heat of vaporisation (kJ kg^{-1})
Aluminium	390	10 500
Alcohol (ethanol)	105	841
Copper	205	4 800
Iron	276	6 340
Lead	25	860
Silver	105	2 350
Water	334	2 260

Specific latent heat (kJ kg^{-1}) is the energy required to change the state of 1 kg of a substance:

Specific latent heat of:	Change of state
Melting	Solid to liquid
Solidification/fusion	Liquid to solid
Vaporisation	Liquid to gas
Condensation	Gas to liquid

WORKED EXAMPLE 1.3

How much energy is required to melt 250 g of ice at 0°C to 250 mL of water at 0°C? (3 marks)

Answers	Logic	
$\Delta Q = mL$	Find the energy needed using this formula.	1 mark
$\Delta Q = 0.250\,\text{kg} \times 334\,\text{kJ kg}^{-1}$	Put the known variables into the formula. *L* is found in the table of latent heat constants (Table 1.5).	1 mark
$\Delta Q = 83.5\,\text{kJ}$ or $8.35 \times 10^4\,\text{J}$	Calculate the answer.	1 mark

Try these yourself

1 A sample of lead of mass of 505 g is heated to its melting point. How much heat energy is needed to completely change its phase to liquid lead at its melting point temperature? (Use the latent heats in Table 1.5.) (3 marks)

2 Different substances have different melting points and boiling points, but the shapes of their heating curves are very similar. The solid iron is heated constantly from its solid state to its gaseous state in a furnace. The heating curve for the duration of this process is shown in Figure 1.22. Use the curve to answer the following questions.

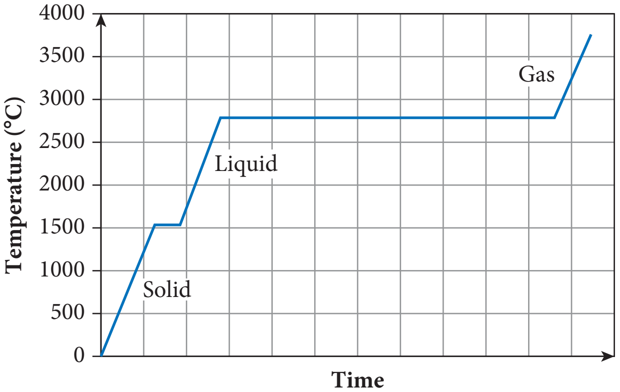


Figure 1.22 ◀ Heating curve for iron

- **a** What is the melting point of iron? (1 mark)
- **b** What is the boiling point of iron? (1 mark)
- **c** Explain why two sections of the graph are parallel to the time axis. (2 marks)
- **d** What does the length of the longest horizontal section of the graph tell you about the latent heat of vaporisation of iron? (2 marks)
- **e** Redraw Figure 1.22 to scale. On the same axes, draw the heating curve for iron when the heating rate was doubled. (2 marks)

INVESTIGATION 1.2

LATENT HEAT OF FUSION OF SOLIDS

Modify Experiment 1.1 'Specific heat capacity of metals', to find the latent heat of fusion of solids such as ice and other substances that melt at similar temperatures.

When crushed ice at 0°C is added to water it will melt. The temperature change in the water can be used to find the latent heat of fusion of ice. Explain how.

What is your aim?

Consider the range of possible substances, then write your aim. The aim should be concise, indicate what quantity is to be measured and provide information about the way the measurement is to be undertaken.

What will you need?

Use the resources suggested in Experiment 1.1 'Specific heat capacity of metals'. Do you need other resources?

What are the risks?

Construct a table similar to the one below. Identify specific risks involved in the investigation and ways that you will manage the risks to avoid injuries or damage to equipment.

What are the risks in doing this investigation?	How can you manage these risks to stay safe?

How will you carry out your investigation?

Consider the similarities and differences between Experiment 1.1 'Specific heat capacity of metals' and your investigation.

Construct a flowchart of the procedure you intend to follow. Refer to Chapter 13 for more information.

Submit your flowchart and risk assessment to your teacher for approval before beginning the work.

What results will you collect?

Show on your flowchart the data that you intend to collect.

How will you analyse your results?

Consider the value of using graphs and equations in your analysis. Justify your analytic processes.

What have you found?

Provide quantitative values for all latent heat of fusions measured. Ensure the uncertainties claimed are justifiable.

What do you conclude?

Draw together all your data and make a conclusion that relates directly to the specific details of your aim.

Ideas for improvement or further investigation

How could you obtain more precise data?

What other substances could you study?

How would you refine the experiment to measure the specific latent heat of fusion of substances that, for a range of temperatures above 0°C, are gases or solids?

Solving problems that have state changes

Many real-life situations involve temperature and state changes occurring in a process that is being studied; for example, cooling down a hot barbecue plate with water or perspiring to remove excess heat from your body involve liquid to gas phase changes. To solve problems that involve multiple temperature and state changes, the temperature versus energy input graph becomes a very handy tool.

Water below 0°C is heated to its melting point, at which it changes state to liquid. The liquid phase heats until it reaches 100°C, at which it changes state to steam. Then the steam superheats.

SUPERHEATED STEAM

Watch steam light a match!

Different sections of the curve require different calculations to find the energy input needed to heat or change the temperature or state of ice, water or steam. Cooling requires the release to the surroundings of the same amounts of energy. Calculations depend on whether the water remains in its state or its state is changing.

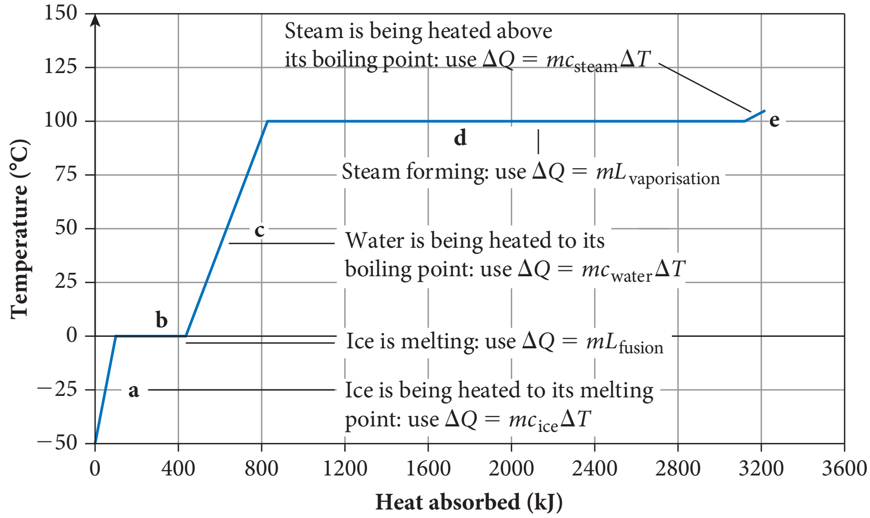


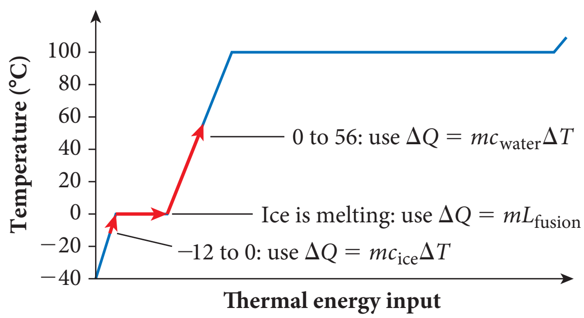
Figure 1.23 ▶ The heating curve for 1 kg of pure water

WORKED EXAMPLE 1.4

430.0 g of ice at −12°C is heated to water at 56°C.

a Sketch the heating curve for water and identify the section of the curve that the problem covers. (2 marks)

b How much thermal energy is required? (5 marks)

Answers	Logic	
a 	Use the heating curve to identify the parts of the solution.	2 marks
b $Q_{total} = mc_{ice}\Delta T + mL_{fusion} + mc_{water}\Delta T$	Find the energy using this composite formula.	2 marks
$Q_{total} = (0.430\text{ kg} \times 2.10 \times 10^3\text{ J kg}^{-1}\text{ °C}^{-1} \times 12\text{°C}) + (0.430\text{ kg} \times 3.34 \times 10^5\text{ J kg}^{-1}) + (0.430\text{ kg} \times 4.2 \times 10^3\text{ J kg}^{-1}\text{ °C}^{-1} \times 56\text{°C})$	Substitute the correct values and the correct units.	2 marks
$= 2.56 \times 10^5\text{ J} = 256\text{ kJ}$	Calculate the answer.	1 mark

Try this yourself

Copper has a melting point of 1084.62°C. How much energy does it take to completely melt 35.6 g of copper that is initially at 25.0°C? (5 marks)

WORKED EXAMPLE 1.5

1 A 420 g block of lead at 93°C is placed in a hole made in a large block of ice at 0.0°C.

a What is the most important assumption you must make to answer the question? (1 mark)

b How much of the ice will melt? (4 marks)

2 A 2.5 kg iron barbecue plate at 328°C is too hot for cooking. It needs to be cooled to 200°C.

a This is to be done by placing the plate on a block of ice. Show what happens to the ice on a sketch of the heating curve for water from below 0°C to above 100°C. (3 marks)

b How much ice at 0°C is needed to use to cool the barbecue plate to the required 200°C? (6 marks)

Answers | Logic

1 a From the phrase 'in a large block of ice', the most important assumption is that all the energy transferred to the ice from the lead only melts the ice.

Give the correct assumptions. — 1 mark

b Heat lost by lead = heat gained by the ice

ΔQ_{Pb} lost = ΔQ_{ice} gained

Use the correct equation. — 1 mark

$m_{Pb}c_{Pb}\Delta T = m_{ice}L_{fusion}$

$$m_{ice} = \frac{m_{Pb}c_{Pb}\Delta T}{L_{fusion}}$$

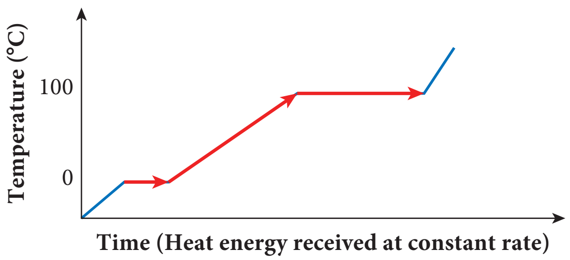
$$m_{ice} = \frac{0.42\text{ kg} \times 130\text{ J kg}^{-1}\text{ °C}^{-1} \times 93\text{°C}}{3.34 \times 10^5\text{ J kg}^{-1}}$$

Substitute the correct values and the correct units — 2 marks

$= 0.015$ kg or 15 g

Calculate the answer. — 1 mark

2 a — 1 mark



All the energy transferred to the ice from the plate goes into melting the ice, heating the water and converting it to steam.

Identify the correct line sections. — 2 marks

b Heat lost by iron = energy gained by the ice/water/steam

Use the correct equation. — 1 mark

$\Delta Q_{Fe} = \Delta Q_{ice\text{-}water} + \Delta Q_{water} + \Delta Q_{water\text{-}steam}$

$m_{Fe}c_{Fe}\Delta T = m_{ice}L_{fusion} + m_{water}c_{water}\Delta T + m_{water}L_{vaporisation}$

Mass of the ice, water and steam is the same.

$m_{Fe}c_{Fe}\Delta T = m_{ice/water/steam}(L_{fusion} + c_{water}\Delta T + L_{vaporisation})$

Complete the correct algebraic steps. — 2 marks

$$m_{ice/water/steam} = \frac{m_{Fe}c_{Fe}\Delta T}{L_{fusion} + c_{water}\Delta T + L_{vaporisatio}}$$

Substitute the correct values. — 2 marks

$$m_{ice/water/steam} = \frac{2.5\text{ kg} \times 450\text{ J kg}^{-1}\text{°C}^{-1} \times 128\text{°C}}{3.34 \times 10^5\text{ J kg}^{-1} + (4.2 \times 10^3\text{ J kg}^{-1}\text{ °C}^{-1} \times 100\text{°C}) + 2.26 \times 10^6\text{ J kg}^{-1}}$$

$m_{ice} = 4.78 \times 10^{-2}$ kg

Calculate the answer. — 1 mark

Try this yourself

240 g of copper filings are removed from boiling ethanol (boiling point 78°C) and placed in a light, insulating polystyrene cup containing 150 g of ice at 0°C. How much of the ice will melt? (6 marks)

Scientific literacy: Steam weeding

Flame weeding is a long established technology used in organic horticulture and agriculture and also by municipal authorities. While flame weeders are useful and productive tools they are not as efficient and effective as steam weeders.

Scientific research has shown that steam is superior to open flames for transferring heat into weeds, principally due to the large latent heat of condensation of water, which contains a large amount of energy, and is therefore more efficient at transferring heat into weeds. The steam weeder only needs to raise the temperature of the weeds to about 70–80°C to kill them.


Copyright © Steam Weeding Ltd

Figure 1.24 ▲
A 6-metre wide 1.2-megawatt steam weeder

Until recently it has not been practical to use steam for weed control (as opposed to soil steaming) in agricultural and horticultural settings. This is because the standard steam boiler design is heavy, expensive, needs regular maintenance and often has to comply with a range of safety legislation. In the last few years new approaches of steam generation have been used to develop commercial steam weeders for orchard and vineyard intra-row weed control.

Direct-fired steam weeders (DFSW) work by spraying water as a fine mist directly into the exhaust gas from a burner, causing it to flash into steam. This simple but highly effective approach has a number of advantages over standard flame weeder designs and existing steam weeders. These include:

- No open flames: the flame in the DFSW is contained in an insulated refractory chamber, the steam–hot gas mix it produces has a lower temperature than an open flame and contains little oxygen, which means it would be very difficult for the DFSW to start a fire;
- The lower temperatures and low fire risk means that the DFSW can be used over polythene and paper based mulches;
- A wide range of fuels can be used, including diesel, which is available on practically all farms and even bio-fuels such as those made from rapeseed oil (canola) or even used chip shop oil;
- The separation of the steam production unit from how and where steam is applied creates huge flexibility. For example, it could be used under-vine, across a whole vegetable bed, or applied in strips over crop rows;
- A wide range of heat outputs can be catered for – e.g. from 100 to 600 kW per steam generator unit – allowing accurate matching of heat output to the job in hand. Individual steam generator units can even have two, easily switchable heat outputs, e.g. 200 or 500 kilowatts;
- As the burner and steam production are protected within combustion chambers they are totally unaffected by wind or other weather conditions, and, depending on the application method, e.g. closed hood over a vegetable bed, the application of the steam will also be unaffected by weather conditions;
- Direct-fired steam avoids the use of heat exchangers containing superheated water, so reducing complexity, cost, and increasing safety, as there is no potential for a superheated steam explosion.

Non-chemical weed management for commercial farmers and growers is becoming increasingly important now that significant problems are emerging with herbicides, especially herbicide resistance: physical weeding is the plan B for a post-herbicide world.

Questions

1. How is the steam produced by direct-fired steam weeders?
2. Why is steam a very effective way to transfer heat energy to the weeds?
3. What happens to the heat energy in the exhaust gas when they are used to 'flash' the fine mist of water into steam?
4. Why is this method of weeding, ecologically sound?
5. Why can the steam weeder but not the flame weeder be used on polystyrene sheeting and paper based mulches?
6. What evidence can you find in the article that the development of the steam weeder addresses economic considerations?

Case study

Professor David Mee

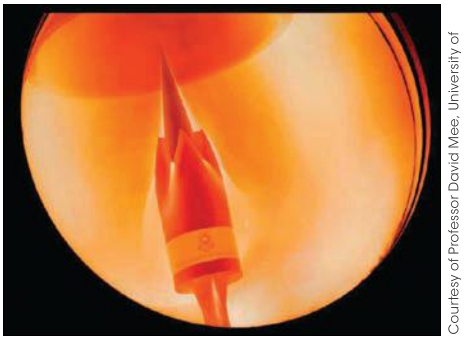
Professor David Mee is the Head of the School of Mechanical and Mining Engineering at the University of Queensland. His research fields include hypersonic aerodynamics and thermodynamics. Professor Mee developed an interest in aerodynamics and thermodynamics while studying to become a mechanical engineer. An experiment he conducted in one of his undergraduate courses to investigate whether hydrogen could burn in a supersonic (faster than the speed of sound) airstream sparked his interest in the analysis of combustion. This set him on a path that led to research on Rolls Royce gas turbine engines in the UK and, eventually, a return to Australia to help develop scramjet technology. This technology can be used to propel aircraft at speeds up to ten times the speed of sound.

A scramjet is an engine that 'breathes' air, just like a car engine or a jet engine. A car engine takes petrol from the fuel tank and mixes it with air drawn in through an air filter. The air–fuel mixture is then fed into the cylinders where it is compressed and then ignited. The fuel burns using the oxygen in the air. The energy released drives the pistons. The power delivered by the engine is then transferred to the wheels to propel the car.

Courtesy of Professor David Mee, University of Queensland

▲ **Figure 1.25**
Professor David Mee, University of Queensland

A jet engine is similar but it uses a compressor to draw air into the engine, raise its pressure and temperature and slow the flow down. Fuel is then injected into the combustion chamber where the flow is travelling much slower than the speed of sound (subsonically). The fuel burns in the air and releases its energy. This energy is used to drive a turbine, the sole purpose of which is to drive the compressor. The products of the combustion are then expanded downstream of the turbine and exhausted at high speed to propel the jet engine forward.



Courtesy of Professor David Mee, University of Queensland

▲ **Figure 1.26**
A scramjet engine in the T4 Shock Tunnel at University of Queensland

When an aircraft travels very fast, just slowing the flow down raises its pressure and temperature. The faster you fly, the more the pressure and temperature increase when you slow the flow. If you are flying fast enough, such as a few times the speed of sound, just slowing the air to subsonic speeds can give temperatures and pressures high enough for good combustion. So unlike a jet engine there is no need for a compressor and therefore a turbine. A specially designed intake will slow the air flow. Fuel is then injected into the subsonic flow in the combustion chamber. The fuel burns and resultant gas particles expand rapidly at high speeds. The fast moving hot gas particles are ejected through a nozzle providing a powerful thrust that propels the vehicle. Such an engine is called a ramjet (a subsonic combustion ramjet).

Professor Mee explained that when you are flying faster than about five times the speed of sound, just slowing the flow down a little but keeping it supersonic lets you reach good conditions for combustion. So, if you can inject fuel into a supersonic air stream in the combustion chamber and get it to burn and release its energy, you can expand the flow through a nozzle and gain a powerful thrust that propels the vehicle. This is called a scramjet (a supersonic combustion ramjet).

Scramjet-powered vehicles have the potential to be much lighter than rocket-powered vehicles. This is because a rocket has to carry both its fuel and its oxygen whereas a scramjet only has to carry its fuel – it can 'breathe' the oxygen in the air through which it flies. He and his fellow researchers in the field hope to develop the scramjet so it can be used as the stage of a launch vehicle to put satellites into space more economically than with rockets alone.

Questions

1 Define 'supersonic'.
2 What are the major similarities and differences between a scramjet and a petrol engine?
3 Give two reasons why the scramjet research is being funded?
4 Will the development and introduction of scramjet technology reduce the emissions of **greenhouse gases** if it is used to launch satellites? Explain.
5 Create a flowchart to show how a scramjet engine works.

QUESTION SET 1.4

Remembering

1 Define an 'isolated system'?

2 What does the law of conservation of energy state?

3 What is the key premise for solving method of mixtures problems?

Understanding

4 Why does the temperature of a liquid at its boiling point remain constant when it is changing state?

5 The line on a heating curve for a pure substance being heated between state changes has a constant linear slope. What does this imply about the heat supply?

6 The lines on a heating curve for a pure substance receiving a constant heat energy input during state changes are parallel to the horizontal axis. What does this imply?

Applying

7 Draw a flow diagram to show how visible light can be transformed and transferred to mechanical energy in two stages.

8 A solar hot water system contains 150 L of water at 23°C. How much solar energy is required to heat the water to 63°C if the solar hot water system is 75% efficient?

9 A 1.5 kg copper electric kettle supplies energy at a rate of 500 $J\,s^{-1}$. The kettle contains 250 g of water and the kettle and water are at a temperature of 24°C. If the system is 100% efficient how long will it take for all the water to boil away?

Analysing

10 A constant volume of hydrogen gas has a specific heat capacity of $1.43 \times 10^4\,J\,kg^{-1}\,K^{-1}$. Hydrogen gas is sometimes used to enclose electrical generators. What purpose do you think the gas may have?

11 A soft drink at room temperature is too warm to drink. Would it be best to add 50 g of water at 0°C or 50 g of ice at 0°C for the most effective cooling? Give a reason for your answer.

Reflecting

12 Explain why you found using heating and cooling graphs useful in solving complex heating and state change problems.

CHAPTER SUMMARY

- Kinetic particle model states that matter consists of particles that are in constant motion. Collisions between particles may result in the transfer of energy.
- Temperature measures the average kinetic energy of particles in a substance.
- Internal energy is all the kinetic energy and all the potential energy in a system.
- Absolute zero: the temperature at which all particles have no kinetic energy (0 K or −273.15°C).
- Thermal equilibrium occurs when two isolated systems at different temperatures come into contact and the final temperature is the same for both systems. At thermal equilibrium there is no net transfer of energy from system to system.
- Law of conservation of energy states that energy can be neither created nor destroyed. Energy can change forms, and energy can flow from one place to another. The total energy of an isolated system remains constant.
- Heat is the energy transferred from one region to another due to a difference in temperature.
- Specific heat capacity c is the energy required to raise the temperature 1 kg of a substance by 1°C. Different materials have different heat capacities:

  $\Delta Q = mc\Delta T$
- Latent heat capacity is the energy needed to cause 1 kg of a substance to change state at constant temperature. Different materials have different latent heat capacities:

  $\Delta Q = mL$

CHAPTER GLOSSARY

absolute zero the theoretical lowest possible temperature −273.15°C on the Celsius scale or 0 K on the absolute or kelvin scale

anthropogenic human derived; caused by human activity

average temperature a measure of the average kinetic energy of the matter in a defined system

boiling point the temperature at which a liquid changes state from liquid to gas

condensation the phase change from a gas to a liquid

core temperature the temperature of the internal organs in the chest cavity, abdominal region and head in animals

cryogenics the study of low-temperature phenomena

dependent variable the variable that changes as a result of changes to the independent or controlled variable

diffusion the spontaneous movement of substances from regions of high concentration to regions of low concentration

elastic collision a collision between two or more objects in which there is no loss of total kinetic energy

energy security the ability to acquire and protect national energy resources

evaporation the process in which some particles with high kinetic energy escape the surface of a liquid at a temperature below its boiling point

first law of thermodynamics in the universe energy can be neither created nor destroyed; however, energy can change forms and energy can flow from one place to another within the universe. The total energy of an isolated system remains constant

fluid a gas or a liquid

greenhouse gases any gas that traps heat in the atmosphere; the primary greenhouse gases are water vapour, carbon dioxide, methane, nitrous oxide and ozone

heat (n.) energy transferred due to a difference in temperature

heat, to heat (v.) the process of transferring energy due to a difference in temperature

heat sink an object or material that moderates the temperature of its surroundings due to its large specific heat capacity; also an electronic component used to transfer heat effectively from components to the surrounding air

independent variable a variable upon which another variable depends; also called the controlled variable

internal energy the sum of the kinetic energy of the particles in the system and the potential energy stored in the system

internal processes processes that occur within a defined system

isolated system a system that neither energy nor matter can enter or leave

joule (J) SI unit of energy, $1\,J = 1\,kg\,m^2\,s^{-2}$

kinetic energy the energy a body possesses due to its motion, it can be in the form of translational, rotational or vibrational energy

kinetic particle model explains the properties of the different states of matter. The particles in solids, liquids and gases have different amounts of energy, are arranged differently and move in different ways

latent heat the heat required to change the state of a substance at its melting or boiling point without a change in temperature; unit $J\,kg^{-1}$

law of conservation of energy in the universe energy remains constant; it cannot be created or destroyed

LED light-emitting diode

melting the phase change from a solid to a liquid

melting point the temperature at which a substance undergoes a phase change from solid to liquid (melts)

metabolic activity the set of internal chemical reactions that maintain an organism's life

objective measurement measurement that has a numerical value; the result of the measurement does not depend on the person taking the measurement

phase change a change of state (e.g. solid to liquid)

plasma a collection of free-moving electrons and ions that can be accelerated by magnetic and electric fields

potential energy energy that can be considered to be 'stored' within a body due to its position, composition or molecular arrangement

solidification phase change from liquid to solid

specific heat capacity the amount of energy required to increase the temperature of 1 kg of a substance by one degree Celsius (or kelvin). It can be thought of as the resistance of a material to an increase in temperature; unit $J\,kg^{-1}\,°C$ or $J\,kg^{-1}\,K$

specific latent heat of fusion the heat required to change the state of 1 kg of a specific substance at its melting point without a change in temperature

specific latent heat of vaporisation the heat required to change the state of 1 kg of a specific substance at boiling point without a change in temperature

subjective indication an estimate by our senses that depends on the person making the measurement

temperature a measure of the average kinetic energy of the particles in a sample of matter, expressed in terms of degrees designated on a standard scale

thermal equilibrium the condition under which two substances in physical contact with each other do not exchange heat energy; the two substances are at the same temperature

thermodynamics the physics of heating and cooling phenomena and their explanation

vaporisation occurs when a liquid changes to a gas; there is no temperature change during vaporisation

work energy transferred due to the action of a force acting through a distance

CHAPTER REVIEW QUESTIONS

Remembering

1 Distinguish between temperature, kinetic energy and internal energy.

2 What happens to temperature during a phase change?

3 What are the 'fixed points' on a thermometer?

4 Define 'thermodynamics'.

5 Define 'energy transformation'.

Understanding

6 You are eating a hot pie you have just removed from the oven and find that the pie filling appears much hotter than the crust. Was the crust much cooler than the filling? Explain your answer.

7 Why is it better to take an insulated hot water bottle to bed rather than an insulated hot brick?

8 A scalpel blade that needs to be sterilised is heated in a flame until it is glowing red-hot. It is then held above a glass of water in which it will be cooled.

- a Which has the highest average kinetic energy, the blade or the water?
- b The scalpel blade is dropped into the water. What happens to the average kinetic energy of the blade and of the water?

9 A hot cup of coffee is left to stand for a couple of hours. Explain on a particle level how it reached thermal equilibrium with the surroundings.

10 Why is it that the water going over a waterfall is warmer at the bottom of the fall than at the top?

Applying

11 Perth recorded its highest temperature of 46.2°C on 23 February 1991. What is this temperature in kelvin?

12 In an espresso coffee machine, steam at 100°C is passed into milk to heat it.

- a Calculate the energy required to heat 150 g of milk from 20°C to 80°C (c_{milk} = 4010 J kg^{-1} K^{-1}).
- b Calculate the mass of steam condensed.

13 A vacuum flask is designed to keep liquids hot or cold. A student seals 200 g of ice-cold water in a glass vacuum flask and finds that it warms up by 3.5 K per hour. Calculate the average rate of heat flow into the flask.

14 To deaden pain in minor operations, a liquid that vaporises easily (vapocoolant) is sometimes sprayed onto the skin. Explain how vapocoolant cools the skin.

15 Explain why food cooks more quickly when it is steamed rather than boiled.

16 If you blow moist warm air from your mouth onto a mirror on a cool day the mirror fogs. Has this process cooled or warmed the mirror? Explain.

Analysing

17 To test the temperature of an oven, a 500 g aluminium tray was placed in the oven and allowed to reach thermal equilibrium. It was then dropped into a plastic bucket that contained 2 L of water at 0°C. The temperature of the water rose by 7°C.

- a What was the temperature of the oven?
- b Why was a plastic bucket used?

18 An electric hotplate is required to heat a 1.6 kg block of ice at 0°C to steam at 100°C. If it takes 10 minutes to boil the water away and 15% of the heat generated by the electric hot plate is lost to the surroundings, calculate the rate at which the electric hot plate supplies energy to convert the ice into steam.

19 A lake can absorb huge amounts of radiation on a sunny summer's day and yet the temperature of the water can remain virtually constant. Explain how this is possible.

20 Suppose two substances, lead and water, are in thermal equilibrium. Equal amounts of heat energy are now put into each. Assuming the mass of each substance is the same, will they still be at thermal equilibrium? Explain in detail.

21 An industrial furnace uses electrical energy to melt iron. A large coil surrounds a crucible containing the metal. The whole furnace is surrounded by high-grade insulation to reduce operating costs. In spite of this, the high operating temperatures mean that some energy is still lost. In a particular situation 100 kg of iron is loaded into the furnace at an initial temperature of 40°C. Under these conditions the furnace has been found to be 70% efficient.

What total energy must be supplied to the heating coils to melt the 100 kg of iron?
[Assume $c_{iron} = 450\,J\,kg^{-1}\,K^{-1}$, $L_{iron} = 2.75 \times 10^5\,J\,kg^{-1}$, melting point of iron = 1540°C]

Evaluating

22 Why is it important to keep the lid on a simmering pot of water?

23 You are to run a stall at a school fair that sells cold soft drinks. The organisers expect to sell 500 cans (375 mL). Assume that 1 L of soft drink has a mass of 1 kg and the same specific heat capacity as water.

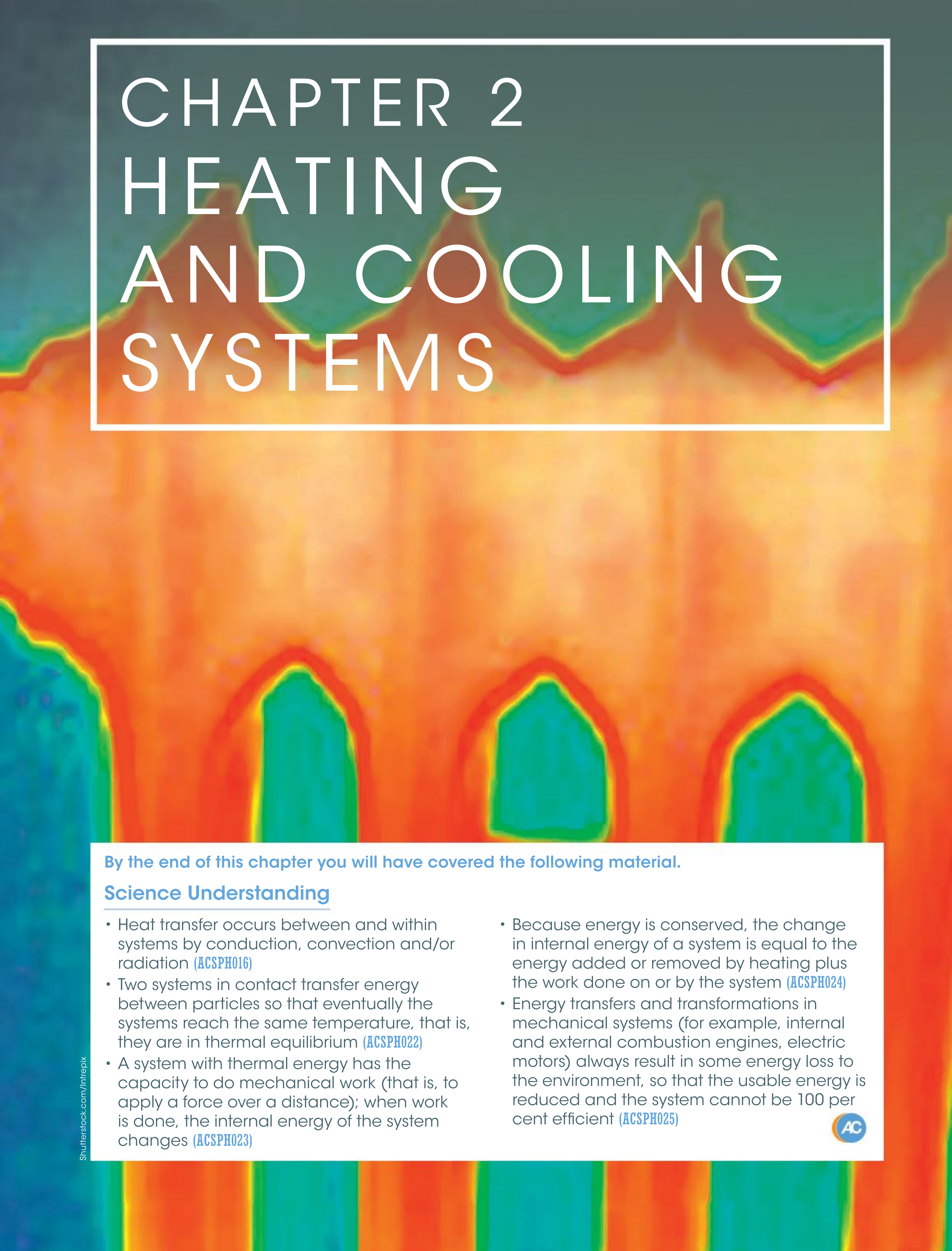
- a Calculate the mass of ice at −5°C needed to cool the drinks from 33°C to 2°C.
- b List the precautions you would take so that you can maintain the drinks at about 2°C over the day.

24 Explain in detail how you would evaluate a manufacturer's claim that their electric kettle was (95±1)% efficient.

Reflecting

25 How has your understanding of the words 'heat' and 'temperature' changed as you studied this chapter?

CHAPTER 2 HEATING AND COOLING SYSTEMS



Shutterstock.com/Intrepix

By the end of this chapter you will have covered the following material.

Science Understanding

- Heat transfer occurs between and within systems by conduction, convection and/or radiation (ACSPH016)
- Two systems in contact transfer energy between particles so that eventually the systems reach the same temperature, that is, they are in thermal equilibrium (ACSPH022)
- A system with thermal energy has the capacity to do mechanical work (that is, to apply a force over a distance); when work is done, the internal energy of the system changes (ACSPH023)
- Because energy is conserved, the change in internal energy of a system is equal to the energy added or removed by heating plus the work done on or by the system (ACSPH024)
- Energy transfers and transformations in mechanical systems (for example, internal and external combustion engines, electric motors) always result in some energy loss to the environment, so that the usable energy is reduced and the system cannot be 100 per cent efficient (ACSPH025)

Introduction

Understanding heat and controlling the transfers and transformations of heat energy is vital for the survival, health and wellbeing of all living things. Humans have a unique responsibility to use that knowledge wisely. Earth is in the 'Goldilocks zone', where conditions are 'not too hot and not too cold'. Its rotation causes day and night and its tilt causes the seasons. Incoming **electromagnetic radiation**, or **insolation**, is distributed through land, sea and air. Some of this incoming energy is reflected, but some is absorbed by Earth's surface and atmosphere. The atmosphere also prevents this absorbed energy being re-radiated into space. Consequently, Earth's average temperature is 15°C, 35°C higher than it would be otherwise.

Humans maintain a constant body temperature of 37°C by a combination of adaptations and physical control, such as heating and cooling systems and clothing. Steam engines powered the 19th century Industrial Revolution. Refrigerators changed the way we produce and preserve food. Other living things rely on adaptations alone. Emperor penguins survive at −70°C; the single-celled microbe, Strain 121 (*Geogemma barossii*), survives and reproduces in a hydrothermal vent at 121°C.

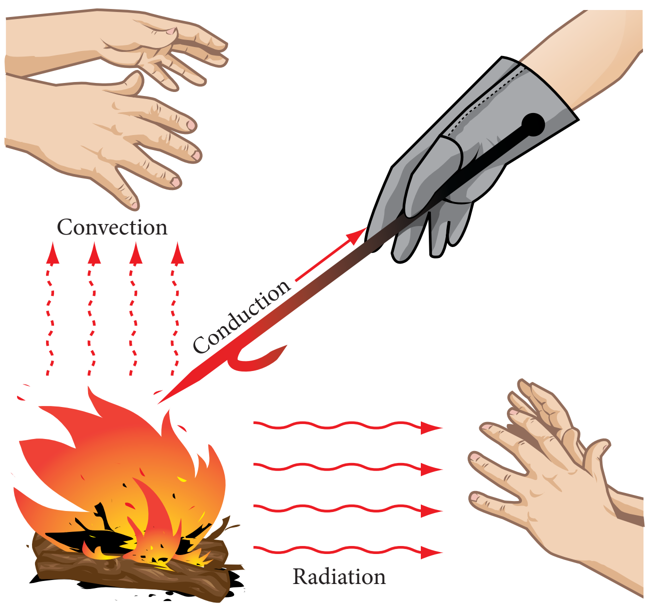
We continue to produce and consume large amounts of energy, much of which is wasted as heat. Reducing the effects of anthropogenic climate change is increasingly focused on insulation, efficiency and transition to renewable resources such as solar, wind, wave, hydro and fusion.

Figure 2.1 ▲ Extreme weather conditions are caused by temperature differences in the atmosphere.

Energy transfer models

In this chapter we use a simple kinetic particle model (see Chapter 1) to explain a wide variety of phenomena. Heat energy always moves from a region of high temperature to a region of low temperature. It can be transferred by conduction, convection or radiation. These transfer processes are vital to the existence of Earth as we know it.

Figure 2.2 ▶ Heat can be transferred by conduction, convection or radiation.



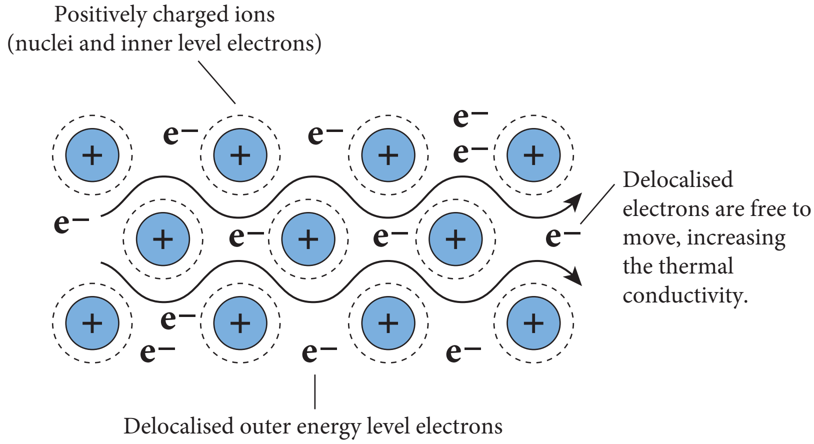
Conduction

Conduction is the transfer of heat energy through a substance by particle collision. There is no net movement of particles. The particles in the hottest part of a material, such as a metal rod, vibrate further from their usual positions. Collision with lower-energy particles transfers energy. Ultimately, the average kinetic energy of all the particles becomes the same – the substance reaches thermal equilibrium.

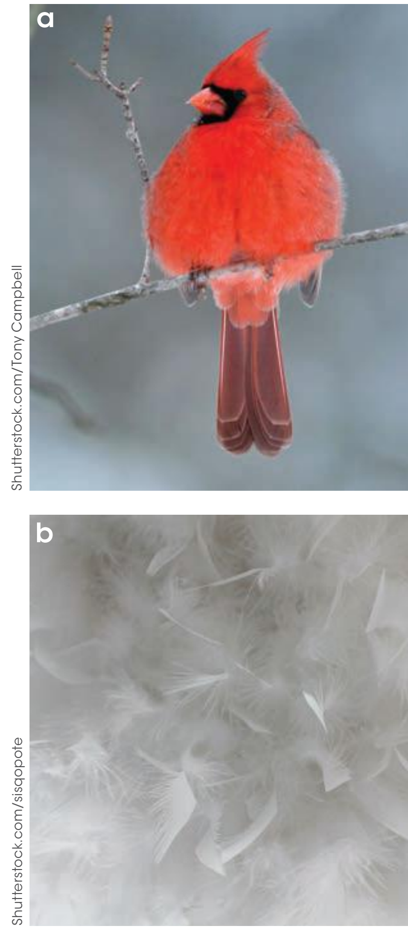
Thermal conductivity

Different materials have different conducting properties. **Thermal conductivity** measures how much energy per second flows through 1 metre of a material per degree temperature difference between the two ends of the material. The unit of thermal conductivity is $W\,m^{-1}\,K^{-1}$. Solids are better **heat conductors** than liquids or gases. **Heat insulators** are poor heat conductors.

Metals are particularly good heat conductors. They have large thermal conductivities. The large numbers of relatively unattached electrons in metals, which are relatively free to move, transfer kinetic energy quickly. The **delocalised valence electrons** transfer energy to other electrons and atoms at a faster rate than electrons that are tightly bonded.



▲ Figure 2.3
Delocalised electrons are free to move in metals and can conduct thermal energy quickly.

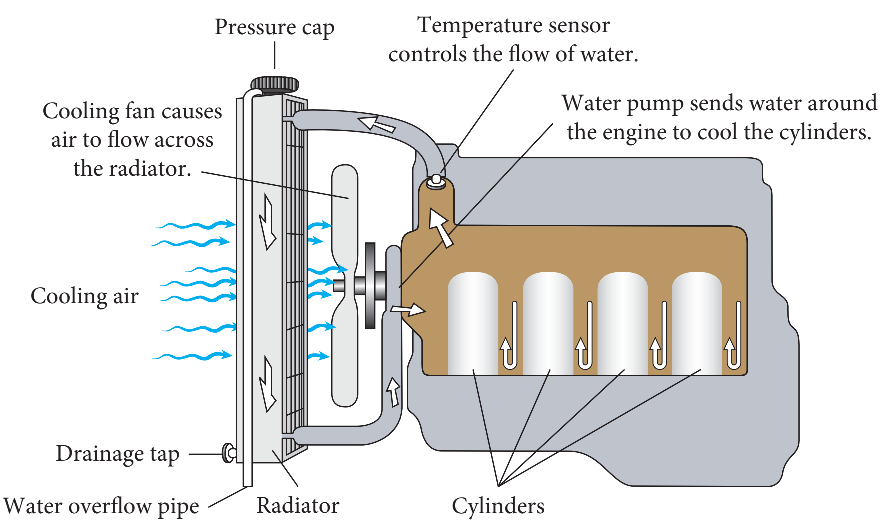


▲ Figure 2.4
a) Birds fluff up their feathers to retain body heat when it is cold.
b) Fine down feathers found under the tougher exterior feathers trap air in their fine structures.

Good heat conductors, such as liquid sodium, are used in some nuclear reactors to transfer heat to water. Other good conductors are used in refrigerators, disc brakes, computer heat sinks and car engine radiators.

Almost all non-metal materials, including gases, are insulators. Unlike metals, non-metals do not have free, delocalised electrons. Energy transfer occurs between relatively fixed neighbouring particles. When they are cold, birds and cats fluff their feathers and fur to trap air. Consequently, less heat is transferred from their bodies.

Good insulators are used in house insulation, thermos (Dewar) flasks, padded jackets and duvets.

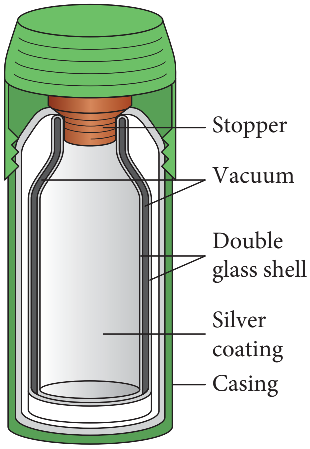


▲ Figure 2.5
Cooling systems for a car's engine use conduction, radiation and convection to remove the waste heat.

▲ Figure 2.6
Arctic First Peoples use fur insulation to trap air.

WOW

Dewar flask



Sir James Dewar (1842–1923) designed a flask to minimise energy transfers by conduction, convection and radiation. Dewar flasks are used to store hot or cold liquids such as liquid nitrogen (boiling point 77 K) and liquid oxygen (boiling point 90 K). They have a double-walled Pyrex glass vessel with silvered walls to reflect heat. The space is evacuated. The small neck also helps reduce heat transfer.

◀ **Figure 2.7**
A diagram of a Dewar flask, used to store hot or cold substances

EXPERIMENT 2.1

INSULATION

Aim

To compare the insulating performance of dry and wet materials

Materials

- three metal cans of the same size
- measuring cylinder or beaker
- polystyrene lids or sheet
- sharp knife (to cut the polystyrene, if necessary)
- time-measuring device
- flexible insulating materials such as cloth, doona or blanket material
- temperature-measuring equipment, such as temperature probe or thermometer
- source of very hot water

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
Scalding or burns can occur when handling very hot water.	Use safety gloves.
The knife is sharp.	Take care when using sharp knives.
Water on the floor can cause slipping.	If water spills, clean up immediately.

Procedure

1. Make a lid for each can. In the centre of the lid, make a hole to fit the temperature-measuring device snugly.
2. Fill each can with the same, measured volume of very hot water.
3. Surround one can with dry material and one can with an equivalent amount of wet material. Leave one can uncovered.
4. Record time and temperature in each can every 2 minutes for at least 12 minutes.

▶

Results

Plot temperature versus time graphs for each can on the same axes.

Analysis of results

Rank the cans in order of effective insulation of heat.

Discussion

1 A person in wet clothing is suffering hypothermia. Should you get them into dry clothes immediately?
2 How could your results be made more accurate?
3 What would you do differently next time?

Taking it further

Compare your results with those of others in the class.

Conclusion

From your results how do you compare the insulating performance of dry and wet materials?

Convection

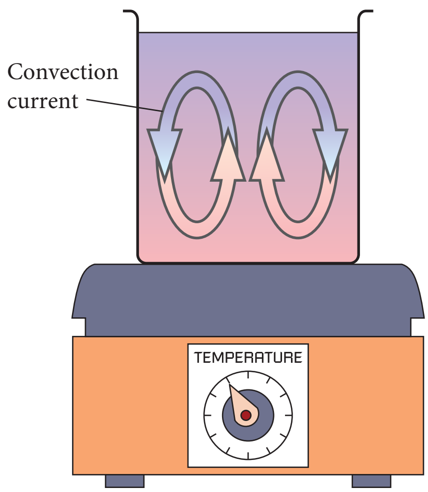
Birds soar gracefully on **convection currents**, or **thermals**, caused by temperature differences between air masses.

Convection is the transfer of heat energy by bulk movement of particles. The flow of particles away from a warmer to a cooler region produces a convection current. Convection currents only occur in fluids (liquids and gases), which have relatively weakly connected particles, but more so in gases than liquids because the particles in a gas are less tightly connected.

In Figure 2.9, warm, less dense water at the bottom flows upwards, while more dense water at the top sinks. A **convection cell** is produced.

Convection currents and convection cells occur where warm and cold fluid masses intersect, for example, in the atmosphere and oceans and in hydronic home heating systems.

▲ Figure 2.8
Soaring in thermals



▲ Figure 2.9
Convection currents transfer heat energy in water.

EXPERIMENT 2.2

BOTTLED CONVECTION CURRENTS

Aim

To explore convection in liquids

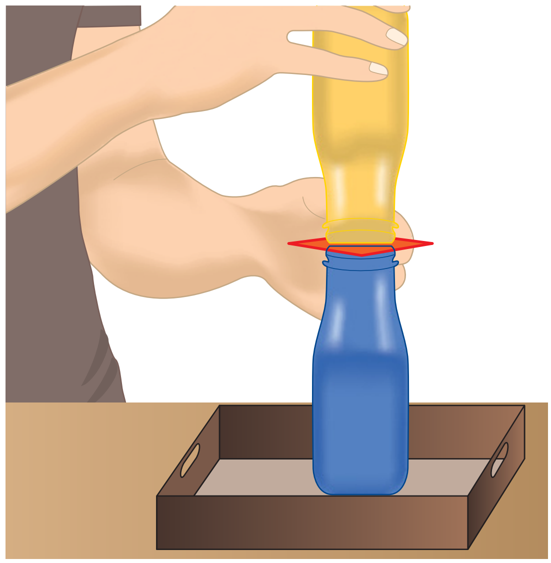
Materials

- four empty identical transparent bottles with a mouth at least 8 cm across
- warm and cold water
- food colouring (yellow and blue)
- an old playing card or card of similar dimensions and strength
- digital camera (optional)

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
Water on the floor can cause slipping.	Perform the experiment in a sink or large tray to avoid spills. Clean up any spills immediately.

Procedure

1 Fill four bottles, two with warm tap water and two with cold water.
2 Colour the warm and cold water with yellow and blue food dye respectively.
3 Place the playing card over the mouth of one of the warm water bottles.
4 Over a sink or large tray turn the bottle upside down and rest it on top of a cold water bottle. Make sure that they are exactly aligned mouth to mouth with the card separating the two liquids.
5 Bring the two liquids into direct contact by sliding the card out carefully.
6 Observe what happens to the coloured liquids.
7 Repeat steps 3–6, but this time place the cold water on top of the warm water.



▲ **Figure 2.10**
Experimental set-up

Analysis of results

1 For each situation describe any changes to the:
  - cold water.
  - warm water.
2 Support your descriptions with annotated diagrams or annotated photos, Voki or video clip.
3 Use the kinetic particle model to explain what you observed.

Discussion

Which of the two experiments could be used as a model to explain:

a ocean currents?
b the formation of thunder clouds?
c why water feels colder the deeper you dive?

WOW

Thunderstorms

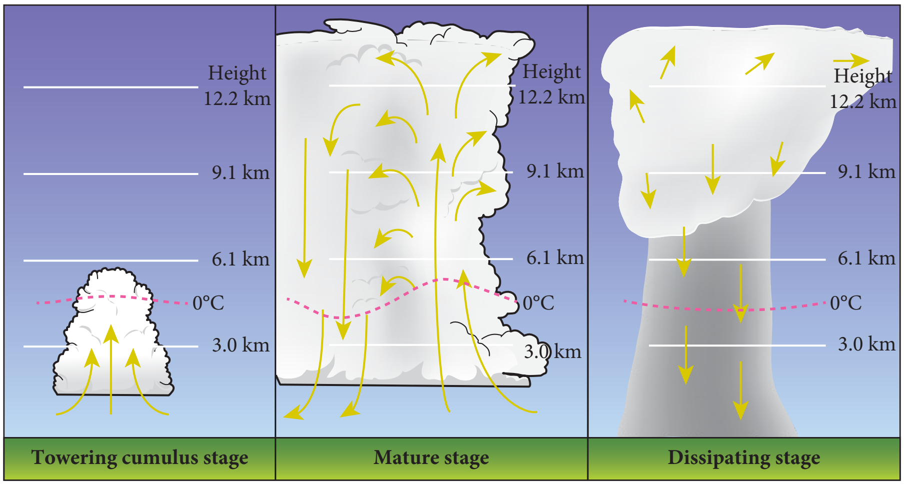


Diagram adapted from NOAA National Weather Service training materials

▲ Figure 2.11 Three stages of a thunderstorm

Thunderstorms undergo three stages.

1 Development: A bubble of air, heated by the ground, starts to create an updraft.

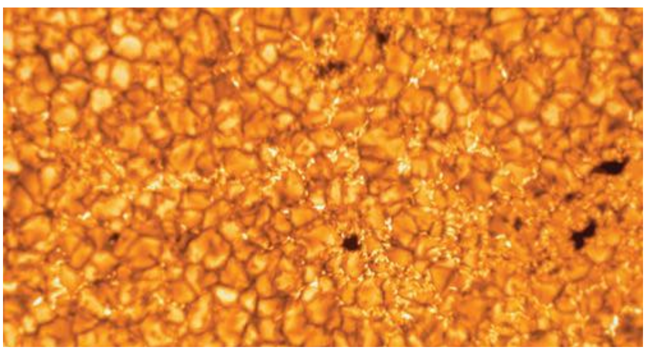
2 Maturity: Water moisture in the air mass condenses and falls as rain or hail.

3 Dissipation: The air mass cools and moves downwards.

On average thunderstorms are 24 km across. The three stages take an average of 30 minutes.

Solar convection

Hot plasma cycles from the centre of the Sun to the surface and back due to convection.



Science Photo Library/National Optical Astronomy Observatories

◀ Figure 2.12 Convection cells on the Sun's surface

Thermal lift

Thermals were first used for glider flight in 1921 by William Leusch in Germany, 20 years after the first powered flight, The pilot uses a thermal to increase altitude by flying in a spiral pattern before flying off to the next thermal. Thermals appear over towns, freshly ploughed fields, sealed roads and, occasionally, over power stations and fires.

Image by Laurence Wright, courtesy of the London Gliding Club

◀ Figure 2.13 An early glider being launched. This is the 'gull wing' Göppingen Gö 3 Minimoa (Germany, 1936).

Radiation

Radiation is the transfer of energy that does not need a medium. Unlike conduction and convection, radiation does not involve particles of matter. Except at 0 K, all objects emit electromagnetic radiation (Figure 2.14).

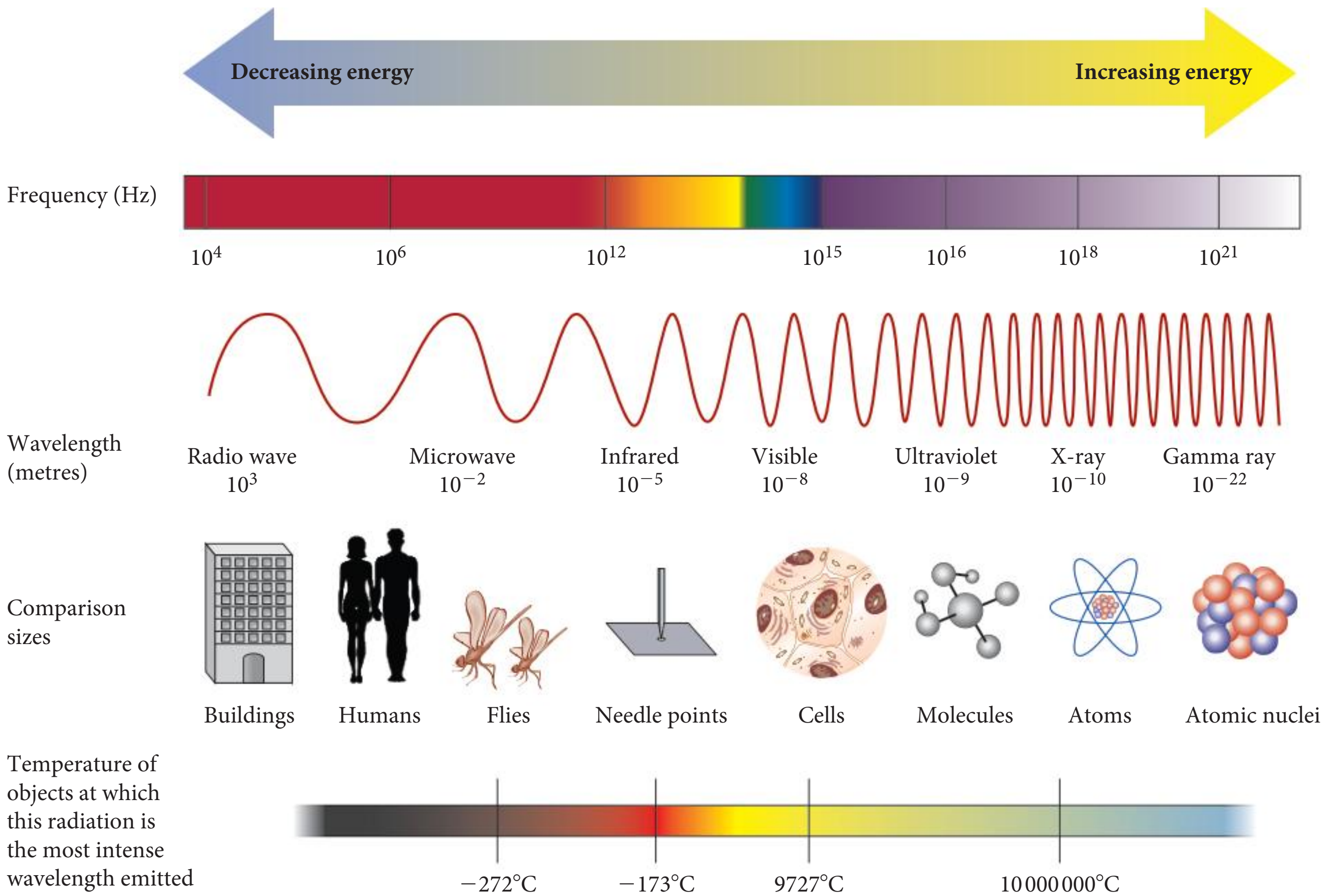


Figure 2.14 ▲
The electromagnetic spectrum

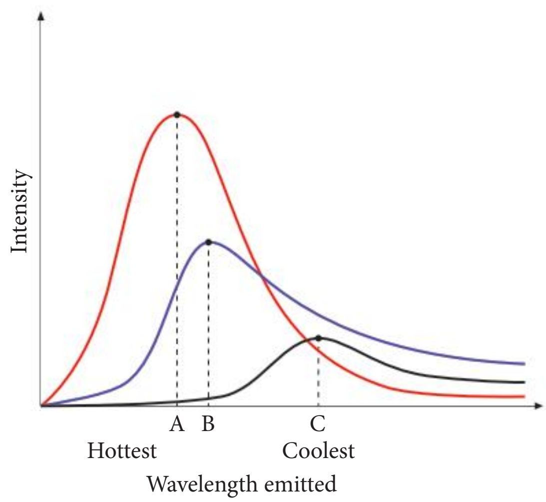


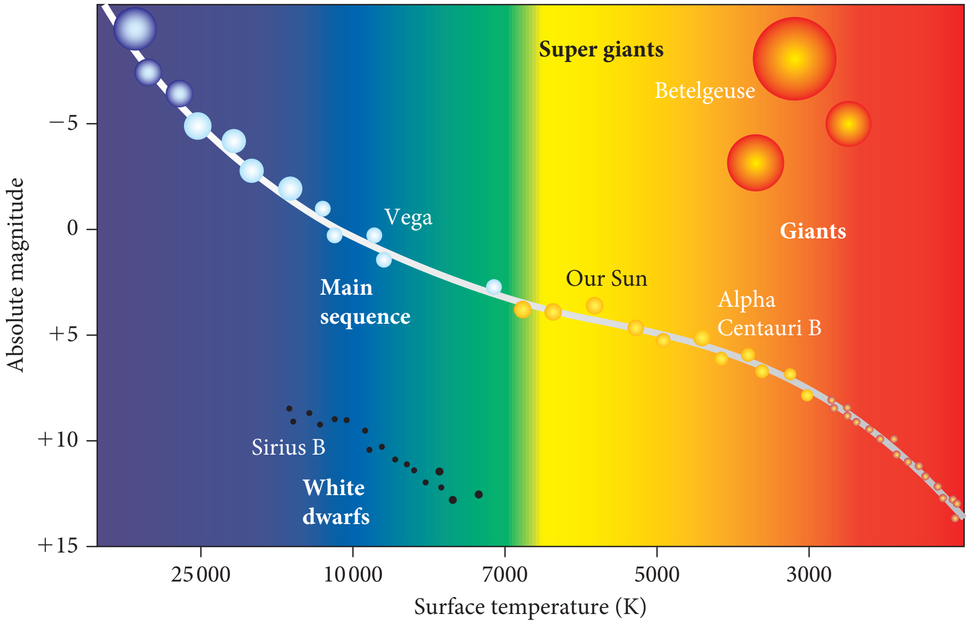
Figure 2.15 ▶
Planck curves showing peak intensities for three objects at different temperatures. Note that the temperature scale reads from right to left.

The intensity of radiation from an object clusters around a peak temperature on a Planck curve. In space, gas clouds (around 0 K) emit radio waves, stars (3000–30 000 K) emit ultraviolet and visible light. Warm bodies mostly emit infrared radiation. Planck curves are used to measure temperature in furnaces and stars. At 800°C, objects glow dull red. Stars such as Spica, which mostly emit ultraviolet light, have temperatures of about 22 000 K.

WOW

Hertzsprung–Russell diagram

The colours of stars represent their peak temperatures. This is used on the Hertzsprung–Russell diagram, which represents stars of different sizes and ages.



▲ Figure 2.16 Hertzsprung–Russell diagram

Electromagnetic radiation is detected when it interacts with matter. Rod and cone cells in our eyes interact with visible light, ultimately sending electrical signals to the brain, which interprets the information. Heat-sensitive nerve cells in our skin detect infrared radiation.

Alamy/Ted Foxx

▲ Figure 2.17 Hands are warmed by a radiant energy source.

Thermal imaging cameras

Thermal imaging cameras show temperature variations across the surface of objects. Some cancers appear as surface hot spots. Identification of heat losses from buildings can lead to greater thermal efficiency. Firefighters use infrared cameras to detect flames obscured by smoke, and astronomers use them to see stars forming deep within nebulae.

Newspix/Lawrence Pinder

▲ Figure 2.18 Thermal imaging cameras are used in firefighting.

Emission and absorption of radiation by surfaces

When radiated energy (radiant heat) is incident on a surface, some of that energy is absorbed and the rest is reflected. The fraction that is absorbed depends on the type of surface material, its texture and its colour. Black and dark-coloured surfaces absorb more radiant heat than white or light-coloured surfaces. Hence a black car gets hotter inside than a white car on a sunny day.

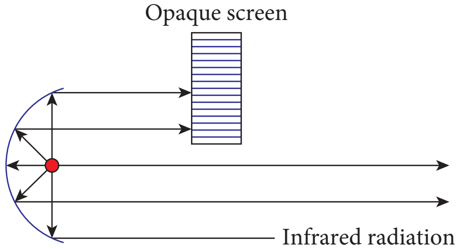


Figure 2.19 ▲ A curved radiator reflects infrared radiation.

Surfaces also radiate heat whenever they are at a temperature above absolute zero – so this is all the time. The amount of heat radiated depends on the temperature and the properties of the surface. Colour is again important. A black surface will radiate more heat than a white surface at the same temperature. Hence dark clothes will radiate more energy, and on a cool night will be cooler than white clothes of the same material. White clothes are cooler during a hot sunny day when heat is mainly transferred to you from the environment.

Desert-dwelling Bedouin people wear either lightweight or heavy, loose-fitting black or white clothing. The loose fit enables convection currents to cool the body between skin and cloth. White cloth reflects energy. Black cloth emits and absorbs energy. Light cloth allows heat to transfer away more quickly than heavy cloth. The choice of cloth depends on factors such as time of day, purpose and cultural norms.

Figure 2.20 ▶ A desert-dwelling Bedouin dressed in a) heavy black robes and b) light white robes.

Corbis/Sergio Pitamitz/Robert Harding World Imagery

Alamy/Prisma Bildagentur AG

QUESTION SET 2.1

Remembering

1 What makes a good absorber of radiant heat?

2 What method of heat transfer best distributes heat energy in:

a solids? b liquids?

Understanding

3 Use the kinetic particle model to explain how thermals may form over towns, freshly ploughed fields and sealed roads.

4 What is the purpose of the curved, silver metal material that makes up the back of an electric bar heater?

Applying

5 How could you remove heat energy from a car's engine?

6 What colour roof tiles would you choose for a home built in a hot climate? Explain your answer.

7 Describe how your understanding of the kinetic particle model helps you to explain the formation of thunderstorms.

Analysing

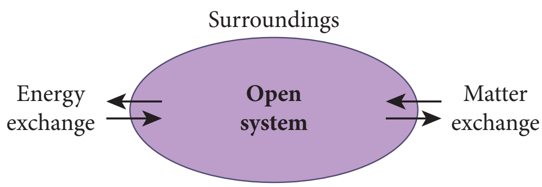
8 Use the kinetic particle model to explain convection currents in a kettle that is cooling.

Reflecting

9 How has your understanding of heat and energy transfer processes changed?

Conservation of energy in open systems

Energy cannot be introduced into, or emitted from, an **isolated system**. The only truly isolated system is the universe. Many systems approximate an isolated system. The calorimeter in a calorimetry experiment is a good approximation of an isolated system once you have closed the lid. It is well insulated so very little heat can enter or leave the system.



▲ **Figure 2.21** A schematic representation of an open system

A **closed system** is one that energy can enter or leave, but the amount of matter in the system remains constant. The matter may change state, for example from a liquid to a gas, due to heat transferred into the system but no matter enters or leaves the system. The water and coolant in a car's cooling system is an example of a closed system. Heat is transferred into the cooling water by the hot engine, and transferred out of the system by radiation at the radiator. Work is done on the water by the water pump that keeps it circulating. Hence energy is transferred in and out of the system by both heat and work, but no matter enters or leaves.

An **open system** can have both energy and matter moving across its boundaries. The ocean is an open system. Energy enters and leaves mainly by heat, and matter enters and leaves by evaporation, rain and the flow of rivers into the ocean.

The human body is also an open system. It receives energy and matter from food and oxygen, and transfers energy by convection and radiation and matter by excretion.

Energy efficiency

The **energy efficiency, η**, of an open system is the fraction of the input energy that produces a useful output. It is usually expressed as a percentage:

$$\eta = \frac{\text{energy output} \times 100\%}{\text{energy input}} \text{ or } \eta = \frac{\text{useful energy transferred} \times 100\%}{\text{total energy used}} \text{ or } \eta = \frac{\text{useful work output} \times 100\%}{\text{total work input}}$$

A car is, at best, about 30% efficient because only 30% of the chemical energy released by combustion of the fuel is used for moving the car. The 70% of energy that is transferred to the environment is mostly in the form of heat and sound.

High-grade energy resources, such as solar, chemical and electrical energy, transform energy relatively efficiently. **Low-grade energy resources**, principally heat, transform energy inefficiently. When producing a desired output, all open systems transform energy from high-grade resources to low-grade resources.

WORKED EXAMPLE 2.1

A family car produced 570 kJ of energy from combusting some petrol with air. The car converted only 190 kJ of this energy to useful mechanical energy. What is the efficiency of the car? (3 marks)

Answer	Logic	
$\eta = \frac{\text{energy output}}{\text{energy input}} \times 100\%$	Find the energy efficiency using this formula.	1 mark
$= \frac{190\ \text{kJ}}{570\ \text{kJ}} \times 100\%$	Substitute the known variables into the formula.	1 mark
$= 33.3\%$	Calculate the answer.	1 mark

Try this yourself

A modern electric motor working at a rate 150 J s^{-1} is 81% efficient. At what rate was the motor transforming electrical energy? — 3 marks

INVESTIGATION 2.1

EFFICIENCY OF MICROWAVE OVENS

Microwaves are quite effective at jiggling water molecules. Water molecules are polar. They have a positive end and a negative end. Microwaves cause these ends to oscillate. The resultant movement causes heating.

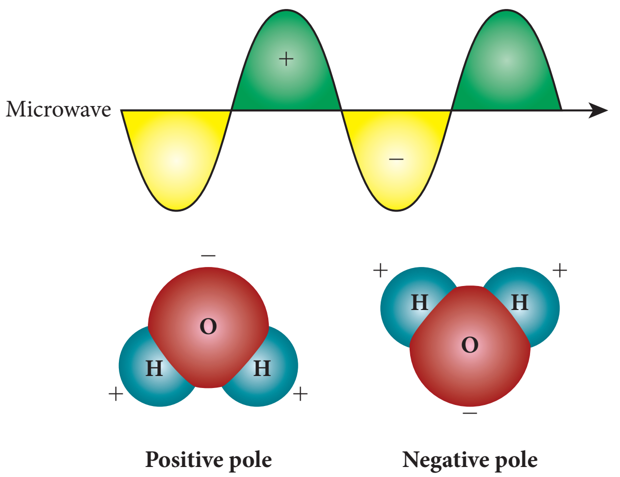


Figure 2.22 Polar water molecules are jiggled by microwaves as they pass by them.

Most foods have a high water content. Heating the water molecules heats the rest of the food chemicals as well. Microwaves penetrate 5 cm into food, so a combination of jiggling and conduction transfers heat through the food quickly.

The amount of energy transferred by a microwave oven to water can be measured using the quantities mass, specific heat capacity and temperature rise. The energy input to the microwave oven can be calculated from its power rating and time.

What is your aim?

To compare the efficiencies of three microwave ovens

What will you need?

Write a list of materials you will need to carry out your investigation.

Team up with other students so that you have a range of different microwave ovens that can be tested.

What are the risks?

Construct a table similar to the one below. Identify specific risks involved in the investigation and ways that you will manage the risks to avoid injuries or damage to equipment.

What are the risks in doing this investigation?	How can you manage these risks to stay safe?

How will you carry out your investigation?

Work out a research question and formulate a hypothesis to be tested.

Create a flowchart to show how data will be collected and analysed.

Draw a diagram to show the equipment in place for data collection.

What results will you collect?

Consider the data and the uncertainties you will collect.

Will you rely on the power rating on the microwave oven or check the manufacturer's specifications?

How will you ensure the data from other students is accurate?

How will you analyse the results?

How will you calculate the efficiency and related uncertainty?

How will you compare microwave ovens quantitatively?

What have you found?

Is there a relationship between efficiency and the power of a microwave?

What do you conclude?

Did you support or disprove your hypothesis? Write a conclusion based on this.

Ideas for improvement and further investigation

How could you improve your experiment?

Is the efficiency the same for different liquids and in different parts of the microwave oven?

QUESTION SET 2.2

Remembering

1 What is the difference between open, closed and isolated systems?

2 What is the difference between high-grade and low-grade energy resources?

Understanding

3 The human body is considered an open system. Why?

4 Is a sealed vacuum flask an open, closed or isolated system? Justify your answer.

Applying

5 A plane is 42% efficient. It transforms 85 000 kJ from fuel to useful work on a flight. How much energy is supplied to the engines by the aviation fuel?

6 Explain why an open system cannot be 100% efficient.

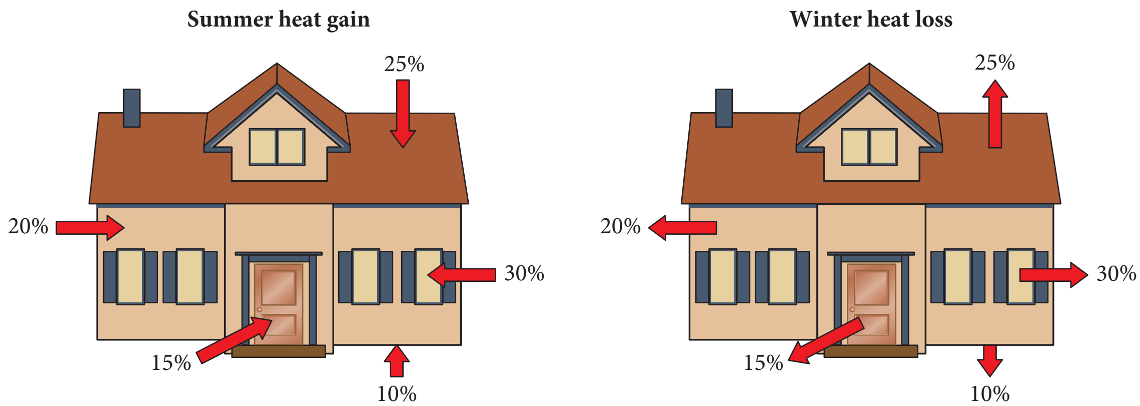
Analysing

7 How does the law of conservation of energy help you understand the efficiency of an open system?

Energy-efficient houses

Daily insolation, Earth's major external energy source, is the amount of energy incident perpendicular to an area of one square metre in one second, $\mathrm{J\,m^{-2}\,s^{-1}}$. Heat transfer from a region of higher temperature to a region of lower temperature is greatest when the temperature difference between two spaces is greatest. If the indoor temperature is less than the outdoor temperature, heat flows into the house. If the indoor temperature is greater than that outdoors, heat flows out of the house. In winter, therefore, heat loss from a house is greatest. In summer, heat gain from the environment is greatest. Comfortable indoor temperatures range between 18°C and 30°C. Fuel use and related costs can be minimised by sensible precautions.

Figure 2.23 ▼ Approximate heat transfers for a house in summer and winter



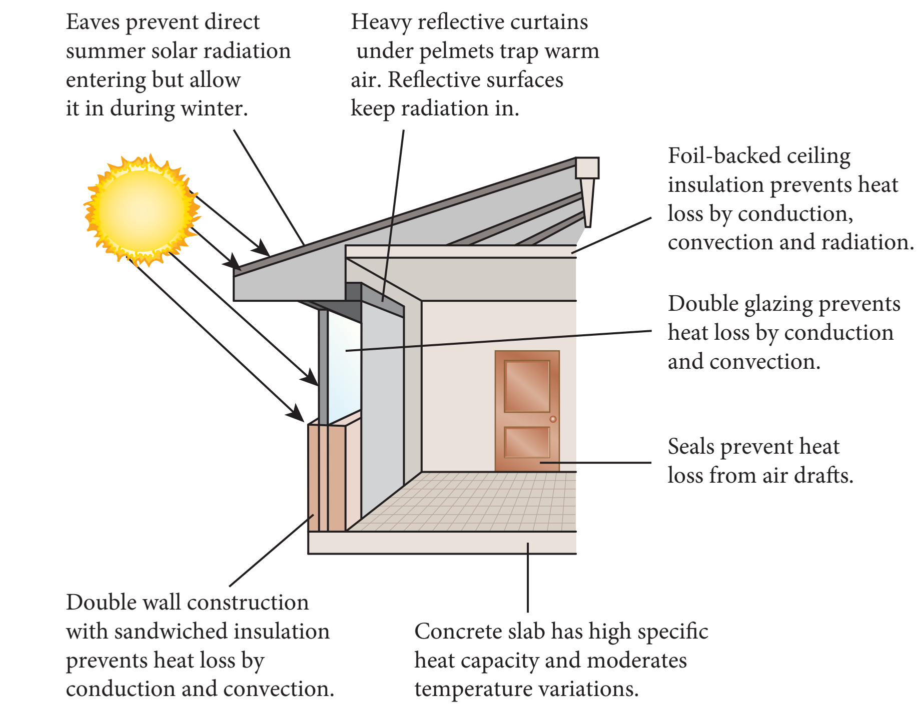
Reducing heat transfers between house and atmosphere

Ways to prevent energy transfers into or out of house roofs and walls rely on knowledge of conduction, convection and radiation. Foils, batts and air cells reduce conduction, convection and radiation. Reflective foils return radiation to the atmosphere. Insulation batts, with or without reflective coatings, trap air between interwoven fibres of non-flammable, insulating materials such as fibreglass or wool. Thinner, plastic air cells covered in reflective aluminium are also used. Batts and air cells inhibit conductive and convective transfer of heat; air cells and reflective materials are effective inhibitors of radiative heat transfer.

The roof cavity can become excessively hot on a summer's day. Many houses are fitted with a ceiling air extraction system that is driven by convection. Low vents under the eaves allow cooler air in. This cool air displaces the hot, less dense air through vents placed higher near the top of the roof.

Figure 2.24 ▶ Reflective aluminium bubble foil insulation is installed to reduce energy transfers.

Alamy/Beepstock



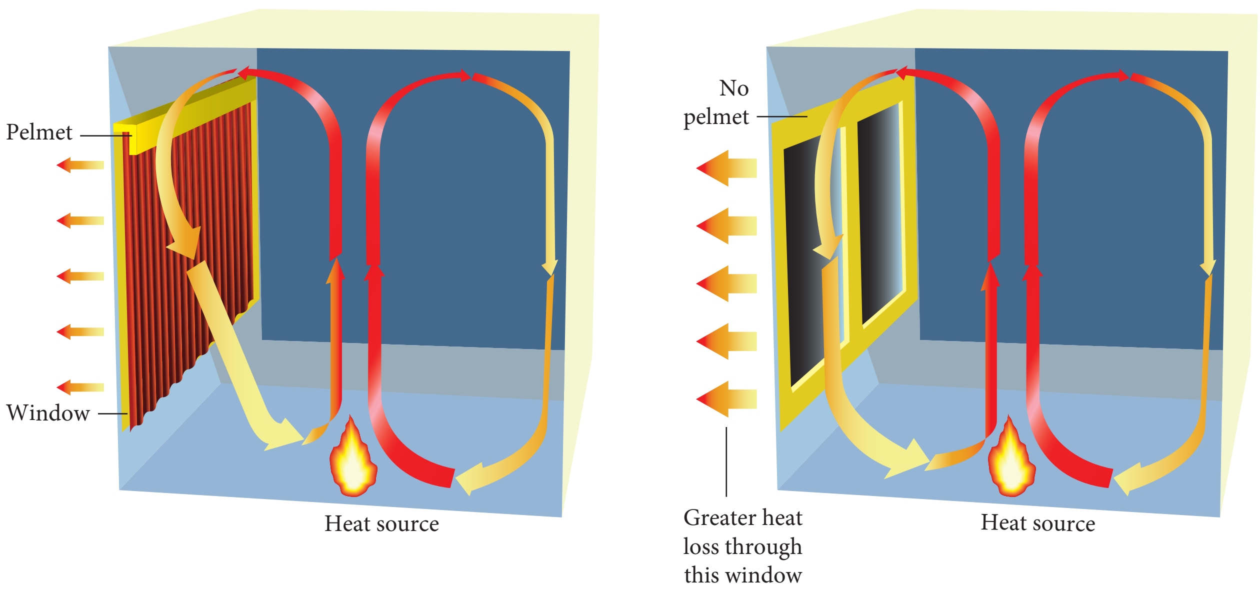
◀ **Figure 2.25**
Some energy-efficiency measures that can be used in house design

Windows

Significant energy transfer occurs through windows, which take up large areas of wall space. Double-glazed windows, comprising air sandwiched between glass panes, reduce heat transfer by conduction and convection, but not radiation.

Eaves are used to shade windows in summer but allow sunlight to warm the room in winter. For maximum effect, eaves must be designed for the latitude of the building. This is because the maximum elevation angle, that is, the maximum height of the Sun in the sky at a particular time of year (solar zenith) varies during the year and depends on latitude.

▼ **Figure 2.26**
Convection in a room with pelmets and heavy curtains compared with one without



In summer, reflective backings enable window curtains to reduce solar energy penetration into rooms. In winter, pelmets and heavy curtains limit convective losses. Convection currents arise in a room. Cooler air falls nearest the curtain, which gradually transfers energy to the window side. A second convection current between curtain and window carries that energy to the cooler window side, where the cooler air falls. The difference in temperature between inside and outside has been minimised. This is a good outcome, because the rate of energy transfer is proportional to the temperature difference. The energy transfer to the outside is also minimised.

ENERGY-EFFICIENT HOME DESIGNER

Find out about ways to improve energy efficiency by design.

Thermal mass

Thermal mass relates to the ability of building materials to absorb or release thermal energy. Buildings with a large thermal mass absorb thermal energy when the surroundings are at a higher temperature than the mass. This thermal energy is released when the surroundings are cooler, without reaching thermal equilibrium.

Materials that have good thermal mass, such as concrete, stone and brick, have large heat capacities. In cold climates, solar heated water is passed through pipes embedded in the concrete foundations of the house. This stores large amounts of energy for slow release during the winter months. Large water reservoir tanks placed under buildings serve the same purpose.

QUESTION SET 2.3

Remembering

1 When is the rate of heat transfer greatest?

2 What is the purpose of the reflective coating found on some insulation materials and curtain backings?

3 What is 'thermal mass'? Where is it used? Why is it used?

Understanding

4 Re-state the unit of insolation, $J\,m^{-2}\,s^{-1}$, in words.

5 Why is it not possible for heat to be transferred through a solid wall by convection?

6 Which is better at retaining heat energy in a house: double-glazed or single-glazed windows? Explain.

Applying

7 How would you place deciduous trees (those that lose their leaves in winter) around a house in Hobart to make it more comfortable in both winter and summer?

8 Air is a very poor conductor of heat energy. Insulation batts, consisting of fibre and air, conduct heat energy better than air.

a Which is the better insulator, air or insulation batts?

b Why do we install batts in the ceiling when it is already full of air?

Analysing

9 A 30000L water reservoir tank under the foundations of a house was solar heated from 22°C to 32°C during the summer. How much extra energy was stored in the water?

Reflecting

10 Use a visual organiser to summarise your understanding of heating and cooling processes to make houses comfortable.

Heating and the environment

Earth is solar powered. The Sun has a surface temperature of about 6000°C, and emits colossal amounts of energy, mostly as ultraviolet, infrared and visible light. About 31% of the solar energy reaching the atmosphere is reflected directly back into space, mainly by clouds, ice and snow. Oceans, land and vegetation reflect smaller amounts. A much smaller proportion is absorbed by atmospheric particles.

Climate scientists rely on temperature and other measurements to develop complex models that connect data with predictions, to explain the energy interactions that drive Earth's climate. These models are continually being refined and improved. Data and explanations are used to influence decisions about world energy use.

Earth's energy balance and climate

Earth's energy balance is the sum of incoming solar energy, energy emitted by Earth, and energy retained in or leaving the Earth's system. When the sum of incoming and outgoing energy is zero, Earth's temperature remains constant. This temperature depends critically on the energy retained in the atmosphere.

Earth has its own energy sources: heat from the molten core and heat released by radioactivity. The atmosphere absorbs 90% of this energy and 70% of the solar energy. Like all objects at temperatures above 0 K, the air and clouds emit radiant energy, most of which is absorbed. Latent heat and **sensible heat** are internal energies.

Earth emits more radiation into the atmosphere than it receives extra solar energy. Nevertheless, by the time energy is reflected, absorbed, radiated and re-radiated from all sources the amount leaving the Earth's system is the same as the additional amount received from the Sun. Currently, the energy retained in the atmosphere is greater than in the past, so the temperature of the atmosphere is increasing. This global warming leads to climate change.

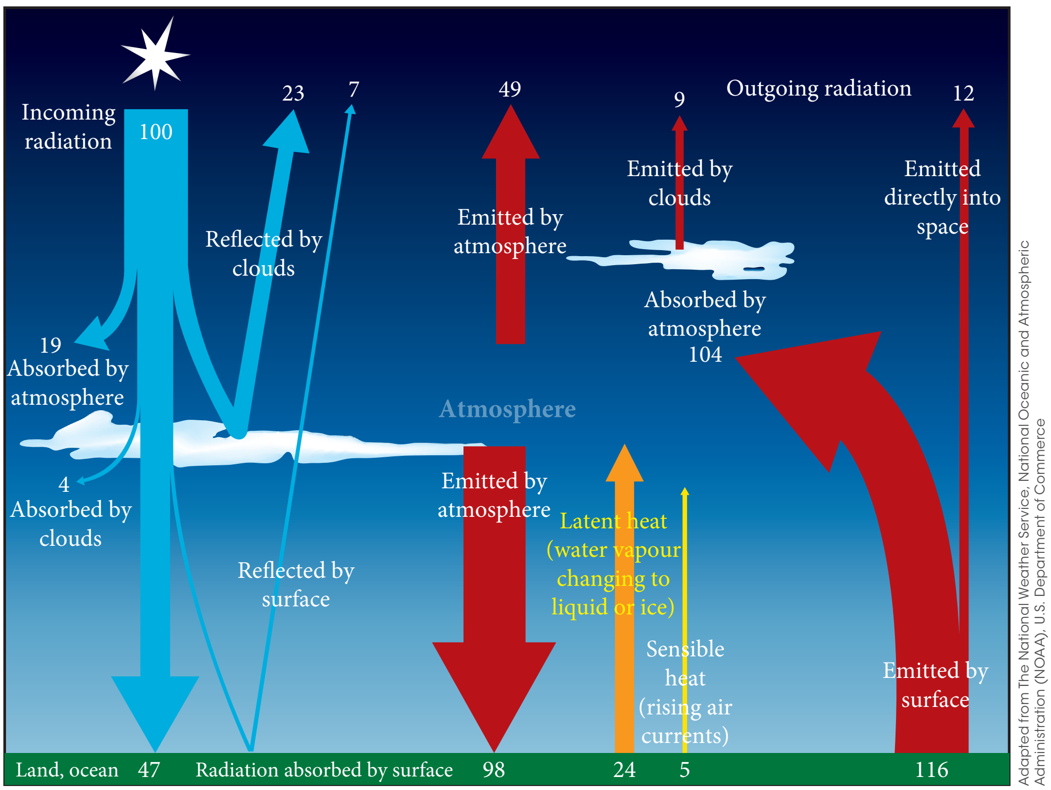
Figure 2.27 shows that, for every 100 units of solar energy (large blue arrow), 116 units of energy radiate from Earth (very large red arrow) and 29 units are internal energy (yellow arrows) comprising latent heat (24), related to condensation and freezing, and sensible heat (5), which causes air currents. The energy returned to space is 100 units: cloud reflections (23), Earth reflections (7), atmospheric emissions (49), cloud emissions (9) and direct emissions to space (12). About 145 units of energy is absorbed by the ground: solar energy (47) and emissions from the atmosphere (98).

CLIMATE SCIENCE

Check for updated summaries and academically rigorous comment on significant reports such as that of the United Nations Intergovernmental Panel on Climate Change (IPCC).

LATENT AND SENSIBLE HEAT

Learn more about these two types of heat.



◀ **Figure 2.27** A Sankey diagram showing Earth's energy budget: 100 units of incoming solar radiation are balanced by a range of outgoings.

Earth's tilt

Earth is tilted at an angle of about 23.5 degrees compared to the plane of Earth's orbit around the Sun. For 6 months it is more tilted towards the northern hemisphere (spring–summer), most of which receives solar radiation for longer than the southern hemisphere (autumn–winter). The reverse applies for the next 6 months.

Polar regions are colder because the energy is more spread out (less intense).

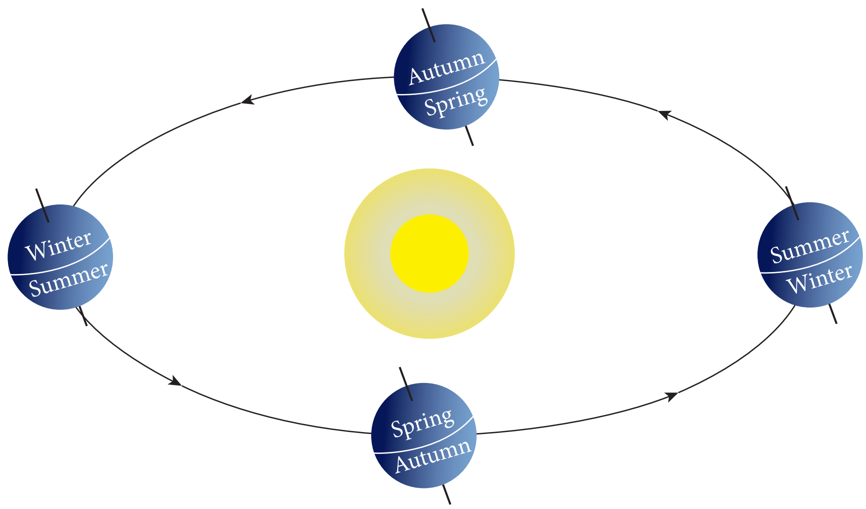


Figure 2.28 ▶ Seasons arise from Earth's tilt

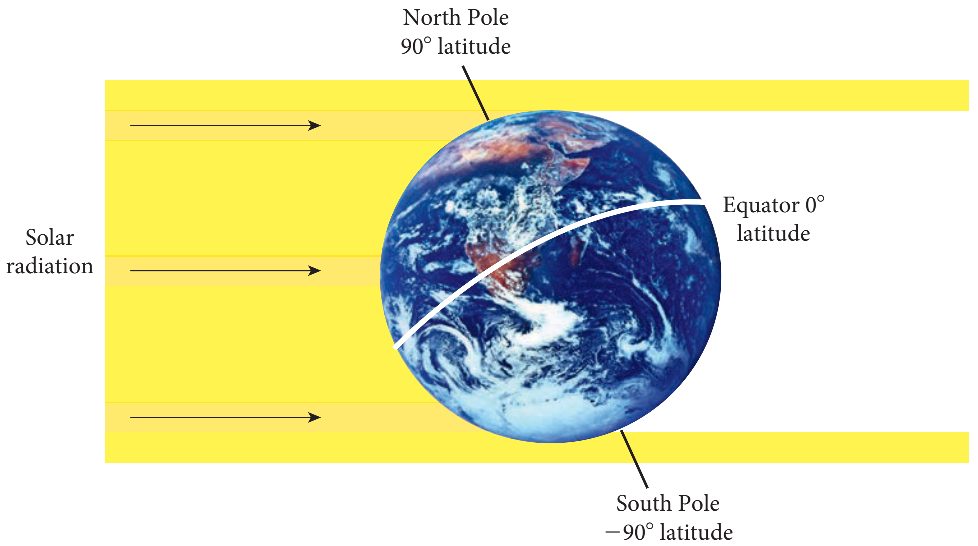
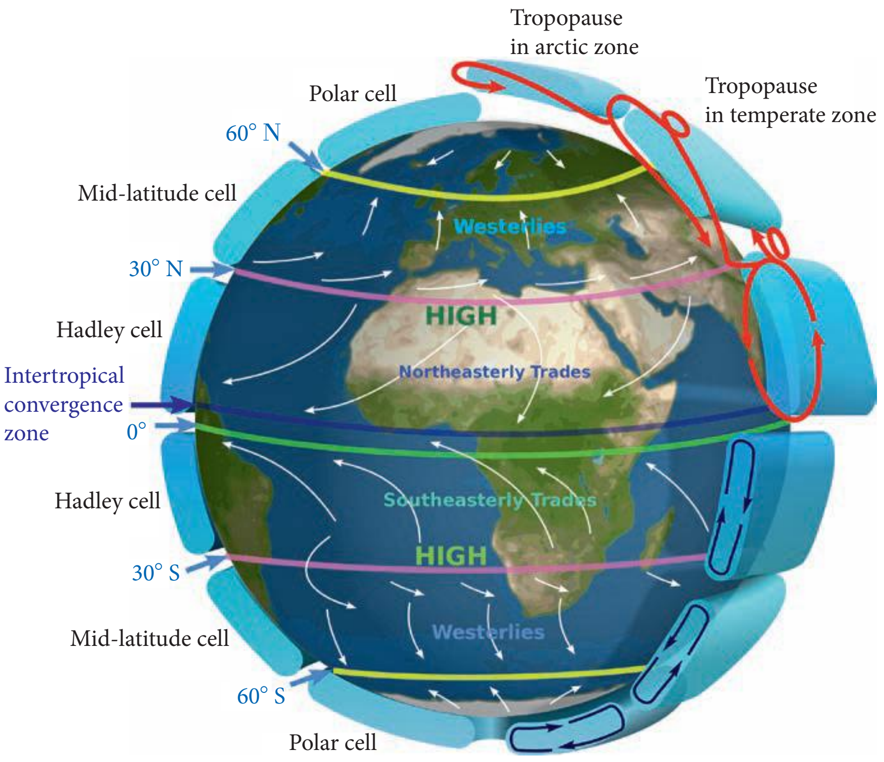


Figure 2.29 ▶ Solar energy distributions on Earth for a northern winter and corresponding southern summer

Earth's wind patterns

Temperature differences between the poles and the tropics cause large convection currents to form. The colder, more dense air over the poles sinks and moves towards warmer, less dense air in the tropics. Warmer tropical air moves up and towards the poles distributing heat and moisture throughout the atmosphere. They contribute to cloud and storm development (where rising motion occurs) and dissipation (where sinking motion occurs).



Vectorised image by Wikimedia Commons/Kaidor created from NASA public domain content

◀ **Figure 2.30**
The circulation of the air in Earth's atmosphere is driven by the temperature differences between the poles and the equator. The rotation of Earth causes the air to be deflected.

WOW

Coriolis effect

If a convection cell causes a wind to move towards the south (northerly wind), the wind rotates relative to an observer on the ground. This **Coriolis effect** is caused by the accelerating motion of Earth underneath the wind.

CORIOLIS EFFECT

Watch the Coriolis effect in a playground.

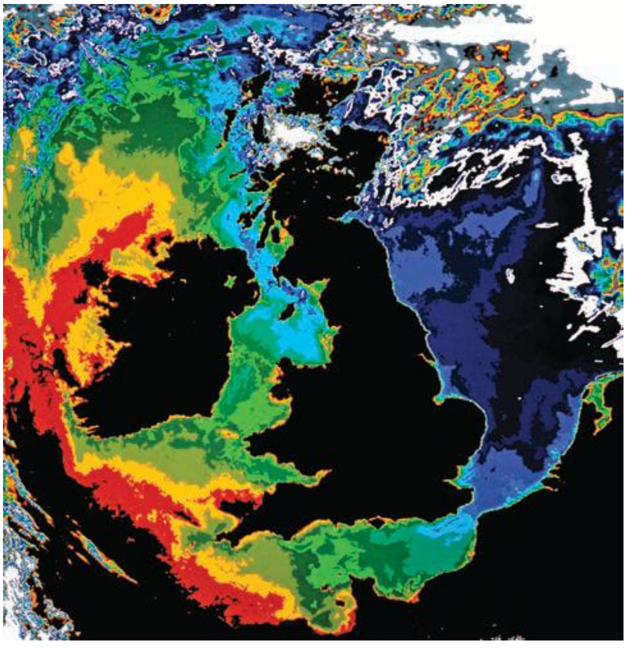
Climate and water bodies

Large bodies of water such as lakes and oceans influence weather and climate. As in the atmosphere, convection currents are important.

Oceans

The top few metres of oceans are solar heated. About 92% of the solar energy is absorbed, 8% is reflected. Solar energy is transported to greater depths by surface wave turbulence. Cooling mainly occurs via evaporation and by radiation from surface particles, which are in constant rapid motion. Evaporation causes the density of the increasingly salty water to increase. The denser, salt water layer sinks slowly over a long period of time, taking energy to greater depths.

The movement of large, deep ocean currents, the **thermohaline circulation**, is influenced by Earth's rotation, the differences in water temperature between the poles and the equator, and the prevailing winds.



Science Photo Library/NRSC LTD

▲ **Figure 2.31**
Thermal imaging of sea temperatures around the British Isles

Lakes

Large bodies of water moderate temperatures. This is a result of the high specific heat capacity of water, which enables lakes and oceans to gain or lose large amounts of heat energy with only small temperature changes.

WORKED EXAMPLE 2.2

During 8.0 hours of daylight, about 8% of 156 $J\,m^{-2}\,s^{-1}$ of insolation is reflected. The remaining energy penetrates to a depth of 2.0 m. Assume the energy is absorbed in the first 2.0 m.

The specific heat capacity of water is 4200 $J\,kg^{-1}\,°C^{-1}$. The density of water, ρ, is 1000 $kg\,m^{-3}$.

1 How much of the insolation is reflected? (1 mark)

2 In 8.0 h, what is the total solar energy, incident on a 1.0 m^2 surface:

a received? (2 marks) b absorbed? (1 mark)

3 For a 1.0 m^2 surface area and 2.0 m depth of penetration, into what mass is the total energy absorbed? (2 marks)

4 a What is the temperature rise over 8.0 h for a lake surface of 1.0 m^2? (2 marks)

b What assumption was required to calculate this temperature rise? (1 mark)

Answers	Logic	
1 8% of 156 $J\,m^{-2}\,s^{-1}$		
$= 12.5\,J\,m^{-2}\,s^{-1}$	Calculate the answer.	1 mark
2 a $E = 156\,J\,m^{-2}\,s^{-1} \times 1.0\,m^2 \times 8.0\,h \times 60\,min\,h^{-1} \times 60\,s\,min^{-1}$	Substitute the known variables.	1 mark
$\Rightarrow E = 156 \times 1.0 \times 8.0 \times 60 \times 60\,J$		
$\Rightarrow E = 4.5 \times 10^6\,J$	Calculate the answer.	1 mark
b E(absorbed) $= 92\%$ of $4.5 \times 10^6\,J$		
$\Rightarrow E$(absorbed) $= 4.1 \times 10^6\,J$	Calculate the answer.	1 mark
3 $\rho = \frac{m}{V}$		
$\Rightarrow m = \rho V$		
$\Rightarrow m = \left(1000\frac{kg}{m^3}\right) \times (1.0\,m^2 \times 2.0\,m)$	Substitute the known variables.	1 mark
$\Rightarrow m = 1000 \times 1.0 \times 2.0\,kg$	Calculate the answer.	1 mark
$\Rightarrow m = 2000\,kg$		
4 a $\Delta Q = mc\Delta T$		
$\Rightarrow \Delta T = \frac{\Delta Q}{mc}$	Transpose to make ΔT the subject.	1 mark
$\Rightarrow \Delta T = \frac{4.1 \times 10^6\,J}{(2000\,kg) \times (4200\,J\,kg^{-1}\,°C^{-1})}$	Substitute the known variables.	1 mark
$\Rightarrow \Delta T = 0.49°C$	Calculate the answer.	1 mark
b You need to assume that all the energy absorbed in the 2.0 m^2 volume each second remains in that volume for the whole day.	Give the correct assumption.	1 mark

Try these yourself

At a beach, the daily insolation is $180\,J\,m^{-2}\,s^{-1}$ of which 10% is reflected. The remaining energy penetrates to a depth of 10 cm.

The specific heat capacity of sand is $800\,J\,kg^{-1}\,°C^{-1}$. The density of sand, $\rho = 1200\,kg\,m^{-3}$.

1 How much of the insolation is reflected from a $1.0\,m^2$ area of sand:
   a every second? (1 mark)
   b in 2.0 h? (1 mark)

2 In 2.0 h, what is the total solar energy, incident on a $1.0\,m^2$ surface:
   a received (1 mark)
   b absorbed? (1 mark)

3 For a $1.0\,m^2$ surface area and 10 cm depth of penetration, into what mass of sand is the total energy absorbed? (1 mark)

4 a What is the temperature rise over 2.0 h for a sand surface of $600\,m^2$? (1 mark)
   b What assumption was required to calculate this temperature rise? (1 mark)

WOW

Daily desert temperatures

Sandy deserts heat rapidly during the day and cool rapidly at night. Maximum temperatures in the Tirari-Sturt Stony Desert in South Australia can reach 50°C during the day and fall sharply to below 0°C at night.

▲ **Figure 2.32**
Daily temperatures can vary by 50°C or more in the Tirari-Sturt Stony Desert.

See 'Specific heat capacity' on pages 15–16 in Chapter 1.

Climate and weather

Climate is the average atmospheric condition over large areas for long periods of time. Weather is the state of the local atmosphere in short time periods.

Land and sea breezes

Local weather conditions near the sea are affected by convection currents. Land and sea receive the same insolation, but the lower specific heat capacity of land causes it to heat and cool more quickly. A convection cell forms, driving air up over land during day and cooler air to flow in from the sea (sea breeze). At night, the process reverses, causing a land breeze towards morning. In between, the air remains relatively still.

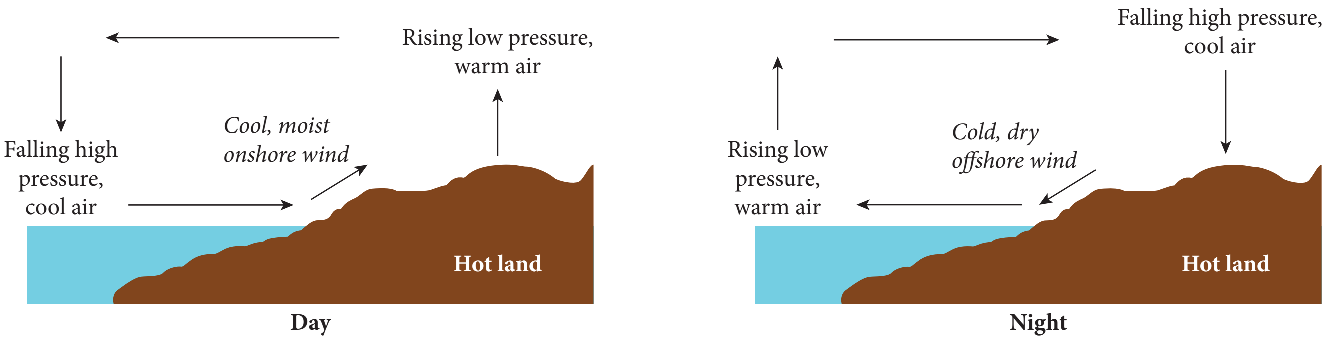


Figure 2.33 ▲ Convection cell causing land and sea breezes

Adiabatic heating and cooling

We have seen that the temperature of a system can be increased or decreased by the three mechanisms of heat transfer – conduction, convection and radiation. There is another way in which temperature (and hence internal energy) can be changed, and that is by work being done on or by the system. Remember that work is energy transferred due to the action of a force.

If your hands are at the same temperature and you rub them together, they will get warmer. This is not because heat is being conducted from one to the other, as they are at the same temperature. The skin of each hand is exerting a force on the other – friction. The work done by this force results in the increased temperature of your hands as energy is transferred to the skin of your hands.

A process in which energy is transferred only by work, and no heat enters or leaves the system, is called an adiabatic process.

Pressure, volume and temperature changes are important in chemical reactions. This is studied in Unit 2 Chemistry.

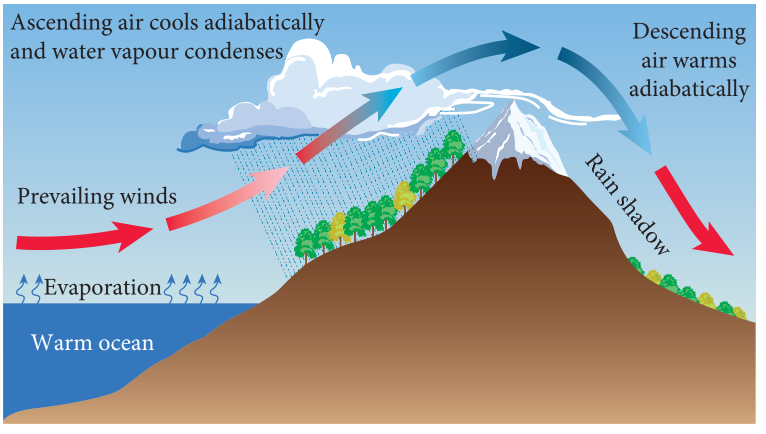
We can model any process that occurs in a system that is well insulated, or that occurs so rapidly that heat transfer is negligible, as being adiabatic.

When the system of interest is a gas, then the properties of the gas that change are its temperature, pressure and volume. When work is done on a gas by compressing it, then the volume of the gas decreases and its pressure and temperature increase. This is what happens inside a bicycle pump when the air is rapidly compressed.

When gas under pressure is allowed to expand and increase in volume, its temperature decreases. This is what happens when the combustion gases inside the cylinder of a car engine expand, pushing the piston and driving the crankshaft. Work is done by the gas as it expands, and at the same time the gas cools.

Atmospheric adiabatic heating and cooling occurs when a parcel of air is forced rapidly up a mountainside by a prevailing wind. The parcel of air expands, its temperature drops and moisture starts to condense out into clouds. The moist air drops its water load. On the other side, a rain shadow occurs. The air compresses and warms as it falls through mostly higher-density surroundings (Figure 2.34).

Figure 2.34 ▶ Mountain ranges affect rainfall and the warming and cooling of prevailing winds.



Very local influences

Human activity directly affects local weather. Buildings, roads and factories absorb and re-radiate more of the insolation than the vegetation they replaced. Vegetation adds to the water content of air via transpiration. Any reduction in vegetation limits the amount of transpired water, so evaporative cooling is reduced. These changes affect energy balance, and hence local weather.

EXPERIMENT 2.3

ENERGY ABSORPTION BY DIFFERENT SURFACES

Different surfaces absorb and emit radiation. Comparisons between surfaces and materials must be based on the same conditions; that is, they must be 'fair' comparisons.

Aim

To compare the temperature rise in a known mass of water in a metal can that is illuminated by the same heat source over a fixed time when the can is (a) painted in different colours (b) covered in different materials

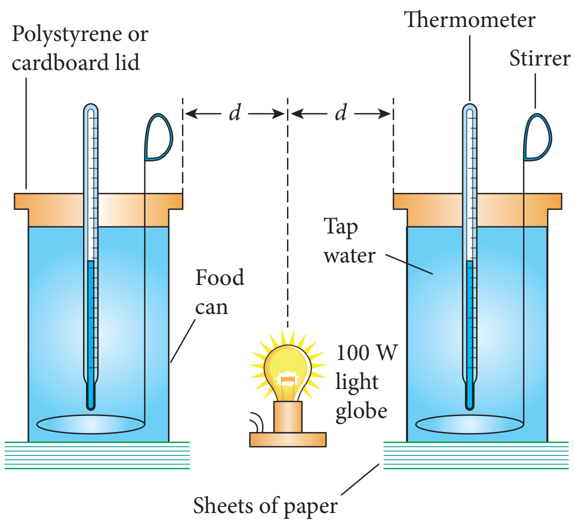
Materials

- four similar metal cans painted black, white, red, blue
- measuring cylinder or beaker
- polystyrene lids or sheet
- sharp knife (to cut the polystyrene, if necessary)
- ruler
- time-measuring device
- flexible insulating materials such as carpet, cloth and cardboard
- temperature-measuring equipment, such as temperature probe or thermometer
- 100 W globe in a suitably insulated mount
- scales
- stirrer
- sheets of paper

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
Globes can get very hot.	Do not touch the globe.
The Stanley knife is sharp.	Take care when using sharp knives
People can slip on spilt water.	If water spills, clean up immediately

Procedure

1. Use the polystyrene to make a lid for each can. In the centre of the lid, make two holes to fit the temperature-measuring device and the stirrer snugly.
2. Fill each can with the same, measured amount of water. Weigh the water.
3. Place the 100 W globe in its mount.
4. Place each can at the same measured distance from the globe.
5. Turn on the globe and start timing.
6. Record time and temperature every 2 minutes for 12 minutes. Remember to stir each can just before making the readings.
7. Wrap a different material around each can, making sure all materials are the same thickness.
8. Repeat steps 1–6.



▲ **Figure 2.35** Two coloured cans equidistant from a heat source

Results

1. Plot temperature versus time graphs for each coloured can and each material.
2. Plot histograms to show the maximum temperature reached for each coloured can and each material.

Analysis of results

Rank the cans in order of effective absorption of heat with respect to:

a colour. **b** material.

Discussion

1. How could your results be used for practical purposes?
2. How could your results be made more accurate?
3. What would you do differently next time?

Taking it further

Compare your results with the results of two other groups.

Conclusion

Write a short paragraph in which you compare the temperature rise in a known mass of water in a metal can that is illuminated by the same heat source over a fixed time when the can is:

a painted in different colours.

b covered in different materials.

Scientific literacy: Upsetting the energy balance

All matter is above absolute zero and therefore radiates energy. If it were not receiving energy from the Sun, Earth would become a frozen wasteland. The more the Sun heats Earth, the more Earth radiates energy back into space. There is an average temperature that Earth reaches at which the amount of energy received is equal to the amount of energy it radiates into space. Then Earth is in energy balance.

Over geological time, the amount of greenhouse gases in the atmosphere has varied greatly, matching the cycles of the ice-ages. These changes happened over time scales of between 40 000 and 100 000 years, enabling life on Earth to evolve and adapt. Currently, the levels are lower than they have been in the very distant past but much higher than over the last 3 million years.

Greenhouse gases are critical to reducing the rate of radiant cooling of Earth. Without them Earth would be a lot colder, the oceans would freeze, and life as we know it would not exist. Greenhouse gases strongly absorb Earth's thermal infrared radiation, so they effectively blanket Earth. They are able to absorb infrared energy by internally vibrating, bending and rotating. They re-radiate a large proportion of the absorbed energy back down to Earth.

The primary greenhouse gases are water vapour, carbon dioxide, methane, nitrous oxides and ozone. The atmosphere consists mainly of nitrogen and oxygen. They are not greenhouse gases because their diatomic molecules, O_2 and N_2, are too rigid to absorb infrared radiation. Triatomic molecules, such as H_2O and CO_2, and the pentatomic methane molecule, CH_4, are readily able to flex, and so absorb and re-radiate infrared energy.

Water vapour is the most important greenhouse gas. When the temperature rises, the rate of evaporation of water accelerates, increasing the amount of water vapour in the atmosphere. The water vapour traps more of Earth's thermal radiation and increases the average temperature of Earth's biosphere. The increase in temperature also increases the rate of growth of plants. Faster-growing plants take up more carbon dioxide and water through the process of photosynthesis. In the short-term, the amount of carbon dioxide and water in the atmosphere is reduced. This feedback cycle helps maintain a constant proportion of greenhouse gases.

Levels of greenhouse gases in the atmosphere are influenced by human activities. Since the 18th-century European Industrial Revolution, human activities have increased the amount of carbon dioxide in the atmosphere by approximately 38%. This rate of increase continues to grow. The extraction and burning of fossil fuels such as oil, gas and coal have been the major contributors to the increase. The clearing of huge swathes of Earth's forests has removed many of the trees that could have stored carbon via photosynthesis. This has reduced the capacity of the biosphere to remove CO_2 and H_2O from the atmosphere.

The rapid current change in global temperatures is one of the biggest issues facing humans across the globe. Without urgent worldwide action a 5–6°C increase in average surface temperature is predicted by 2100. A change of this magnitude will dramatically affect the climate and vegetation distribution, raise sea levels and produce more severe weather events including droughts and floods.

There have always been processes that lock away some of the carbon dioxide from the atmosphere. Short-term storage occurs through growth of plants, both on land and in the sea. Plants extract carbon dioxide as they grow. This is returned to the atmosphere when they decay after dying. It has no net effect on the carbon dioxide levels. To take the excess CO_2 from the atmosphere, plant growth must outstrip the clearing of plants. Unfortunately, we are cutting down the forests and returning the stored carbon more rapidly to the atmosphere as carbon dioxide.

In the long term, some of the dead plant material can become trapped in an oxygen-free region. This trapped material can be changed into coal and oil by compression. This process takes many thousands of years, but does permanently sink the carbon. But when this carbon that has been stored for millions of years is returned rapidly to the atmosphere, the biosphere cannot react quickly enough to remove it. This new carbon load adds to global warming.

Green phytoplankton in the oceans account for half the take-up of carbon by all plants. Via the food chain, the shells of marine organisms form from phytoplankton carbon. Eventually, the shells become part of ocean sediments which, over millions of years, form limestone and dolomite.

Large amounts of carbon dioxide dissolve in the oceans, forming carbonic acid. This increases the acidity of the water and stresses the phytoplankton and damages the shells of marine organisms. The uptake of carbon dioxide is reduced further.

World scientists have checked all data and subjected it to rigorous scrutiny. They agree: anthropogenic (human-caused) warming is occurring. They face the challenge of understanding the driving forces of this warming. There are many possible complex interactions and causes. Scientists are currently working to understand them. Science never has all the answers. Research mostly throws up more questions than it answers. Ongoing research is needed.

Questions

1 Name three significant greenhouse gases.

2 What happens to Earth's energy balance when greenhouse gases:
  a increase?
  b decrease?

3 How does the interaction of green phytoplankton and shellfish in the oceans contribute to the reduction of CO_2 in the atmosphere?

4 Explain why increases in the acidity of the oceans results in the reduced uptake of carbon dioxide.

5 What anthropogenic changes have caused global warming? Give your answer in terms of forest and farm practices.

6 Reflect on whether scientists have successfully convinced the world that climate change is real and that we need to act now. What evidence from the article would you use to convince someone that anthropogenic climate change is occurring?

QUESTION SET 2.4

Remembering

1 What are the main types of solar radiation?

2 Define these terms.
  a Climate
  b Weather
  c Land and sea breeze
  d Adiabatic

Understanding

3 a Define 'Earth's energy balance'.
  b Why must Earth remain in energy balance?

4 Use one or more diagrams to explain how the seasons in the southern hemisphere occur.

Applying

5 Earth receives about 156 $J\,m^{-2}\,s^{-1}$ of solar energy on average during the day. A solar hot water system is 80% efficient and has an absorbing surface area of 1.2 m^2. How long will it take for its 60 litres of water to heat up by 23°C?

6 Sea breezes are less likely to occur during winter. Use the kinetic particle model to explain why.

Analysing

7 Use the kinetic particle model to explain why a prevailing wind undergoes adiabatic heating as it flows down the rain shadow side of a mountain.

8 If Earth's axis of rotation were at right angles to the plane of its orbit (that is, not tilted), how would the seasons change?

9 Explain how evaporation contributes to the slow mixing of solar energy to lower levels of the ocean.

Case study

Professor Hui Tong Chua

Hui Tong Chua is a Winthrop Professor, Research Theme Leader and the Chemical Engineering Program Chair at the University of Western Australia. He is currently working in the fields of geothermal energy and waste heat utilisation. Professor Chua said that he was attracted to these fields of research by his desire to address the world's present and future needs for clean energy and water security in an environmentally responsible way. He is currently collaborating with world-class scientists and engineers in the areas of geothermal engineering, low-grade heat driven desalination, energy-efficient air conditioning and zero-emission hydrogen production from natural gas.

Professor Chua says that freshwater resources around the world are currently in short supply. The development of desalination technologies that produce fresh water from sea water and which are less expensive and use low-grade or waste energy to operate them are currently in great demand.

Courtesy of Professor Hui Tong Chua

▲ **Figure 2.36**
Professor Hui Tong Chua

His current project in this area is the development of low-grade heat desalination. This is a multi-effect distillation process that is especially tailored to use natural geothermal heat energy available from hot sedimentary aquifers, or the copious amount of low-grade waste heat energy that is available from large industrial processes such as refinery plants. Compared with conventional processes, Professor Chua said the process he and his team have developed can produce 20–40% more fresh water for the same flow of heat. This process is particularly suitable for the remote regions of Australia.

In the area of energy efficient air conditioning, his development of a multi-bed adsorption chiller that makes better use of low-grade heat to produce chilled water for air conditioning has been a significant advancement in efficient air conditioning. The temperature of the low-grade heat can be as low as 65°C. This means that the higher-temperature geothermal energy can first be used for power generation and then for air conditioning. This design could result in a big boost to the uptake of geothermal energy for district power production and air conditioning in places with abundant geothermal energy, such as Western Australia.

Professor Chua said that a third area of research he is involved with is the development of a catalytic process to convert natural gas into hydrogen without any carbon dioxide emission. He pointed out that the current level of carbon emissions are causing rapid global warming and climate change. He said the carbon produced in the production of the hydrogen is sequestered as highly graphitic carbon nano-onions. These solid, pure carbon products can be permanently stored or used as a substrate material for the loading of platinum nanoparticles in fuel cell applications aiming at clean power production. This will reduce the amount of carbon emitted into the atmosphere.

He pointed out that hydrogen can then be used as a non-polluting energy source for developing technologies such as hydrogen fuel cells, or in the production of chemical fertilisers such as ammonia.

Professor Chua said the technical challenges that arise from the activities in this field also actively stimulate him intellectually. Being able to provide answers to these challenges always gives him immense satisfaction.

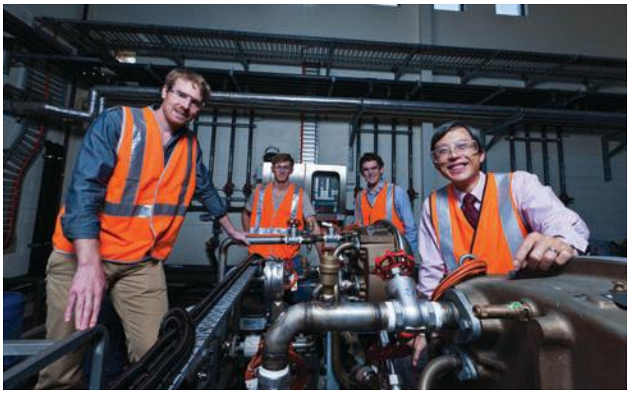


Photo by Matthew Galligan

▲ Figure 2.37
The engineering team for the prototyping of a new low-grade heat driven desalination process. The project is funded by the National Centre of Excellence for Desalination Australia. From left: Alexander Christ (a PhD student), Cameron Bruce McKenzie and James Maddock (both are Final Year Project students), and Hui Tong Chua.

Questions

1 Scientists worldwide are putting large resources into the development of cost-effective freshwater desalination plants. What reasons does Professor Chua give for this trend?

2 A major area of his research is developing technologies that can use low-grade waste heat. What do you think the major benefits can be from using these developing technologies?

3 From the interview, what can you deduce about why low-grade heat desalination is particularly suitable for the arid inland of Australia?

4 How can the process of converting natural gas into hydrogen contribute to the reduction of greenhouse gases and hence to the slowing of climate change?

5 Why do you think international collaboration between scientists is important?

6 Do some of your own research to find out what carbon nano-onions are and their possible uses. Write two paragraphs on your findings.

Work, heat and energy

We have seen that heat is one way in which energy can be transferred from one object or system to another object or system. The other way in which energy can be transferred from one object or system to another object or system is by work being done. Work, W, is energy transferred by the action of a force. When a force, F, acts on an object and moves the object through some distance, s, in the same direction as the force, the energy transferred to the object is equal to:

$$W = F \times s$$

This is the work done by the object applying the force. It has units of newton times metre (N m), which is equivalent to $\text{kg m s}^{-2}\text{m} = \text{kg m}^2\text{s}^{-2}$. This is the unit J, as we expect for any form of energy.

Work is energy transferred by the action of a force.
Heat is energy transferred due to a temperature difference.

The rate at which energy is transferred, either by heat or work, is called **power**. Power, P, is energy, E, transferred per unit time:

$$P = \frac{E}{t}$$

Power has units of J s^{-1}, which is given the name watt (W) after the Scottish engineer James Watt (1736–1819), who did important work on developing steam engines.

Figure 2.38 ▲
George Stephenson (1781–1848) built one of the first efficient steam engines in 1829 called 'Stephenson's Rocket'.

Figure 2.39 ▲
Coal heats the boiler to produce the steam that drives the train engine's pistons.

The power of steam drove the Industrial Revolution of the 18th and 19th centuries. It made mining, manufacturing, travel and transport very much more effective. Water was pumped from mines more efficiently so mines could be dug deeper. Long-distance travel and transport by rail and water improved markedly. Mass production of goods in factories concentrated employment in cities.

Demand for fuels for the energy needs of steam engines rose sharply. Employment patterns changed as new jobs were created and many of the older ones disappeared. Steam continues to be used today for electricity production and manufacturing.

Many of the models and theories of thermodynamics were also developed in parallel with the steam engine. This is an example of the interplay between theory, experiment and technology. Advances in any one of these usually leads to advances in the other two.

Consider the steam engine of a train as an example of a useful system. Steam engines are external combustion engines. Coal or wood or some other fuel is burnt (combusted) outside the engine. If we define our system as the engine, then the engine is not an isolated system – energy both enters and leaves the system.

Energy enters the system as heat, Q, due to the temperature difference between the combustion chamber (fire box) and the engine. This increases the energy of the system by an amount:

$$\Delta E = +Q_{in}$$

This heat is used to boil water in the boiler which creates steam. The steam is hot and at high pressure. It pushes on pistons that, in turn, push on the wheels, which push on the ground and make the engine move, pulling the carriages behind it. Hence the system is doing work by applying a force. The energy transferred out of the system is:

$$\Delta E = -W$$

The negative sign indicates that energy is lost from the system. If no energy is lost as heat, then the total energy change is:

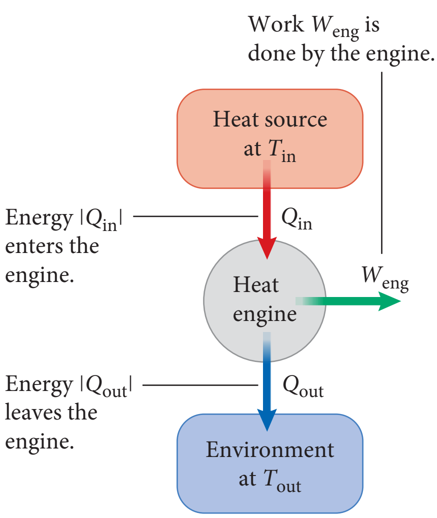
$$\Delta E = Q_{in} - W$$

In an ideal engine with 100% efficiency, the work done would be equal to the heat input. However there is no such thing as an ideal engine. There is always some heat lost from the system to the environment, $-Q_{out}$. Hence we can write our energy equation for any real engine as:

$$\Delta E = Q_{in} - Q_{out} - W$$

This tells us that the change in energy to the system is equal to the net heat added to the system minus the work done by the system. If the net heat in ($Q_{in} - Q_{out}$) and the work done (W) are equal, then the total change in internal energy will be zero. This is the case for an engine working at constant temperature; in other words, once it has reached its stable operating temperature.

Figure 2.40 shows these energy transfers. The heat source (the fire box) supplies heat to boil the water to make steam. Heat energy Q_{in} moves from here into the engine. Work is done by the engine; this is energy leaving the system. Heat is also lost by the system to its surroundings; this is Q_{out}. We call a system that converts heat into work a **heat engine**. Steam engines and petrol and diesel car engines, as well as the petrol engines that power generators and pumps, are all examples of heat engines.



◀ Figure 2.40
A schematic representation of a heat engine

Now let us consider another system – one of the carriages. The carriage, to a first approximation, is not gaining or losing heat. However, if it is being pulled by the engine then a force is applied to it and work is done. In this case the change in energy is positive, energy is coming into the system, hence the sign of the work is positive:

$$\Delta E = +W$$

In general, any energy change, whether it is heat or work, is positive for energy coming into a system and negative for energy leaving the system.

It is important to carefully define our system boundaries so that we know which sign to use. If we define our system as the universe, which is an isolated system, then the total energy change must always be zero. The heat going into our steam engine must have come from somewhere. It came from the internal energy of the fuel that was burned. This loss of internal energy to the environment outside the engine is equal to the gain in heat inside the engine. The total energy of the universe is conserved.

Efficiency of a heat engine

There are many different types of engines, but they all work by converting heat from some fuel source into useful work. How much of the heat is converted into useful work is characterised by the efficiency, η, of the engine. Recall from page 45 that:

$$\eta = \frac{W}{Q_{in}} \times 100\% = \frac{\text{energy output}}{\text{energy input}} \times 100\%$$

Improving the efficiency of the engines that we use, particularly car engines, is very important. Fossil fuels are a non-renewable resource, and currently most of our petrol comes from fossil fuels. So using more efficient engines that consume less fuel to travel the same distance is a very desirable thing. In addition, cars are a major source of greenhouse gases. As we have seen, these gases contribute to the warming effect of the atmosphere. Decreasing greenhouse emissions by developing and using more efficient engines reduces the anthropogenic effects on the climate.

Heat exchange and conversion systems

A **heat-exchange system** transfers heat from a warmer to a cooler location. For example, numerous capillaries in the human nasal passages maintain a temperature below core temperature. This means that air leaving the nose is cooled and air entering the nose is warmed.

A **heat-conversion system** transforms the internal energy of a system. For example, air expelled from a person's wide-open mouth feels warm, but air blown through a smaller hole feels cooler. The decrease in hole size increases pressure in the mouth. When the air is released through a small hole, it undergoes adiabatic cooling as it expands rapidly, does work on the surrounding air, transfers its energy and cools rapidly. Try it yourself.

Heat exchange and evaporative cooling are common in nature and are used to regulate body temperature. Many modern-day cooling and heating systems use heat exchange, state change, energy release and capture, and energy conversion systems. These processes can control temperature and do useful work.

Countercurrent heat exchanger

'Countercurrent' heat exchangers occur naturally in the circulation systems of fish, whales and other marine mammals. Arteries carrying warm blood from the heart to the skin are intertwined with veins carrying cool blood to the heart from the skin. This allows the warm arterial blood to exchange heat with the cooler blood in the veins. This reduces the overall heat loss in cold waters. Baleen whales sieve vast quantities of cold water to extract plankton and other foods. Their tongues have effective heat exchangers that reduce heat losses. Their blubber is also an effective guard against conductive losses. Wading birds have a similar system that limits heat losses to water from their legs.

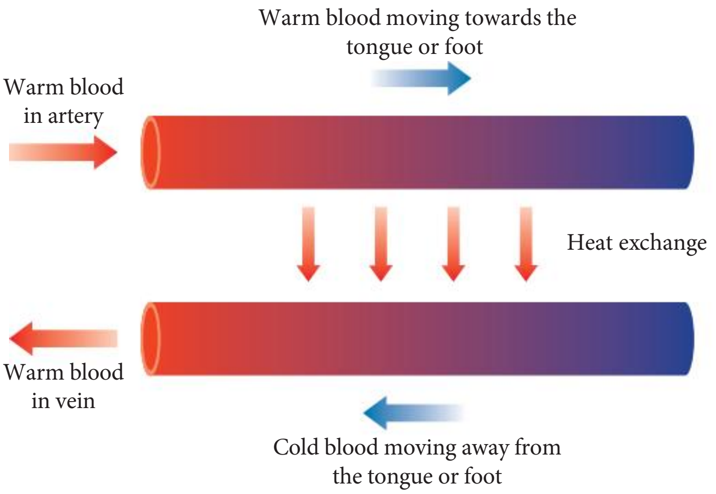


Figure 2.41 ▲ Countercurrent heat exchange between blood vessels reduces heat loss in some animals.

An extreme example of countercurrent heat exchange occurs in the feet of emperor penguins, where arteries and veins are intertwined. The emperor penguins breed on the Antarctic ice during winter. Warm blood in the arteries supplying the feet transfers some of its heat to the veins returning blood to the body. The returning blood brings the gathered heat back to the heart. This helps penguins maintain their core temperature in freezing conditions, while allowing enough warmth to the feet to stop them freezing.

Reverse-cycle heating and cooling

During winter a reverse-cycle air conditioner extracts heat from the outside air, even on very cold nights. The evaporator coil is maintained at a much colder temperature than out side. Energy is transferred from the warmer, though very cold, outside air to the colder evaporator coil. This energy is then transferred inside. In summer, by clever design, this cycle can be reversed and heat is extracted from the house and transferred outside. These systems are fully contained and are relatively cost efficient.

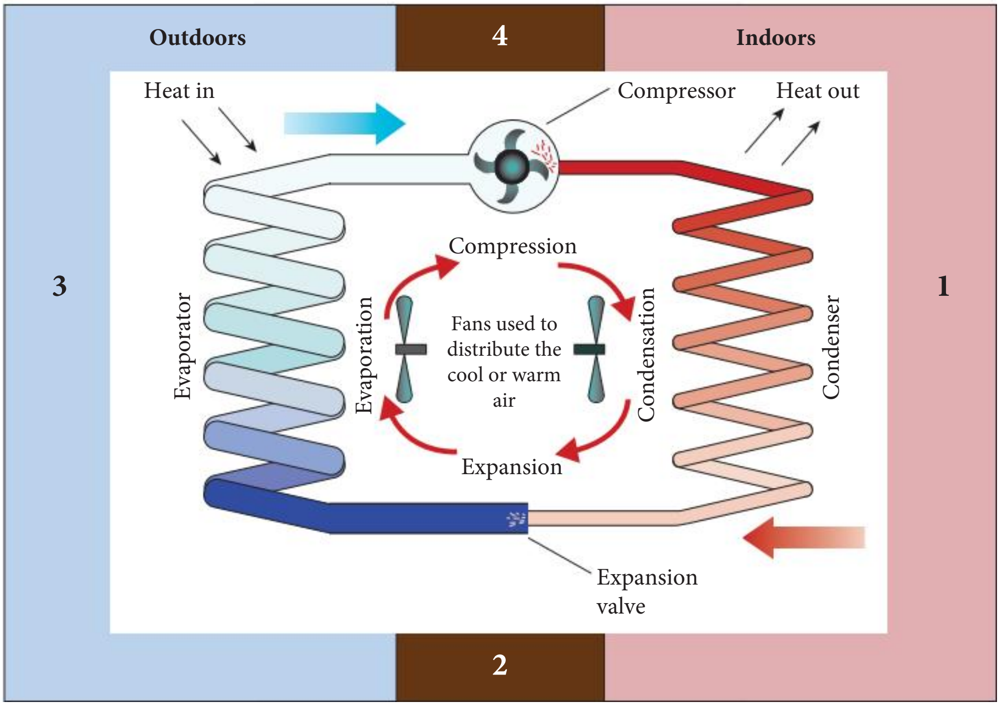


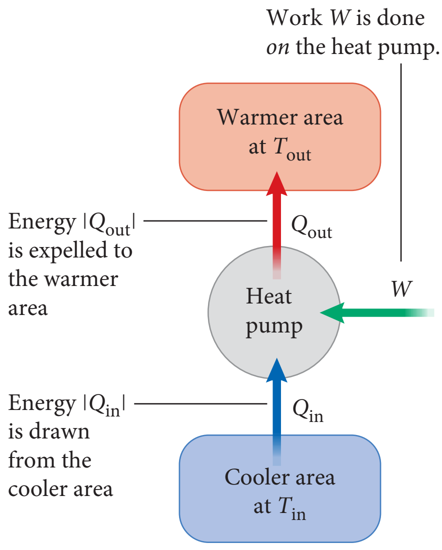
Figure 2.42 ▶ The heating cycle of a reverse-cycle air conditioner. This schematic diagram shows the energy transfers from the outdoor set of coils to the indoor set of coils by a repeating cycle of compression and expansion of a refrigerant.

As the cold gas refrigerant is passed through an external copper coil between (2) and (4) in Figure 2.42, it absorbs heat by collision from the cool outside air particles. The refrigerant gas is then pumped through a compressor (4) where it is compressed and turns from a cool gas into a hot liquid. The compressor has transferred energy to the particles of the refrigerant by forcing them closer together. This increases the internal energy of the compressed refrigerant in the pump. Its temperature increases.

The hot liquid passes into a copper coil with a large surface area, the condenser (1). The hot liquid in the condenser radiates heat energy into the room. The cooler refrigerant liquid continues to pass along the condenser's copper coil until it reaches a constriction at (2) called an expansion valve. As the refrigerant passes through the constriction it expands rapidly into the evaporator. This adiabatic expansion means that the internal energy is re-balanced so that potential energy increases while the kinetic energy, hence temperature, decreases.

The expansion of the vapour in the expansion coil causes rapid cooling of the refrigerant. The cold expansion coil again absorbs heat energy from the outside, warming the coil and refrigerant. The refrigerant is then pumped back into the condenser, starting another heat-exchange cycle. The cycle will continue as long as the compressor continues to operate.

Figure 2.43 shows the energy transfers that occur in this system. Heat Q_{in} enters the system and heat Q_{out} leaves the system. Work, W, is done *on* the system (see page 62).



▲ **Figure 2.43**
A schematic representation of a heat pump

The net effect is that heat is moved from a cooler area to a warmer area. This is impossible without some energy being used to accomplish it. This energy is supplied by the compressor. Work is done on the system by the compressor. This work is used to move the heat energy from the cooler area to the warmer area. The energy supplied to the compressor for a fridge or reverse-cycle heater is usually the electric potential energy that powers the motor.

This sort of system is called a **heat pump**, because it moves energy (heat) from one place to another.

A refrigerator is a heat pump that cools the inside volume, including the food and containers. Work must be done to remove heat from inside the refrigerator. The work is done by an electric motor that compresses the refrigerant gas that passes through external coils. This causes the temperature in the coils to rise so that they radiate energy away into the cooler room (see Figure 2.44). The gas now passes through an expansion valve into coils in the refrigerator, where it becomes cooler than the materials inside the refrigerator. Heat is transferred to the refrigerant, which cycles back via compression to the outside, and so on. The temperature is kept at an appropriate level by a sensor circuit.

▲ **Figure 2.44**
The back of a household refrigerator. The air surrounding the coils is the heat source.

The effectiveness of a heat pump, its coefficient of performance (COP), is calculated as the ratio of heat removed, Q, to the work done, W:

$$\text{COP} = \frac{Q}{W}$$

A 100 W refrigerator that removes 200 W of heat has a COP of 2.0:

$$\text{COP} = \frac{200\,\text{W}}{100\,\text{W}} = \frac{200\,\text{J s}^{-1}}{100\,\text{J s}^{-1}} = 2.0$$

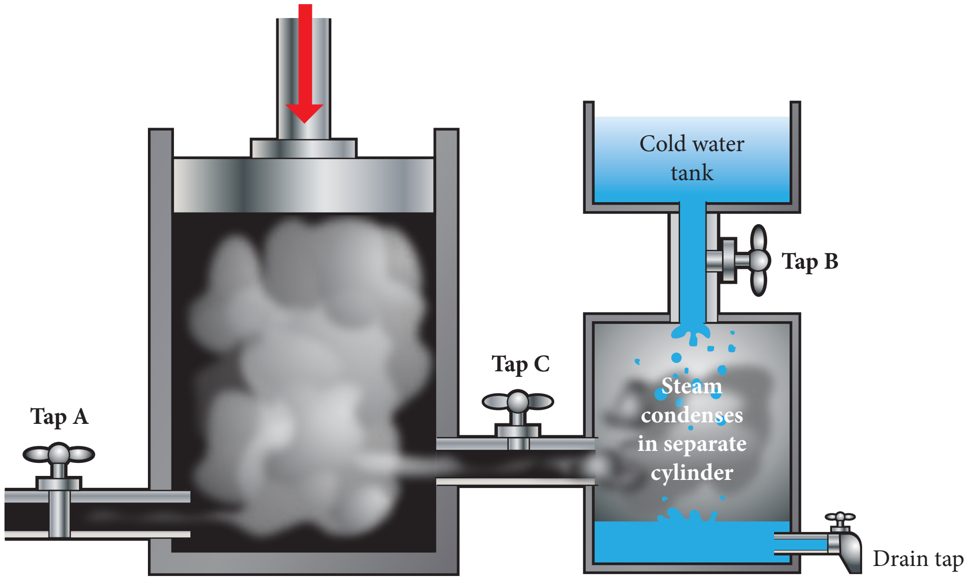
Reverse-cycle heat pumps can act in the same way as refrigerators: they cool the room in summer and send the heat outside or heat the room in winter by cooling the air outside. Coolers that use this process often drip water because the expansion coil gets very cold, causing water vapour in the surrounding air to condense.

The external combustion engine

An external steam combustion engine is a heat engine that uses **superheated steam** under pressure as the working fluid to drive a piston. The water is heated outside the piston cylinder to temperatures well above boiling point. The rapidly expanding high pressure steam transfers energy to the piston to move it. The steam is then expelled from the cylinder where it is condensed, reheated back into steam and recycled.

Figure 2.45 shows how an early James Watt (1736–1819) external condenser steam engine worked. Taps B and C were closed and steam was introduced into the cylinder through tap A, which was open. This pushed

Figure 2.45 ▶
James Watt's external combustion engine with an external condenser



the piston up the cylinder. Then tap A was closed and tap C was opened, allowing steam to escape under pressure into the **condenser**. This reduced the pressure under the piston, and air pressure and gravitational force caused the piston to fall. This expelled all the steam into a separate, external cylinder. Cold water was then added into this steam through tap B to condense it.

The piston's cylinder was always hot under these conditions, so fuel was not required to heat it again before steam was reintroduced. This also meant that the steam could be allowed to expand into the cylinder to do work, rather than continuously feeding in the steam. This was a saving on the quantity of steam needed and, hence, the amount of fuel used. This early design has been developed into the highly efficient engines that are used today, mainly in the generation of electrical energy.

The internal combustion engine

Almost all cars currently use what is called a four-stroke combustion cycle to convert the chemical energy in the petrol into motion. The four-stroke engine was invented by Nikolaus Otto (1832–91) in 1867. The four strokes are the intake stroke, the compression stroke, the combustion stroke and the exhaust stroke. High-energy fuels, such as petrol, contain large amounts of chemical energy. When they combust with oxygen they release this energy mainly in the form of heat. When small amounts of petrol are ignited with oxygen in a confined space, large amounts of energy are released by the hot expanding gases. These expanding gases can do work if they are produced in the cylinders of a piston-driven engine.

The piston is connected to the crankshaft by a rod. As the crankshaft revolves, it uses lifters to open and close the valves at the correct time during each cycle. The engine can complete the cycle thousands of times per minute.

Let us follow the four strokes of the engine cycle. We shall start with the piston at the top. The intake valve opens and the petrol–air mixture is drawn in as the piston moves down. The second step begins as the piston moves up, compressing the mixture. When the piston reaches the top the third stage begins. In this stage the compressed mixture is ignited by the sparkplug or electronic ignition system, causing a powerful explosion that pushes the piston down. This is known as the power stroke. Once the piston reaches the bottom of its stroke, the exhaust valve opens. The exhaust gases are then expelled into the exhaust pipe as the piston moves up. This completes the cycle.

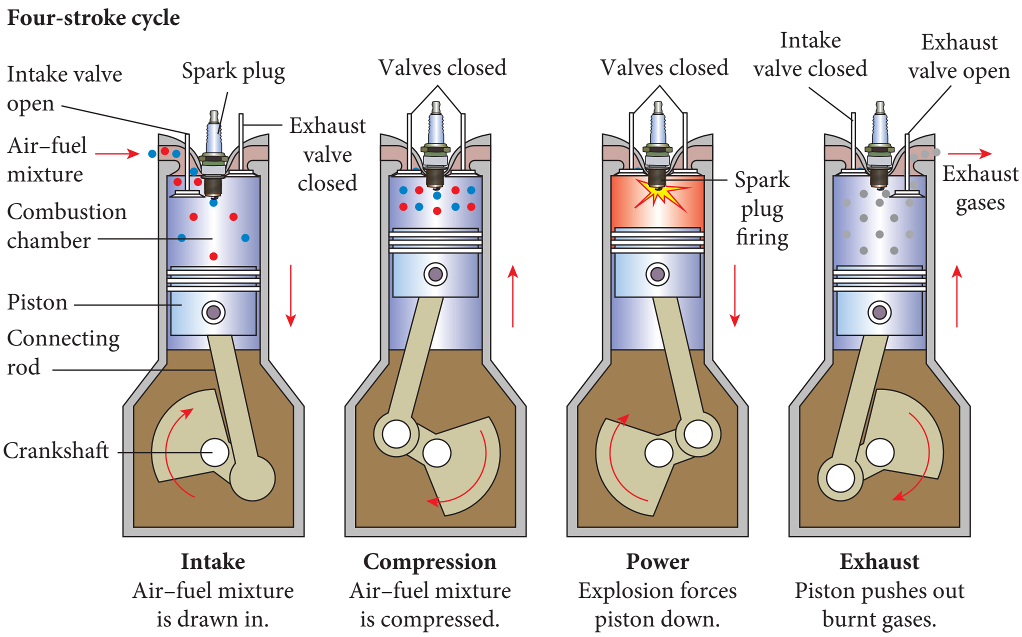


Figure 2.46 The four-stroke cycle internal combustion engine

QUESTION SET 2.5

Remembering

1 What is a heat-exchange system? How does it differ from a heat-conversion system?

2 What is the difference between an external and an internal combustion engine?

Understanding

3 Explain why opening a working refrigerator door will not cool the room it is in.

4 Use diagrams to describe how a four-stroke engine works.

Applying

5 Modern internal combustion engines are about 35% efficient.
- a What happens to the remaining 65% of energy?
- b How can the waste energy be used?

6 Use the kinetic particle model to explain why the air expelled from a wide-open mouth is warmer than air expelled though a smaller mouth opening.

7 How does a countercurrent heat exchanger work? Give an example from the animal kingdom.

8 A heat engine does 100 kJ of work when 250 kJ of heat energy is transferred into it.
- a How much energy is lost as heat by the engine?
- b What is the efficiency of this engine?

9 A refrigerator is rated at 120 W. Most of the energy used by this refrigerator goes into running its compressor, which removes heat at a rate of 560 W from inside the refrigerator.
- a What is the coefficient of performance of this refrigerator?
- b How much heat energy is radiated into the kitchen in which the refrigerator sits?

Analysing

10 James Watt's external condensation combustion engine (see Figure 2.45), improved on earlier models in which the steam was condensed in the cylinder.
- a To what does the expanding steam transfer energy?
- b Why was this engine more efficient than earlier models?

Low-temperature physics

Low-temperature physics, cryogenics (from the Greek meaning 'producing cold'), studies the behaviour of materials at very low temperatures.

Achieving low temperatures

Low temperatures (123 K to 0 K) are achieved by removing heat energy. Liquid nitrogen and liquid air, for example, are produced by repeated compressions and repeated removal of the heat generated by the compressions. Each time this happens the gas becomes cooler until it condenses to its liquid state.

Once these low-temperature liquids have been formed, other matter can be cooled by immersing it in the low-temperature liquid. Heat from matter to be cooled is transferred by conduction to the lower temperature substance it is immersed in. Liquid air and liquid nitrogen are commonly used in science, medicine and industry for cooling to low temperatures. Electronic devices, for example the charge-coupled device (CCD) image sensors used in astronomical telescopes, are cooled to reduce thermal noise, which affects the identification and measurement of very small electrical signals. When a liquefied gas expands back to the gas phase, evaporative cooling occurs. This enables a system to cool significantly. For example, liquid helium boils at 4.2 K but can reach temperatures of less than 1 K by this evaporative cooling.

It is theoretically impossible to reach absolute zero by a finite number of processes. Temperatures as close as one-millionth of a degree above absolute zero are achieved by using the heat energy associated with magnetisation. Using adiabatic nuclear de-magnetisation, temperatures of one-trillionth of a degree kelvin have been achieved.

WOW

The world's fastest train

Magnetic levitation, or maglev, trains use supercooled supermagnets to raise and propel the train. The train floats on air so friction is reduced. The ride is smooth and quiet. Starting and stopping are better than on conventional trains. In 2009, the world's fastest train, the Japanese JR-Maglev train, was clocked travelling at 581 $km\,h^{-1}$.

Shutterstock.com/mamahoohooba

▲ **Figure 2.47** JR-Maglev train uses supercooled supermagnets to levitate the train. This is the Maglev train at the Shanghai Pudung International Airport.

THE WORLD'S FASTEST TRAIN

View how the JR-Maglev train works and its record-breaking run.

Superfluidity

Superfluidity occurs in liquid helium at 2.17 K. Part of the liquid becomes a 'superfluid', a fluid with zero viscosity (the resistance to movement through the fluid) that can move rapidly through any pore in the apparatus. The superfluid liquid can flow over the sides of any container it is placed in, without being stopped by friction or gravity.

Superconductivity

Superconductivity occurs when the electrical resistance of a material drops to zero at a low temperature. The temperature at which this occurs – the **critical temperature** – can be as 'high' as 100 K, and depends on the material. Superconductivity was discovered by Dutch physicist Heike Kamerlingh Onnes (1853–1926) in 1911.

Powerful superconducting magnets are used in MRI (magnetic resonance imaging) machines that give very detailed internal images of the body. Supermagnets have been used to improve control and accuracy in scientific measuring instruments. Research into superconductors for power transmission appears promising because the loss of energy by heating of transmission wires would no longer occur.

QUESTION SET 2.6

Remembering

1 What branch of physics investigates low-temperature phenomena?

2 What property of a material changes when a liquid becomes a superfluid?

Understanding

3 How does releasing liquid nitrogen cool a CCD image sensor?

4 Use the kinetic particle model to explain why it is impossible to reach absolute zero.

Applying

5 Electrical energy is lost during electricity transmission due to the internal resistance of the transmission wires. How could we use superconductors to solve this?

Analysing

6 Use the kinetic particle model to explain how liquid nitrogen is produced from gaseous nitrogen.

7 What are the main advantages of magnetic levitation used by the maglev train?

CHAPTER SUMMARY

- Law of conservation of energy: Energy cannot be created or destroyed, only transformed and transferred. The total energy of an isolated system is constant.
- Neither energy nor matter can enter or leave an isolated system.
- An open system is one that both matter and energy can enter and leave.
- Matter cannot enter or leave a closed system, but energy can.
- Energy can be transferred and transformed.
- Energy as heat is transferred from regions of higher temperature to regions of lower temperature.
- There are three methods of heat transfer: conduction, convection and radiation.
  - Conduction is the transfer of heat energy through a material by particle collision.
  - Convection is the transfer of heat energy through the bulk movement of fluids.
  - Radiation is the transfer of heat energy without the need of a medium.
- Energy can be transferred by the action of a force. The amount of energy transferred is called the work.
- When a system does work, its internal energy decreases. When work is done on a system, its internal energy increases.
- Pockets of contained air are good insulators.
- Metals are good conductors because they contain electrons that are free to move.
- Earth's climate is driven by surface and atmospheric interactions with incoming solar energy.
- The high specific heat capacity of water has a significant effect on climate.
- Anthropogenic changes to greenhouse gas concentrations cause global warming.
- Heat engines convert heat energy into work. Heat energy is transferred into a heat engine, and the engine does work by applying a force to some part of its environment.
- The efficiency η of a heat engine is:

  $$\eta = \frac{W}{Q_{in}} \times 100\% = \frac{\text{energy output}}{\text{energy input}} \times 100\%$$

- A heat pump transfers energy from a cooler region to a warmer region. Work must be done to make this happen. It cannot happen spontaneously.
- Rapid gas expansion and compression are used in refrigeration and reverse cycle air conditioning to move energy from one region to another.
- Superconductivity shows great promise as a way to reduce the cost of electricity transportation.

CHAPTER GLOSSARY

adiabatic process a process that occurs without the exchange of heat between a system and its environment.

closed system a system that matter cannot enter or leave, but energy can

condenser a vessel that removes heat from steam allowing it to condense back to water

conduction the process by which heat energy is transferred by the collision of particles

convection the transfer of energy through the movement of the cooler more dense fluids displacing less dense warmer fluids

convection cell the condition that occurs when there are density differences within a body of liquid or gas; the density differences result in rising and/or falling currents

convection current fluid circulating as a result of heating at a point or localised region; movement of fluids due to convection

Coriolis effect a deflection of moving objects when they are viewed in a rotating reference frame

critical temperature the temperature at which the electrical resistance of a material becomes zero

delocalised valence electrons the outer electrons of metal atoms that are free to move

Earth's energy balance the balance occurring when the amount of solar energy absorbed by Earth is equal to the amount of energy radiated back into space

electromagnetic radiation energy that travels as waves and moves at the speed of light

energy efficiency, η the effectiveness by which one form of energy is transformed into the desired energy

heat conductor a material that readily allows the transfer of heat

heat-conversion system a system in which heat is transformed into another form of energy

heat engine a system that converts heat energy into work

heat-exchange system a system in which heat is exchanged between two or more bodies of matter

heat insulators a material that is a poor conductor of heat

heat pump a system that moves heat energy from a cooler to a warmer area; work must be done on a heat pump

high-grade energy resource energy form that can be converted into other energy forms with a high rate of efficiency

insolation incoming solar radiation

isolated system a system that neither energy nor matter can enter or leave

low-grade energy resource energy form that is converted into other energy forms with a low rate of efficiency

open system a system that both matter and energy can enter and leave

power the rate at which work is being done on an object or the rate of energy transfer

radiation the process by which heat is transferred in the absence of a medium

sensible heat heat energy causing change in temperature

superconductivity occurs when electrical resistance of a material drops to zero at a temperature below its critical temperature

superfluidity the property a material has when it has zero viscosity

superheated steam steam under high pressure that has been heated to a temperature higher than the boiling point of water

thermal conductivity a measurement of how efficiently heat can be conducted through a material ($W\,m^{-1}\,K^{-1}$)

thermal mass a large mass of matter that can absorb and retain heat energy

thermals rising air columns caused by convection

thermohaline circulation the continuous circulation of sea water due to variations in temperature and salinity

CHAPTER REVIEW QUESTIONS

Remembering

1 What state of matter is the most efficient in the conduction of heat energy?
2 Define 'conduction', 'convection' and 'radiation'.
3 Write the work–energy equation for any real engine. Define each term.
4 Define 'Earth's energy balance'.
5 What is the difference between a heat engine and a heat pump?

Understanding

6 How do isolated, open and closed systems differ?
7 In which direction and at what angle should solar hot water panels face in Brisbane, to operate at maximum efficiency?
8 Why are bar heaters more effective than convection heaters in a room that has large doors that are frequently opened?
9 Explain what role the air trapped between the two panes of glass in double-glazed windows plays in controlling heat flow.
10 Why is Earth's atmosphere described as a blanket?
11 Why do good-quality frying pans have plastic handles and thick metal bases?
12 What were the major advantages of Watt's external condenser steam engine over the earlier engine that condensed the steam inside the cylinder?
13 When you leave your car outside on a sunny day, the internal temperature climbs rapidly. Is this an example of the greenhouse effect? Explain.
14 Why are sea breezes more likely in summer?
15 Why is it advantageous to have the steam driving a turbine in a power plant as hot as possible?
16 Explain how the compressor does work on the gas in a refrigerator. Where does this energy to do the work come from?

Applying

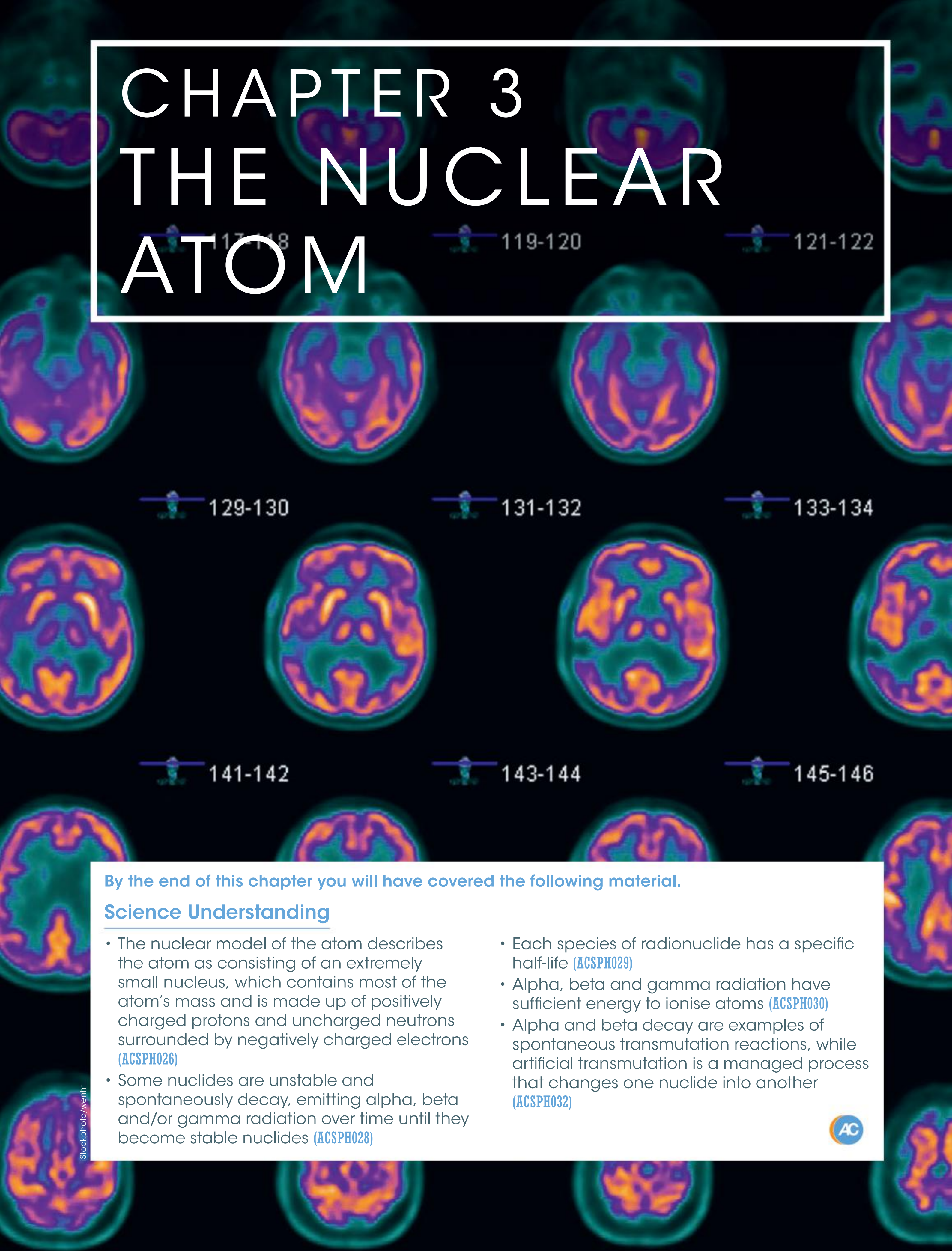
17 All the energy lost from Earth to space is radiant energy. Why is there no loss due to conduction and convection?
18 Air conditioners use a pump and a refrigerant to move energy from the inside of the house to the outside of the house. Where would you place the condenser coil and where would you place the expansion coil? Give reasons for your answers.

Analysing

19 How do you think the total amounts of incoming solar radiation and outgoing radiation from Earth will be affected by:
   a a large volcanic eruption?
   b the continents moving closer to the poles?
   c an increase in the amount of greenhouse gases in the atmosphere?
   d an increase in the amount of water vapour in the atmosphere?
   e an increase in the cloud coverage?
20 The atmosphere of Venus is composed of about 96% carbon dioxide, with most of the remainder being nitrogen. The atmosphere appears to be relatively clear until the cloud deck composed of sulfuric acid starts about 50 km above the surface. The surface temperatures on Venus are around 460°C, exceeding those of Mercury, which is closer to the Sun. Why is Venus hotter than Mercury?
21 Would adding a light reflective layer to the inner layer of outdoor winter clothing help you keep warm? Explain your answer.
22 Explain why air that is sinking in the atmosphere gets warmer.
23 What effect would the melting of the polar ice have on the carbon dioxide levels in the atmosphere?

Reflecting

24 As a result of what you have learnt in this chapter, list three practical things you could do at home to reduce greenhouse gas emission into the atmosphere.



iStockphoto/wenht

CHAPTER 3
THE NUCLEAR ATOM

By the end of this chapter you will have covered the following material.

Science Understanding

- The nuclear model of the atom describes the atom as consisting of an extremely small nucleus, which contains most of the atom's mass and is made up of positively charged protons and uncharged neutrons surrounded by negatively charged electrons (ACSPH026)
- Some nuclides are unstable and spontaneously decay, emitting alpha, beta and/or gamma radiation over time until they become stable nuclides (ACSPH028)
- Each species of radionuclide has a specific half-life (ACSPH029)
- Alpha, beta and gamma radiation have sufficient energy to ionise atoms (ACSPH030)
- Alpha and beta decay are examples of spontaneous transmutation reactions, while artificial transmutation is a managed process that changes one nuclide into another (ACSPH032)

Introduction

Radiation that comes from the nucleus of an atom is called **radioactivity**. Radioactivity is used extensively in medical diagnosis and treatment. Every house is fitted with smoke detectors that use a radioactive source. Leaks in pipes are traced using radioactivity. Radioactivity is involved in nuclear power production and, most unfortunately, the production of weapons of mass destruction.

In this chapter, we develop the particle model of matter further by investigating the internal structure of the atom. As we do so, we will discover more particles and processes. These include protons, neutrons and electrons; however, we will also find that we need to include a number of other particles, such as photons, neutrinos, antineutrinos and positrons. These particles are necessary to account for the mass and energy involved in radioactive decay and nuclear reactions.

Radiation

Generally, **radiation** refers to energy transfer that occurs across empty space. Sunlight, for example, is radiation that travels from the Sun to Earth. It is experienced in a variety of ways, such as light and heat. X-radiation is emitted by stars. Radiation can be classified in several ways. One way is to distinguish between the radiation that affects the charge on atoms and the radiation that does not. Some radiation has its origin in Earth, other radiation bombards us from outer space and the upper atmosphere.

Non-ionising and ionising radiation

Radiation is also used to describe the emanations from radioactive atoms. **Gamma rays** are electromagnetic in nature, but alpha and beta radiation are charged particles with mass.

Some forms of radiation can ionise atoms by removing electrons. Radiation comes in two forms: **non-ionising radiation** and **ionising radiation**. Radiation from radioactive sources is ionising radiation, but electromagnetic radiation can be non-ionising or ionising depending on its energy.

Electromagnetic radiation is pure energy with no mass. It is modelled as continuous waves or massless particles called photons that have an associated wavelength. The **electromagnetic spectrum** goes from high-energy gamma rays to lower-energy radio waves.

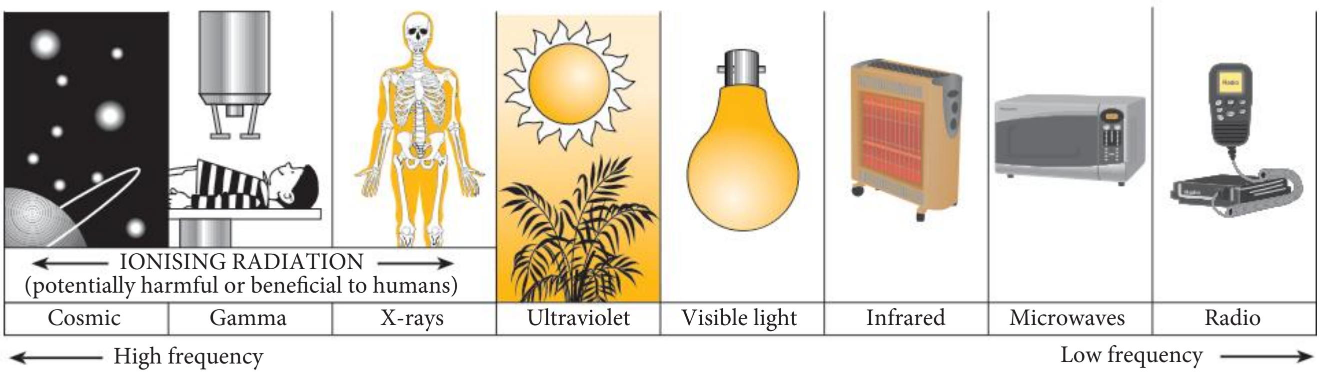


Figure 3.1 ▼ Electromagnetic radiation can be non-ionising (low energy) or ionising (high energy).

Non-ionising radiation

Non-ionising radiation is electromagnetic radiation with low energy. The radiation from electric powerlines and microwave and TV appliances, as well as sunlight and electric globes, is non-ionising radiation. These may affect our bodies for good or ill, but they do not change the electron configuration around the nuclei of atoms.

Ionising radiation

Ionising radiation is high-energy radiation that can affect the electrons surrounding an atom so that a charged ion is formed. Ionising radiation includes alpha particles, beta particles, gamma rays and X-rays. The high energy of these radiations can affect our bodies in significant and unwanted ways, including changing the electron configurations, types of atoms and radioactive materials in our bodies.

We are constantly bombarded by low-level, background ionising radiation.

Background radiation

There are two types of **background radiation**: terrestrial radiation and cosmic radiation.

Terrestrial radiation comes from the decay of radioactive elements such as uranium and thorium in the Earth's crust. The energy from **radioactive decay** of these materials is one of the reasons for the temperature of Earth. Terrestrial radiation can enter our food chain via the naturally occurring radioactive chemicals in soils.

Cosmic radiation comes to us from space. It comprises mainly protons that interact with Earth's atmosphere to produce cosmic showers of radiation, some of which reaches Earth's surface.

The background radiation on Earth's surface varies from place to place. It depends on altitude and proximity to radioactive minerals. Radioactive fallout from nuclear tests and damaged nuclear power stations also adds to background radiation. Radiation doses from these vary with location and are usually small, unless you are close to the site of a nuclear accident or nuclear explosion test site.

COSMIC RADIATION EXPOSURE

Learn about how you are exposed to more cosmic radiation when you are flying.

EXPERIMENT 3.1

BACKGROUND IONISING RADIATION

Ionising radiation comes from terrestrial and cosmic sources. There is a normal level that can be measured.

Aims

To investigate the random nature of background radiation

To measure the background radiation rate

Materials

- Geiger-Muller tube (G-M tube)

For more information on G-M tubes see page 93.

Procedure

Set up the G-M tube to record counts every 15 seconds.

Results

Record the readings in a properly constructed data table.

Analysis of results

1 Produce a frequency table from the data table.
2 Plot a graph of count rate versus time.
3 Show the mean and standard deviation on the graph. (The standard deviation is a measure of the spread of the data.)

Discussion

1 How does the graph of frequency against count rate demonstrate the random nature of the background radiation?
2 Provide an estimate of the average background radiation per minute.

QUESTION SET 3.1

Remembering

1 What are the two main forms of radiation?

2 What are the two origins of background radiation?

Understanding

3 Explain why electromagnetic radiation is classified as both ionising and non-ionising radiation.

Applying

4 How does terrestrial radiation keep Earth's surface warm?

5 In radiation counting experiments, why is it useful to know the background radiation? What must be done to the data before it is analysed?

Analysing

6 Why might airline pilots need to monitor their exposure to ionising radiation?

Reflecting

7 State one way in which your understanding of radiation has changed.

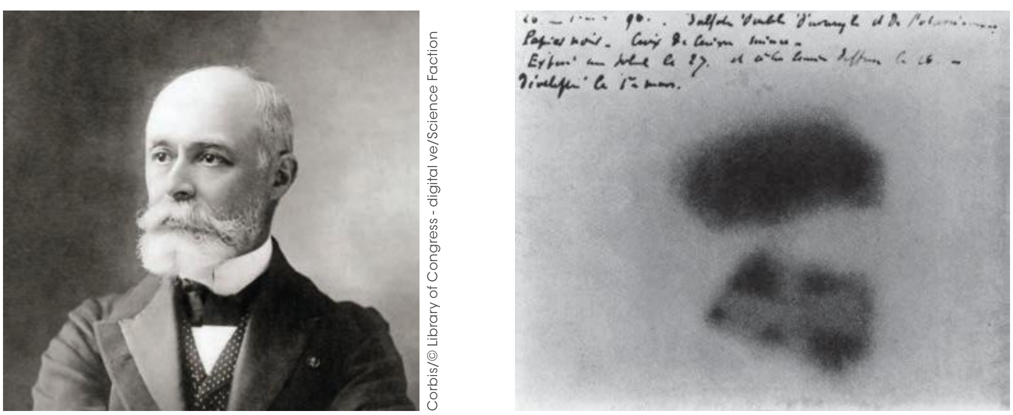
Nuclear model of the atom

In Chapter 1, we used a simple particle model to explain heating processes – particles move about, bump into each other and transfer kinetic energy.

The word, **atom**, was proposed by Democritus (460 BCE–370 BCE) to refer to the smallest, indivisible particle of matter. Democritus believed that matter was composed of these particles. His views were not altogether accepted in his lifetime, or for a long time afterwards. Nevertheless, as people developed and refined their understanding of chemical substances and energy-transfer processes, this powerful idea began to be used to explain a range of experimental results.

Nuclear radiations and the atomic model

In 1896, Henri Becquerel (1852–1908) discovered that uranium salts emitted a previously unknown form of radiation.



Corbis/© Library of Congress - digital ve/Science Faction

Science Photo Library

Figure 3.2 ▶ Henri Becquerel and the first evidence for radiation (Becquerel rays) from a uranium salt

Becquerel called the radiation 'metal phosphorescence' – his training was in studies of phosphorescence – and he thought of these emanations as an invisible type of light.

Becquerel showed that atoms were capable of emitting smaller particles. Atoms must be divisible. This was the beginning of investigations into radioactivity and the use of nuclear radiations to investigate the structure of the atom. Just prior to Becquerel's discovery, Roentgen had discovered X-rays, initially referred to as Roentgen rays.

The new Becquerel radiation was soon shown to be different from X-rays. The first of these radiations to be identified was called **alpha particles**, the second, **beta particles** and the third, gamma rays after the first three letters in the Greek alphabet – α, β and γ. Alpha particles were extensively studied and used by the New Zealander, Ernest Rutherford (1871–1937), after he discovered them while working at Cambridge in 1899. Becquerel identified beta particles as electrons in 1900. Paul Villard (1860–1934), however, is widely credited with identifying the differences between gamma rays, other nuclear radiations and X-rays in 1900.

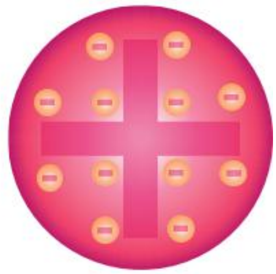
JJ Thomson (1856–1940) discovered the electron in 1897, a year after radioactivity was discovered. Electrons could be emitted from many metals, which meant they were more fundamental particles than an atom. He demonstrated that the mass of the electron was 9.1×10^{-31} kg, a minute fraction of the mass of the hydrogen atom. Electrons have a charge of -1.6×10^{-19} C.

Thomson proposed a model for the atom, comprising a uniformly positively charged region of 'electrification' in which negatively charged electrons were distributed. This model was dubbed the 'plum pudding' model, because the electrons appeared to be like raisins stuck in a plum pudding. Thomson started by considering the electrons to be uniformly distributed. However, in 1904 he published a mathematical physics theory paper in which he showed that the electrons could be arranged in concentric rings. He even discussed how many electrons could be fitted into each ring.

Alpha particles have a positive charge and beta particles have a smaller negative charge. These results were first demonstrated by applying electric and magnetic fields to streams of alpha and beta particles.

BECQUEREL AND RADIOACTIVITY

Click on the Becquerel Bio button. Read the citation and summarise the key points.



▲ **Figure 3.3** Thomson's 'plum pudding' model of the atom shows electrons in a positively charged sphere of electrification.

WOW

Sir JJ Thomson

Sir Joseph John Thomson – he was knighted in 1908 – was influential in the field of subatomic physics from 1890 until about 1930. In 1876, he won a scholarship to Cambridge University where, from 1903 to 1919, he was head of the Cavendish Laboratory. His two major experimental discoveries were the electron and two isotopes of neon, Ne.

Thomson was a great teacher and lecturer – seven of his students went on to win Nobel Prizes – but he did not agree with the atomic models of Rutherford or Bohr. His inability to take them seriously led Rutherford and Bohr to work at other universities.

JJ THOMSON AND THE ELECTRON

Click on the JJ Thomson button. Watch the film. Find and read the related Nobel Prize citation. What were Thomson's two main points in this film?

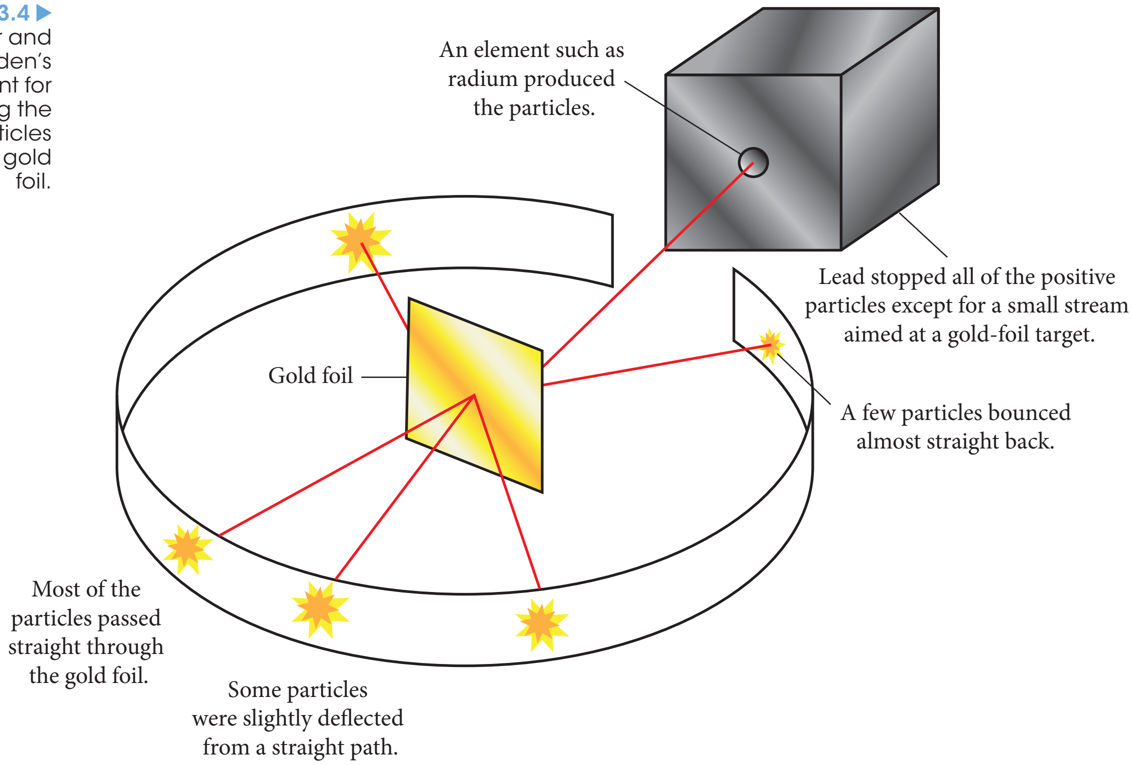
Rutherford develops a new atomic model

Ernest Rutherford pioneered experiments with the positively charged alpha particles, which he used to probe matter at the atomic scale.

In 1909, Rutherford designed an experiment to test Thomson's plum pudding model. If the model was correct, very energetic, fast-moving alpha particles should experience small deflections from the negatively charged particles but come through more or less in a straight line. The possibility of them deflecting through 90 degrees or more was almost zero. Hans Geiger (1882–1945) and Ernest Marsden (1889–1970) carried out the experiment in 1909. Alpha particles from a radium source were fired through a vacuum tube at a sheet of gold foil that was very thin, about 6×10^{-7} m (0.6 μm).

The end of the vacuum tube was coated with zinc sulfide, which makes a tiny **scintillation** (spark) whenever an alpha particle strikes it. They used a small, low-power microscope to see the scintillations more clearly. By moving the microscope they were able to investigate the number and rate at which the alpha particles came straight through the foil or were scattered.

Figure 3.4 ▶ Geiger and Marsden's equipment for counting the alpha particles fired at a thin gold foil.



GEIGER–MARSDEN EXPERIMENT

Work through the simulation to see how Rutherford obtained evidence for his model of the atom.

RUTHERFORD'S INTERPRETATION

Learn more about the Geiger–Marsden experiment and Rutherford's interpretation.

They found the following:

- Most of the alpha particles went straight through the gold foil as though it were empty space. (This was the expected result according to Thomson's model.)
- Some alpha particles came through but were deflected by relatively small, but measurable, angles. (This was unexpected.)
- About 1 in 20 000 alpha particles was scattered almost straight back towards the source. (This was really puzzling!)

When they repeated the experiment with different metals, they obtained similar results.

Until this experiment, physicists thought that the atom was a sphere filled with a continuous electric fluid and populated by very small electron particles. Rutherford and others showed mathematically that the central positive region was about 20 000 times smaller in diameter than the atom. It was this last result that convinced physicists that the model of the atom must be changed to accommodate the new experimental evidence. By 1911 Rutherford was convinced that the evidence did not support Thomson's model.

Now, Rutherford began to think of a refinement of Thomson's model of the atom. He decided that the atom must have a very dense centre that was positively charged, but that most of the atom was lightweight, negatively charged space in which the electrons circulated like planets around the Sun.

Mostly, the positively charged alpha particles were unaffected by this dense, positively charged atomic centre or nucleus. They could travel more or less unaffected through the negatively charged space. But, sometimes, they came close to an atomic centre and were deflected. In some cases, they collided head-on with a nucleus and were scattered almost directly back.

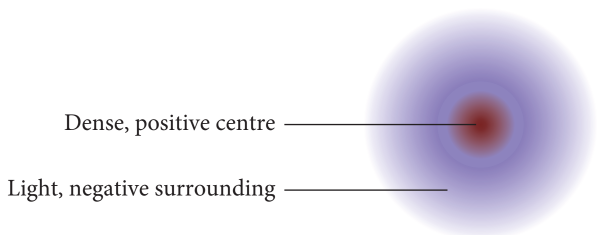
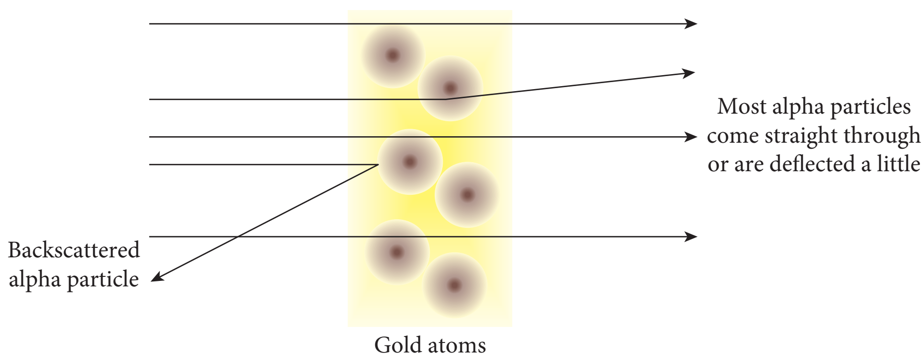


Figure 3.5 ▶ Rutherford's refinement of Thomson's atomic model, which he soon abandoned



◀ **Figure 3.6**
Rutherford's explanation of back-scattering in the Geiger–Muller experiment.

There was another problem to solve. Rutherford's calculations showed that the mass of the positively charged part of the nucleus accounted for about one-half or less of the mass of the nucleus. This suggested that there was another mass, yet to be identified, in the nucleus. He was also concerned to explain where the electrons were situated in the atom.

RUTHERFORD'S DISCOVERIES

Click on the Rutherford Atom button. Read the article. For what discoveries was Rutherford nominated for the Nobel Prize? Why did he get only one Nobel?

WOW

Sir Ernest Rutherford

Rutherford grew up on a farm in New Zealand. He pioneered the transmission of radio waves, used later by Enrico Marconi to develop over-the-horizon wireless communication. After gaining a scholarship to Cambridge University, Rutherford worked with JJ Thomson. He became professor of physics in Montreal, Canada, but in 1908, Arthur Schuster, the then head of physics at Manchester University, resigned because he believed that England needed Rutherford. He was right! Rutherford returned to England to take up Schuster's post. He was knighted in 1914. He led the Cavendish Laboratory from 1919 to 1937, its most influential period.

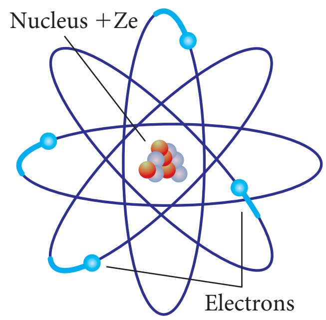
SIR MARCUS OLIPHANT

Use the weblinks to make a timeline to show Oliphant's significant contributions to early particle discoveries.

The Rutherford–Bohr atomic model

Rutherford decided to do away with Thomson's negatively charged region, opting for a planetary model. In this model, the atom comprised a central nucleus in which almost all the mass of the atom was concentrated, as well as the positive charge. Electrons moved around this nucleus in circular orbits.

A new problem now arose. When a charged particle, such as an electron, goes around in a circle it must radiate electromagnetic energy and lose speed. This effect was well understood in Rutherford's time. From 1860 to 1871, the Scotsman James Clerk Maxwell (1831–79) unified the study of electricity and magnetism with that of light into **classical electrodynamics**. Maxwell's equations were, and are, exceptionally effective in describing, explaining and predicting a vast array of electromagnetic phenomena. As the electron slows down and loses energy, the radius of its orbit should get smaller and a full spectrum of colours should be emitted. The electron should spiral into the centre. Calculations showed that this would take about a billionth of a second (10^{-9} s). There was no evidence of the predicted phenomena. Rutherford's model atom was impossible!



▲ **Figure 3.7**
Rutherford's planetary model of the atom (Z is the atomic or proton number)

Niels Bohr (1885–1962), a Dane who worked extensively with Rutherford, came to the rescue. He used his intuition to suggest that insights from a new form of physics – **quantum physics** – might be helpful. Bohr put forward a model in which electrons were only allowed to have specific energies. We call these specific energies the energy states of the electron. While in those energy states, they would not radiate energy. In the lowest energy state they could only accept energy. If they were in higher energy states, they could either accept energy or release energy, but only in **discrete quanta**, by moving between states.

Even though it was not consistent with Maxwell's classical electrodynamics, the Rutherford–Bohr model, as it became known, was very successful. For example, it explained the specific colours of radiation from heated metals (spectra).

The simple Rutherford–Bohr model of the atom proposes that the atom comprises a nucleus surrounded by electrons. The electrons revolve in shells located at specific distances from the nucleus. They are bound to the nucleus in specific energy levels. Electrons nearest the nucleus are most tightly bound, and those further away, especially those in outer shells, are less tightly bound. It is the outer or valence electrons that take part in chemical reactions.

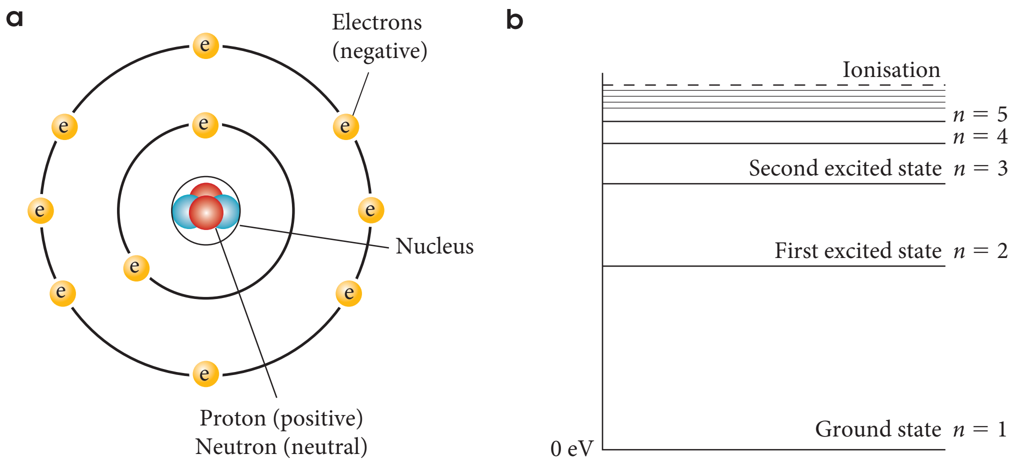


Figure 3.8 a) The Rutherford–Bohr planetary shell model of the atom and b) Rutherford–Bohr energy level model

Refer to page 109 for an explanation of electron volt, eV.

Another way to imagine the Rutherford-Bohr atom is as a ladder with uneven steps (shells). The lowest step is on the ground; hence, the ground state. Higher rungs correspond to higher energies. Electrons cannot exist between the rungs. If they fall down, they must release energy equivalent to the energy difference between the rungs. If they jump up to a higher level, they must accept only the right amount of energy.

The Rutherford–Bohr atom

- **The nucleus contains most of the mass.**
- **The nuclear charge is positive and equal in size to the total electronic charge.**
- **Electrons exist in orbitals that correspond to allowed energy states.**
- **The atom is much bigger than the nucleus.**

ACTIVITY 3.1

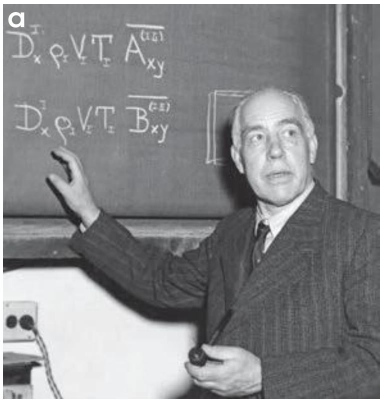
THE GOLDEN AGE OF PHYSICS

1. Create a timeline of scientific development from 1880 to 1930 that includes the year, name of scientist, country of origin and main contribution. Include a column to add interesting scientific or biographical information. (You could add to this timeline as you meet other scientists involved in the development of the nuclear model of the atom.)
2. Explain how the timeline demonstrates the importance of international cooperation in scientific developments during this period.
3. CP Snow, in his 1981 book *The Physicists*, said that from the late 19th century to 1930 'nearly all the scientists … tended to expect other human beings to be as free from class and racial tensions as they were themselves' (p. 76). How did this way of approaching science help to develop knowledge about the atom and radioactivity? In making your case, refer to the timeline, particular physicists and their work.

WOW

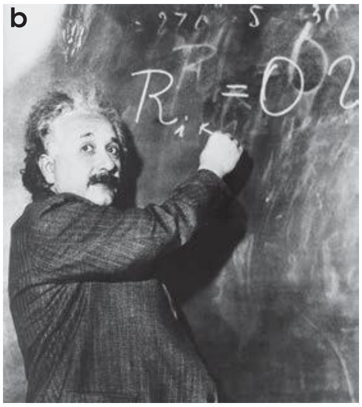
The genius of Niels Bohr

a



Science Photo Library/Princeton University/American Institute of Physics

b



Corbis/© Bettmann

▲ Figure 3.9 a) Niels Bohr (Princeton University, 1948) and b) Albert Einstein (California Institute of Technology, 1931)

The most notable scientist of the twentieth century, Albert Einstein (1879–1955), wrote in 1949 that 'this insecure and contradictory foundation [of the old quantum theory] was sufficient to enable a man of Bohr's unique instinct and sensitivity to discover the principle laws of the spectral lines and of the electron shells of the atoms, together with their significance for chemistry, appeared to me as a miracle – and appears to me a miracle even today. This is the highest musicality in the sphere of thought'.

QUESTION SET 3.2

Remembering

1 a Name each of the three forms of nuclear radiation.
  b Write the symbol for each form.
  c Give the charge of each form.

2 Draw:
  a the planetary model of the atom.
  b the quantum 'ladder' model.

Understanding

3 What was the significance for the model of the atom of Becquerel's 'metal phosphorescence'?

4 Describe the Geiger–Marsden experiment. What evidence from this experiment was used by Rutherford to develop his first atomic model?

Analysing

5 Describe the differences between the atomic models proposed by Thomson, Rutherford and Bohr.

6 Why did Rutherford's atomic model fail? How was Bohr's model different?

Nuclides and the periodic table

In 1911, the Dutch physicist Johannes van den Broek (1870–1926) suggested that the elements in the periodic table should be arranged according to the positive charge on the nucleus, rather than their atomic weights. X-ray studies of atoms by Henry Moseley (1887–1915) in 1914 supported this hypothesis. From the data, Moseley proposed laws to sort chemical elements on a

WOW

Gallipoli and the Nobel Prize

Moseley's work was outstanding and exceptional. He could have been a candidate for the 1916 Nobel Prize but for one thing. He was killed at Gallipoli in August 1915. What a waste!

consistent, physical basis. As a result, the organisation of the periodic table by Dmitri Mendeleev (1834–1907) was refined. Moseley's work was also influential in supporting the Rutherford–Bohr model with accurate physical data.

Atomic number and atomic mass number

In 1919, Rutherford bombarded nitrogen atoms with α particles. Positively charged hydrogen nuclei, **protons**, were emitted. The protons were subatomic particles with a positive charge of $+1.6 \times 10^{-19}$ C and about 1800 times more massive than electrons. In the periodic table, the number of protons in the nucleus, the **atomic number (*Z*)**, is now used to identify elements.

According to Rutherford, the radioactive emission of electrons suggested that a proton in a nucleus might have an electron attached to it so strongly as to form another particle. This neutral particle – the **neutron** – was hard find by deflection in electric and magnetic fields. Subsequently, James Chadwick (1891–1974) discovered the neutron in 1932. The mass of a neutron is slightly greater than the mass of the proton. The neutron is now considered a unique particle, *not* a proton–electron combination.

Naming nuclides

Protons and neutrons are collectively called **nucleons**. The **atomic mass number (*A*)** is the nucleon number. The atomic number (Z) is the proton number. The number of neutrons (N) in the nucleus of an atom is the difference between the mass number (A) and the atomic number (Z):

$$N = A - Z$$

Table 3.1

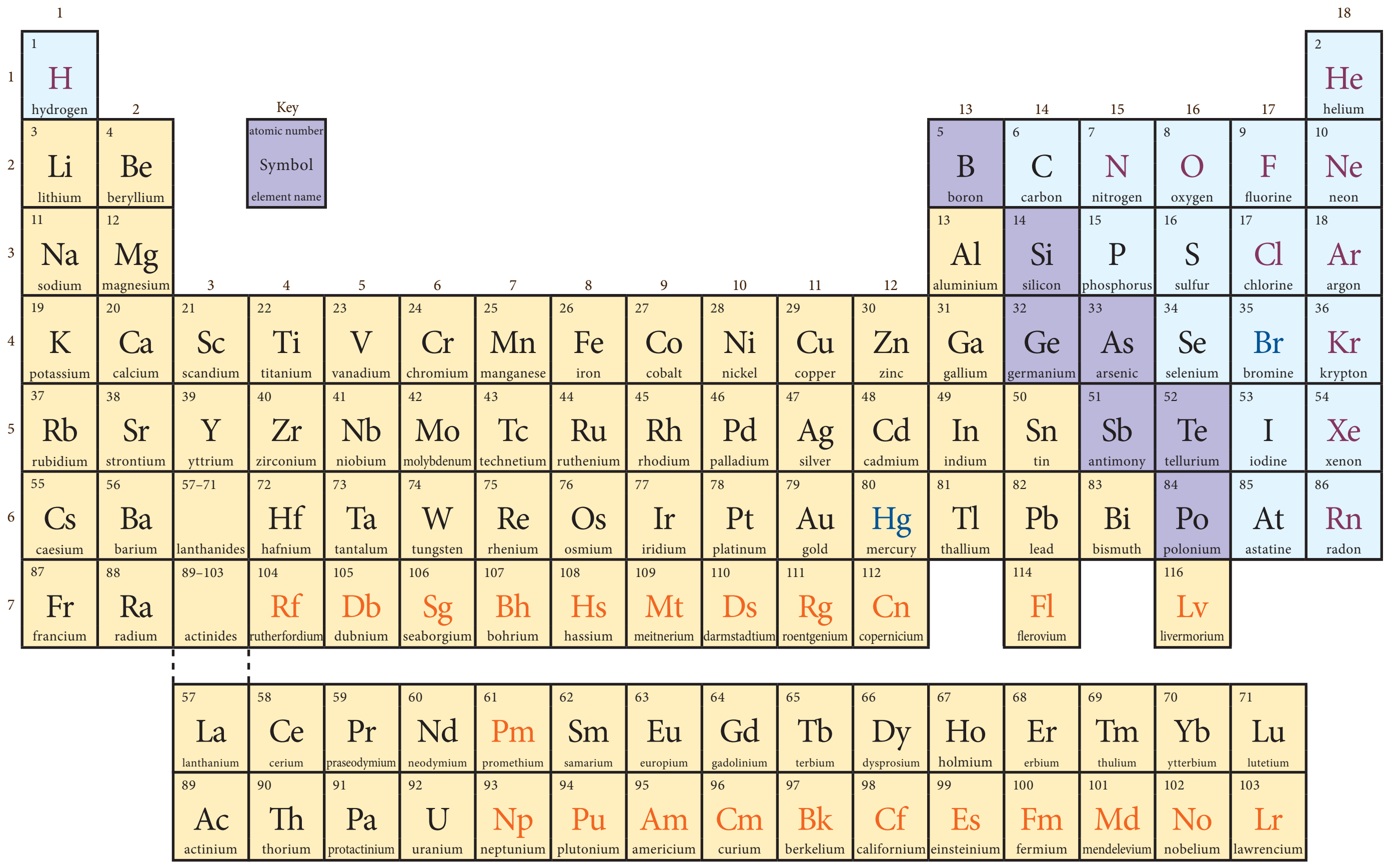
Symbol	Name	Description
A	Mass number (nucleon number)	Number of protons and neutrons in the atom
Z	Atomic number (proton number)	Number of protons
N	Number of neutrons	$N = A - Z$

The number of protons (atomic number) is used to name atoms. For example, all atoms with 8 protons are called oxygen. All atoms with 79 protons are called gold. The periodic table in Figure 3.10 shows the names and symbols for each of the elements, sorted according to their proton number.

Nuclides, 'the element' and isotopes

A **nuclide** is a species of atom classified according to the number of protons and neutrons as well as its energy state. Like atoms, the energy level shell model is used to explain energy transitions in the nucleus. The most stable nuclides exist in their lowest or ground state. Nuclides that are in excited energy states above the ground state are unstable and radioactively decay until a stable nuclide is formed. The energies involved are far greater than for atomic energy-level transitions.

A substance that only has nuclides with the same number of protons is called an **element**. Unfortunately, there is some ambiguity about the use of the word 'element'. When you take a naturally occurring sample of an element, you typically find several different nuclides or **isotopes**. Each has the same number of protons but different numbers of neutrons. The nuclide that is most common is sometimes called 'the element', but this is not the standard meaning of element. All nuclides with the same number of protons are isotopes of the same element.



▲ **Figure 3.10** The periodic table of elements

Figure 3.11 shows the international standard notation for representing a nuclide. The chemical symbol has the atomic mass (nucleon number) as a left superscript and the atomic (proton) number as a left subscript.

The nuclide of beryllium-9, ${}^{9}_{4}\mathrm{Be}$, contains 4 protons and 9 nucleons. The number of neutrons is $N = A - Z = 9 - 4 = 5$.

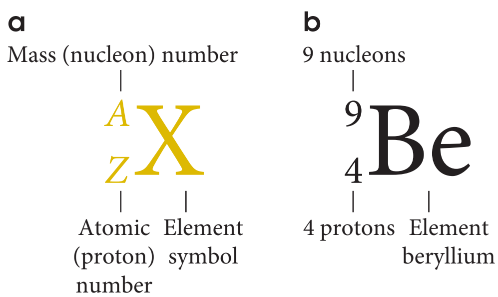
Nuclides can be sorted into families in a number of ways: number of protons (isotope), number of nucleons (**isobar**), number of neutrons (**isotone**) and energy state (**isomer**). This last, energy state, is quite important. If a nuclide can exist in an energy state above its ground state for more than 10^{-12} s it is called a **metastable nuclide**.

Table 3.2 Nuclear families: isotopes, isobars, isotones and isomers

Families	Nuclides with the same:
Isotopes	atomic (proton) number, *Z*
Isobars	mass (nucleon) number, *A*
Isotones	number of neutrons, *A* – *Z*
Isomers	*Z* and *A*, but different energy states

Technetium-99m is a metastable nuclide that is produced from molybdenum-98 by neutron absorption. The resultant molybdenum-99 undergoes beta decays to a relatively long-lived energy state in technetium-99, hence ${}^{99m}\mathrm{Tc}$. Technetium interacts with many different chemicals in the human body. It emits a gamma ray of 140 keV (140 000 eV) energy as it transitions to ground state isomerically. It can be injected to diagnose bone and other disorders in which calcium is involved. A collector on the outside of the body observes the 140 keV gamma ray. Tc-99m is cheap to prepare,

▼ **Figure 3.11** International standard notation for representing a nuclide



ISOTOPES AND RADIOACTIVITY

Use the two weblinks to investigate all the isotopes of elements. Track the radioactivity of different isotopes.

has a wide range of uses, produces a single gamma ray in the optimum energy range for medical application, and becomes relatively harmless in relatively small time periods.

Atomic weight (relative atomic mass)

In a pure sample of an element there is usually more than one nuclide or isotope of the element. The **atomic weight** or **relative atomic mass** is the weighted average of the masses of the different nuclides in a pure, naturally occurring sample of the element. For example, pure silver contains 51.84% of isotope $^{107}_{47}Ag$ and 48.16% of $^{109}_{47}Ag$. Its atomic weight is 107.96. Atomic weight (relative atomic mass) must not be confused with atomic mass, atomic mass number or the atomic number.

WORKED EXAMPLE 3.1

Pure silver contains 51.84% of isotope $^{107}_{47}Ag$ and 48.16% of $^{109}_{47}Ag$. Find its atomic weight. (3 marks)

Answer	Logic	
Weighted average $= 51.84\% \times 107 + 48.16\% \times 109$	Substitute the known values.	2 marks
$= 0.5184 \times 107 + 0.4816 \times 109$		
$= 107.96$	Calculate the answer.	1 mark

Try these yourself

1 Pure boron contains 80.22% of isotope $^{11}_{5}B$ and 19.78% of $^{10}_{5}B$. Find its atomic weight. (3 marks)

2 Pure antimony has an atomic number of 51. It contains 57.25% of antimony-121 and 42.75% of antimony-123. Find its atomic weight. (3 marks)

The **atomic mass** of a single nuclide is the mass compared to the mass of a single ^{12}C nuclide. This mass is given in unified mass units (u), or Dalton (Da). It is assigned the value 1. Compared to this, a silicon-28 nuclide has an atomic mass of 27.976 926 532 46 u. This is not the same as the atomic mass (nucleon number) of 28. Nor is it the atomic weight, 28.11, which is the weighted average of all the atomic masses of the nuclides in a naturally occurring sample. The atomic weight depends on the sample. Boron found in California has a different atomic weight from that found in Turkey, because their isotopic compositions differ. These local differences can be used in forensic analyses for criminal, archaeological and historical studies.

Table 3.3 Summary of terms

Atomic number (*Z*)	Number of protons in a nucleus
Atomic mass number (*A*)	Number of nucleons in a single nuclide
Unified mass unit (u or Da)	Mass of a single carbon-12 nuclide, assigned the value 1 (1 u = 1.6605×10^{-27} kg)
Atomic mass	Mass of a nuclide compared to a single carbon-12 nuclide
Atomic unit (amu)	Obsolete unit originally based on oxygen-18 instead of carbon-12. (If used, must be related to carbon-12)
Atomic weight	Weighted average of all naturally occurring nuclides of an element in a sample
Relative atomic mass	Atomic weight

QUESTION SET 3.3

Remembering

1 For the general nuclide, $^{A}_{Z}X$, what do the letters *A*, X and *Z* represent?

2 Define:

a atomic number.

b unified atomic mass unit (u).

Understanding

3 Distinguish between atomic mass number and atomic weight.

4 How did Moseley's work alter the basis on which the periodic table of the elements is organised?

5 Use Figure 3.12 to model how molybdenum-99m decays to the relatively stable state Tc^{99m} above ground, then decays isomerically.

Applying

6 Find the number of neutrons in $^{136}_{57}La$.

7 Europium comes in two forms, $^{153}_{63}Eu$ (52.18%) and $^{151}_{63}Eu$ (47.82%). Find the atomic weight of europium.

Analysing

8 An isotope of an element is sometimes referred to as 'the element'. Explain why this can lead to confusion about the terms 'element' and 'isotope'.

9 Molybdenum-99 is formed when a nuclide absorbs a neutron. What is that nuclide?

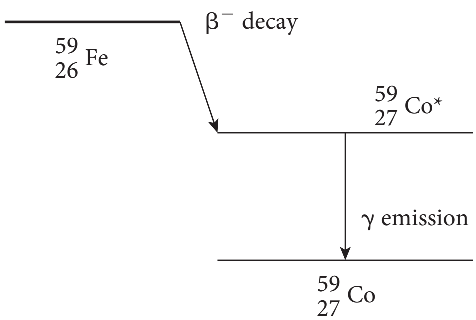
Reflecting

10 How has your understanding of the use of language in physics changed? Refer to the use of 'element', 'isotope' and 'nuclide' in your response.

Radioactive decay

Radioactive atoms can emit α or β particles that have mass, and massless γ rays. As a result of α and β emission, the radioactive atom is transformed into an atom of a different chemical element. Experiments show that the changes in radioactive decay are different from chemical changes. In both nuclear and chemical changes, energy is produced; however, nuclear processes release far more energy than chemical processes.

When the energy conditions are right in a nuclide, it can enter a transitional state with a higher than stable energy. It then releases the energy as a particle (α decay or β decay) or as electromagnetic radiation (γ ray). Figure 3.12 shows the decay of inon-59 to the unstable transition state cobalt-59*. This then γ-decays to stable cobolt-59. Depending on the radioactive decay sequence, **neutrinos** or **antineutrinos** are also emitted.



▲ **Figure 3.12** A beta decay followed by gamma emission results in a stable nuclide in its ground state.

Radioactive decay is a random, spontaneous and uncontrollable process involving the nucleus of an atom. Uranium and thorium are examples of naturally occurring radioactive substances. Artificial radioisotopes are routinely made for medical and industrial applications.

When a radioactive nucleus emits an alpha or beta particle, it breaks into two parts – the lighter, emitted particle and a new nuclide of a different element. This is because, in radioactive decay, it is the nucleus of the atom that changes. In a gamma emitter, the same nucleus emerges, but at a lower energy level. When a nuclide of one element (the **parent nuclide**) decays, it becomes a nuclide of a different element (the **daughter nuclide**). The daughter nuclide and the emitted particles are the decay products.

Radioactive decay is also called **nuclear transformation**, **disintegration** and **transmutation**.

Types of radioactivity

Alpha particles are helium nuclei. Beta particles may be electrons or their anti-particle equivalent, positrons. They are assigned the charge –1 for electrons and +1 for positrons, but they have no nucleon number. Gamma rays have neither nucleon number nor charge. In all nuclear reactions, the nucleon number remains constant. The net charge is also conserved. The nucleus never takes part in chemical reactions; thus, the nucleus goes wherever the atom goes. It is indestructible by means of chemical reactions.

Before and after a nuclear reaction there is the same:

- number of nucleons.
- net charge.

Table 3.4 Summary of terms

Alpha particle	α, ${}^{4}_{2}He^{2+}$	Helium-4 nuclide
Beta particle	β^-, ${}^{0}_{-1}e$	Electron
	β^+, ${}^{0}_{+1}e$	Positron
Gamma ray	γ, ${}^{0}_{0}\gamma$	Electromagnetic radiation
Neutrino	ν_e, ${}^{0}_{0}\nu_e$	Energy carrier
Antineutrino	$\overline{\nu}_e$, ${}^{0}_{0}\overline{\nu}_e$	Energy carrier

Alpha-decay: ${}^{4}_{2}He$ or ${}^{4}_{2}\alpha$ or α

The α particle is a positively charged helium nucleus, which contains two protons and two neutrons. It has been stripped of its two electrons, so it carries a +2 charge. The most common nuclide of uranium, ${}^{238}_{92}U$, undergoes α-decay, resulting in 90 protons and 234 nucleons in the daughter nucleus. This is thorium, ${}^{234}_{90}Th$. The nuclear reaction equation is:

$$ {}^{238}_{92}U \rightarrow {}^{234}_{90}Th + {}^{4}_{2}He^{2+} $$

In all radioactive decays, energy is released. In this decay reaction, the energy is almost all taken away by the α particle.

The mass numbers balance (238 = 234 + 4) and the atomic numbers balance (92 = 90 + 2).

In general, α-decay can be written as:

$$ {}^{A}_{Z}X \rightarrow {}^{A-4}_{Z-2}Y + {}^{4}_{2}He $$

In a chemical reaction equation, the same atoms in equal numbers appear on both sides of the equation. This is because the identity of the atoms does not change. However, in a nuclear reaction equation involving alpha emission, the same atomic symbols do not appear on both sides of the equation. This is because new nuclides are formed as a result of the emission of alpha particles. In a nuclear reaction equation it is the same number of nucleons that appears on both sides of the equation.

WORKED EXAMPLE 3.2

Neptunium-237 decays by emitting an α particle and changes to a completely different element.

a Write a complete nuclear reaction equation that includes the symbol for the daughter nuclide. (3 marks)

b What is the name of the daughter nuclide? (1 mark)

Answers	Logic	
a $^{237}_{93}Np \rightarrow ^{233}_{91}Pa + ^{4}_{2}He^{2+}$	Use the correct symbols for all particles, and conservation of nucleon number and proton number.	3 marks
b Pa = protactinium		1 mark

Try these yourself

1 Francium-211 decays by emitting an α particle and changes to a completely different element.

a Write a complete nuclear reaction equation that includes the symbol for the daughter nuclide. (3 marks)

b What is the name of the daughter nuclide? (1 mark)

2 Polonium-213 decays by emitting an α particle and changes to a completely different element.

a Write a complete nuclear reaction equation that includes the symbol for the daughter nuclide. (3 marks)

b What is the name of the daughter nuclide? (1 mark)

Beta decay

There are two forms of beta decay: electron emission and positron emission. An electron has opposite charge to a proton, but it is not a nucleon. Its symbol is written:

$$^{0}_{-1}e \quad \text{or} \quad ^{0}_{-1}\beta \quad \text{or just} \quad \beta^-$$

A positron is an anti-electron. It has the same charge as a proton, but it is not a nucleon. Its symbol is written:

$$^{0}_{+1}e \quad \text{or} \quad ^{0}_{+1}\beta \quad \text{or just} \quad \beta^+$$

The mass of each of the two beta particles, positrons (+) and electrons (−), is very small compared with the masses of nucleons. They have a mass of 9.11×10^{-31} kg, which is tiny compared with the mass of a proton (1.6726×10^{-27} kg) or a neutron (1.6749×10^{-27} kg). This difference in mass was good evidence that, although particles could be emitted from the nucleus, they were not nucleons.

Electron emission or β⁻ decay

The ejection of an electron from the nucleus, β^- decay, can be modelled, *for the time being,* by regarding a neutron as capable of changing into a proton and an electron. Another particle, an uncharged and almost undetectable antineutrino, $\overline{\nu}$, is also released in this nuclear reaction.

This model of β^- decay can be written as:

$$^{1}_{0}n \rightarrow ^{1}_{1}p + ^{0}_{-1}e + \overline{\nu} \quad \text{or} \quad ^{1}_{0}n \rightarrow ^{1}_{1}H + ^{0}_{-1}e + \overline{\nu}$$

When thorium undergoes β^- decay, it produces the nuclide with 91 protons but an unchanged mass number of 234. The new element is protactinium, Pa:

$$^{234}_{90}Th \rightarrow ^{234}_{91}Pa + ^{0}_{-1}e + \overline{\nu}$$

Note that the electron (β^- particle) comes from the nucleus and not from the atomic shell electrons.

When a nucleus emits a β^- particle, the atomic mass (nucleon) number does not change, but the atomic number (Z) of the daughter nuclide increases by 1; that is, the resulting atom belongs to an element one place forward in the periodic table.

In general, β^- decay can be represented as:

$$^{A}_{Z}X \rightarrow ^{A}_{Z+1}Y + ^{0}_{-1}e + \overline{\nu}$$

Several beta-emitters are used as radiopharmaceuticals to treat a range of diseases. The choice of radiopharmaceutical depends on the problem and location. For example, treatment of a melanoma on the skin requires a different approach to a cancer in the bladder or secondary cancers in the bones.

Positron emission or β⁺ decay

The ejection of a positron from the nucleus, β^+ decay, can be modelled by regarding a proton, or hydrogen nuclide, as capable of changing into a neutron and a positron. Another particle, an uncharged and almost undetectable neutrino, ν, is also released in this nuclear reaction.

This model of β^+ decay can be written as:

$$ {}^{1}_{1}\mathrm{p} \rightarrow {}^{1}_{0}\mathrm{n} + {}^{0}_{+1}\mathrm{e} + \nu \quad \text{or} \quad {}^{1}_{1}\mathrm{H} \rightarrow {}^{1}_{0}\mathrm{n} + {}^{0}_{+1}\mathrm{e} + \nu $$

When thallium-195 undergoes β^+ decay, it produces a nuclide with 80 protons but an unchanged mass number of 195. The new element is mercury, Hg:

$$ {}^{195}_{81}\mathrm{Tl} \rightarrow {}^{195}_{80}\mathrm{Hg} + {}^{0}_{+1}\mathrm{e} + \nu $$

Note that the positron (β^+ particle) comes from within the nucleus.

When a nucleus emits a positron, the mass (nucleon) number does not change, but the atomic number (Z) of the daughter nuclide decreases by 1; that is, the resulting atom belongs to an element one place earlier in the periodic table.

In general, positron decay can be represented as:

$$ {}^{A}_{Z}\mathrm{X} \rightarrow {}^{A}_{Z-1}\mathrm{Y} + {}^{0}_{+1}\mathrm{e} + \nu $$

During positron emission tomography (PET), a patient is given a positron emitter, such as ^{18}F-fluorodeoxyglucose (FDG) or oxygen-15 (^{15}O). The selected positron emitter is known to accumulate in an organ or other tissue of interest. The positrons interact with electrons to produce gamma rays of known energy and direction. Collectors identify these gamma rays and compute an image.

A note on models of beta emission

We have been careful to say that a proton can be *converted into* a positron and a neutron, and that a neutron can be *converted into* an electron and a proton. A neutron *does not contain* an electron and a proton, nor does a proton contain a neutron and a positron.

When β radiation from nuclei was first observed it seemed to physicists that this was the case, as it explained how a neutron could emit an electron. However, it later turned out that this could not be correct. More than just an electron is emitted in β^- decay, and there are other reactions involving protons and neutrons that could not be explained by them containing electrons or positrons.

WORKED EXAMPLE 3.3

What daughter nuclide is produced after:

a alpha decay of a polonium-214 nuclide? (2 marks)

b beta-minus decay of carbon-14? (2 marks)

c positron emission of sodium-20? (2 marks)

d gamma emission from cerium-139? (2 marks)

Answers	Logic	
a ${}^{214}_{84}\mathrm{Po} \rightarrow \mathrm{X} + {}^{4}_{2}\mathrm{He}$	Use correct nomenclature for alpha particle.	1 mark
$\Rightarrow {}^{214}_{84}\mathrm{Po} \rightarrow {}^{210}_{82}\mathrm{Pb} + {}^{4}_{2}\mathrm{He}$	Use correct symbol for daughter nuclide.	1 mark
b ${}^{14}_{6}\mathrm{C} \rightarrow \mathrm{X} + {}^{0}_{-1}\mathrm{e}$	Use correct nomenclature for electron.	1 mark
$\Rightarrow {}^{14}_{6}\mathrm{C} \rightarrow {}^{14}_{7}\mathrm{N} + {}^{0}_{-1}\mathrm{e}$	Use correct symbol for daughter nuclide.	1 mark

c $^{20}_{11}Na \rightarrow X + {}^{0}_{-1}e$	Use correct nomenclature for positron particle.	1 mark
$\Rightarrow {}^{20}_{11}Na \rightarrow {}^{20}_{10}N + {}^{0}_{+1}e$	Use correct symbol for daughter nuclide.	1 mark
d $^{139}_{58}Ce^{*} \rightarrow X + {}^{0}_{0}\gamma$	Use correct nomenclature for gamma ray.	1 mark
$\Rightarrow {}^{139}_{58}Ce^{*} \rightarrow {}^{139}_{58}Ce + {}^{0}_{+1}\gamma$	Use correct symbol for daughter nuclide.	1 mark

Try these yourself

1 What daughter nuclide is produced after:

- **a** alpha decay of bismuth-211? (2 marks)
- **b** beta-minus decay of bromine-82? (2 marks)
- **c** positron emission from gold-190? (2 marks)
- **d** gamma emission from samarium-145? (2 marks)

2 What daughter nuclide is produced after:

- **a** alpha decay of lead-204? (2 marks)
- **b** beta-minus decay of lead-209? (2 marks)
- **c** positron emission of lead-199? (2 marks)
- **d** gamma emission from lead-203? (2 marks)

QUESTION SET 3.4

Remembering

1 What two numbers must be the same before and after a nuclear reaction?

2 Write the general equation for:

- **a** alpha decay.
- **b** beta-minus decay.
- **c** positron decay.
- **d** gamma radiation.

Understanding

3 How are nuclear reactions different from chemical reactions? Consider the two sides of the equation and the energy involved.

4 Fluorine-21 is a beta-minus emitter. What is the daughter nuclide?

Applying

5 Holmium-151 decays by emitting an α particle and changes to a completely different element.

- **a** Write a complete nuclear reaction equation that includes the symbol for the daughter nuclide.
- **b** What is the name of the daughter nuclide?

6 Radon-210 decays to polonium-206. What type of radioactive particle is emitted in this decay? Write the decay equation.

Analysing

7 Write the decay equation for these nuclides.

- **a** ^{15}O
- **b** FDG

8 Use correct symbols to show the decay of terbium-158 by alpha emission followed by gamma emission.

9 Gold-198 is used in liver scans. It radiodecays to mercury-198. Write the nuclear decay equation. Identify the radioactive emission.

Reflecting

10 What is one question about radioactivity that you still want to have answered?

Alpha, beta and gamma radiation

Alpha, beta and gamma rays affect materials in different ways according to their power to ionise and penetrate materials.

Ionising power

Atoms become ions by losing or gaining electrons. When electrons are removed from a neutral atom, it becomes a positive ion, and when a neutral atom gains electrons, it becomes a negative ion.

Slow-moving, positively charged α particles attract electrons from atoms and ionise them. With each ionisation, α particles lose energy.

Negatively charged β^- particles are repelled by electrons in atoms. This causes particles to be bounced between atoms. These collisions may cause some electrons to be ejected from atoms, which become ionised. These collisions transfer less energy than the interactions between α particles and atoms.

Positively charged β^+ particles interact with electrons in atoms. The effect of these interactions is that the atoms become ionised. These interactions transfer less energy than the interactions between α particles and atoms.

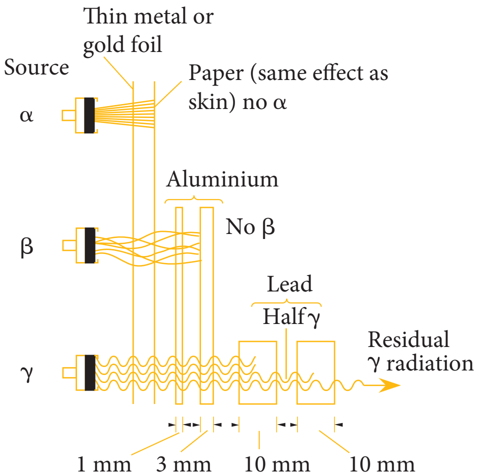
The ionising power of a ray depends upon its energy. High-energy rays can ionise atoms and molecules by colliding with them and transferring energy to their electrons. If an electron gains enough energy it can leave the atom or molecule, leaving behind a positive ion. The electron may then remain free for some time before binding to another atom or molecule to form a negative ion. Low-energy rays are more likely to result in heating of a material than ionisation.

α particles have far more **ionising power** than β particles or γ rays. The ionising power is inversely proportional to the **penetrating power**. This is to be expected because the particles expend their energy in causing ionisation, and so do not penetrate as far. Consequently, the range of α particles in air is much less than that of either β particles or γ rays.

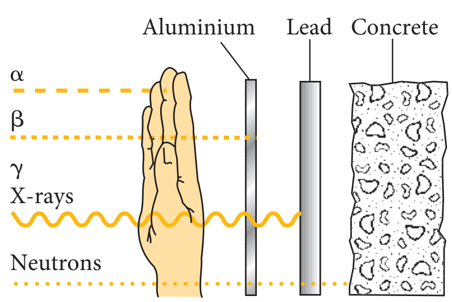
Neutrinos and antineutrinos are weakly interacting particles that do not ionise atoms.

Penetrating power

α particles are the most easily absorbed particles; therefore, the least penetrating. They are absorbed by a thin sheet of paper or outer skin (see Figure 3.13). β particles can be stopped by a few millimetres of aluminium. γ rays can penetrate up to 30 cm of steel. A 1 cm thick sheet of lead reduces the intensity of γ radiation to about half of the original intensity. A comparison of penetrating power is illustrated in Figure 3.14.



▲ Figure 3.13
γ rays are much more penetrating than α or β particles. In turn, β particles are more penetrating than α particles.

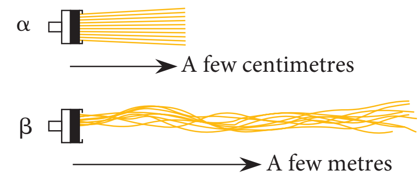


▲ Figure 3.14
Penetrating power for different radiations is indicated by their relative absorptions in materials.

Neutrons are highly penetrating in air and most other materials. They interact with and are absorbed strongly by materials containing a lot of hydrogen. Hence materials such as water and concrete are good neutron absorbers and are used as shielding at nuclear reactions. The core of the Open Pool Australian Lightwater reactor (OPAL) at Lucas Heights in Sydney's south is contained in a large pool of water.

Range in air

There are many ions and other charged particles in air. The greater the charge on a radioactive particle, the more likely it is that it will pick up these charged particles and become neutral. For this reason, α particles penetrate only a few centimetres in air. The range of β particles is difficult to determine, partly because they can have quite widely different energies, but they can certainly travel a few metres in air before stopping. γ radiation is so energetic it is not absorbed in air, therefore has no maximum range in air (see Figure 3.15).



▲ **Figure 3.15**
The range of α and β particles in air. γ rays pass easily through air.

Effects of electric and magnetic fields on radiation

Charged α and β particles moving in straight lines can be deflected in regions subject to electric and magnetic effects. γ-rays have no mass or charge, which is why they are not deflected in either region.

Charged particles are affected by electric and magnetic fields. Electric fields act on all charged particles and magnetic fields act on moving charged particles. Charged particles emitted from nuclei are generally travelling very fast. We can use electric and magnetic fields to distinguish between the different types of radiation.

α, β and γ radiation in electric fields

α and β particles are charged and hence will be affected by an electric field. The α and β^+ particles will be accelerated in the direction of the field, and β^- particles will be accelerated in the opposite direction. This can be observed as a curving of the path of a moving charged particle when it enters an electric field. α and β^+ particles will be deflected in one direction, β^- particles are deflected in the opposite direction. The force depends on the charge of the particle, but the acceleration depends on both the force and the mass, and it decreases with increasing mass. Hence, even though an α particle in an electric field experiences twice the force of a β particle, the curvature of its path is much less because its mass is about 7000 times greater than that of a β particle. A γ particle has no charge, and so will pass through a region of electric field without any **deflection**.

α, β and γ radiation in magnetic fields

A magnetic field applies a force to any moving charged particle, so that it follows a curved path.

The magnitude of the force depends on the speed at which the particle is moving and the magnitude of its charge. The direction of the force depends on the sign of the charge. Hence the force experienced by a β^+ particle is the same size but opposite direction to that experienced by a β^- particle, if they are moving at the same speed. An α particle at the same speed experiences a force in the same direction as a β^+ particle, but the force is twice as large.

How this magnetic force affects the path of the particle depends on the mass of the particle. An α particle has a mass 7000 times that of a β particle. Hence an α and a β^+ particle will be deflected in the same direction by a magnetic field, but the β^+ particle will be deflected far more. A β^- particle will be deflected just as much as a β^+ particle, but in the opposite direction. A γ particle, which has no charge, is not deflected by a magnetic field.

This is shown in Figure 3.16.

▼ **Figure 3.16**
Deflections of α, β and γ radiation in a magnetic field. There is no deflection in the absence of a magnetic field.

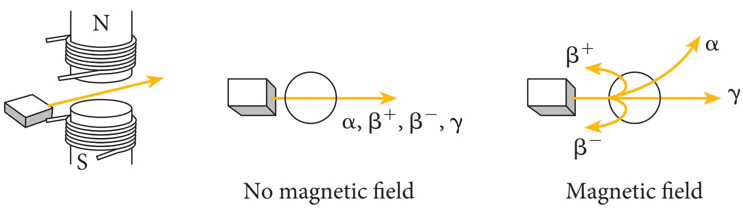


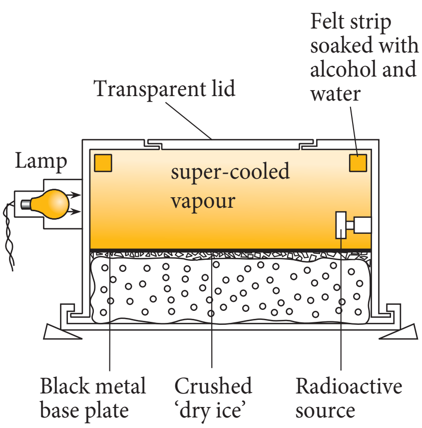
Table 3.5 Properties of α, β and γ radiation

	α particles	β particles	γ ray
Nature	A helium nucleus (i.e. 2 protons and 2 neutrons)	A fast-moving electron or positron	High-frequency (short wavelength) electromagnetic radiation (i.e. a high-energy photon)
Charge	+2 elementary charges	−1 (electron) +1 (positron) elementary charge	Uncharged
Mass	4 atomic mass units (i.e. 4 u); $4 \times 1.66 \times 10^{-27}$ kg	0.0005 u; 9.11×10^{-27} kg	No mass
Ionising effect	Strong	Weak	Very weak
Penetration	Few centimetres in air	Few metres in air	Very weakly absorbed in air (most radiation absorbed by a few centimetres of lead)
Effect of electric and magnetic fields	Very small deflection	Large deflection	No deflection
Effect on photographic plate	Blackens	Blackens	Blackens
Typical emission velocity	5–7% of speed of light	30–90% of speed of light	Speed of light (3×10^8 m s^{-1})

Detection of radioactivity

Radioactive decay radiation is invisible. A number of different devices have been developed to detect the radiation. A charged electroscope can be used easily for this purpose. Solid-state detectors, dosimeters and thermoluminescent dosimeters (TLDs) are also used to detect and measure radiation. Other devices include the cloud chamber and the Geiger–Müller (G-M) tube.

Figure 3.17 ▼ In a cloud chamber, ionising particles cause tracks of condensed alcohol particles that can be observed.

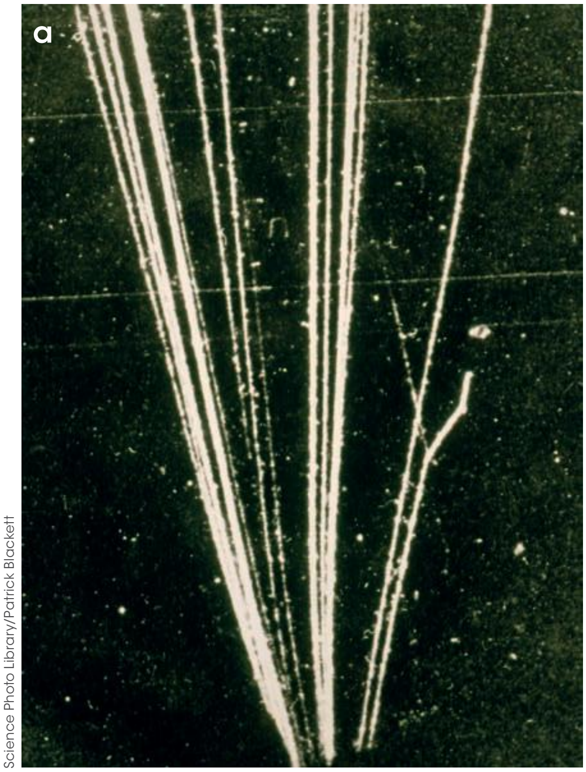


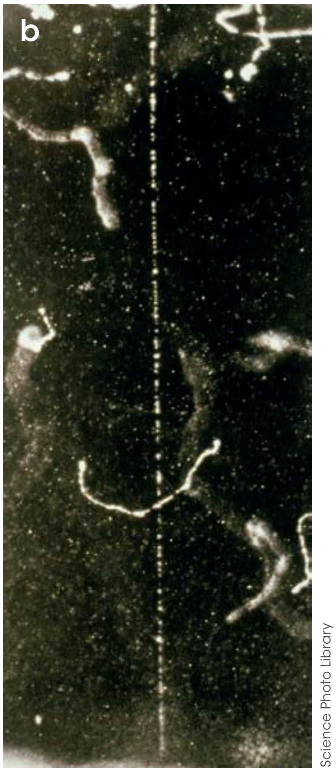
Cloud chamber

If a weak radioactive source is placed in a cloud chamber, ionising particles cause ionisation of the air in their path. The track becomes visible because the alcohol vapour condenses around these ions to produce many thousands of tiny alcohol droplets.

In a cloud chamber, α particle tracks are relatively thick compared with β particle tracks. α particles are the most strongly ionising particles, with the largest mass and momentum. The condensation lines are relatively straight except, occasionally, when an α particle collides directly with an air molecule.

In air, β particles are deflected by the electrons in molecules such as N_2, CO_2 and O_2. The paths are not straight, an effect that is particularly noticeable near the ends of the tracks when the β particles have transferred most energy and are slowing down.





◀ Figure 3.18
Cloud chamber tracks for different radiations: a) alpha particles from a radioactive source are shown travelling up the picture; image taken by Patrick Blackett at the Cavendish Laboratory, Cambridge, in 1921-4; b) the track of a fast beta ray traversing the cloud chamber from top to bottom. The short, thick tracks in the picture are not caused by the beta ray electron; they are due to other, less energetic electrons that have been knocked from atoms in the gas filling the chamber by invisible X-rays. The photograph was taken by CTR Wilson, the inventor of the cloud chamber.

In a cloud chamber, thin, short tracks are observed from an intense beam of γ rays. γ rays do not produce tracks directly. When a γ-ray photon is absorbed by a nuclide, an electron is ejected. It is these ionising particles that form the tracks.

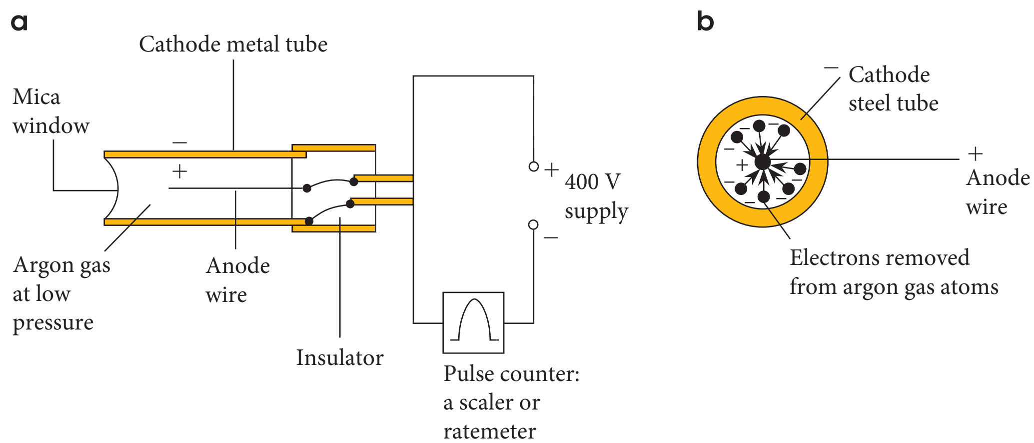
Liquid hydrogen bubble chambers replace alcohol with liquid hydrogen at low temperatures. This increases the intensity of the condensation tracks, which are produced by tiny hydrogen bubbles.

Geiger–Müller tube

The basic design features of a G-M tube are shown in Figure 3.19.

Radiation that enters through the mica window ionises the argon gas inside the tube by removing electrons. These electrons ionise the gas even more. The directly and indirectly produced electrons are collected on the positively charged central wire (the anode), which is at a 400 V potential difference with respect to the wall of the G-M tube. Lots of electrons, each of which causes an electric pulse, arrive at the anode. These are used to count the number of radioactive particles entering the G-M counter. Used as a **scaler**, the G-M tube counts all pulses. When used as a **ratemeter**, it records the number of counts per second:

count rate $= \dfrac{\text{number of counts}}{\text{time interval}}$. The uncertainty associated with radiation-counting experiments relates to the randomness of decay events. For N counts, the uncertainty is $\pm\sqrt{N}$.



◀ Figure 3.19
a) A Geiger–Müller tube and circuit; b) electrons are pulled towards the anode wire.

Scientific literacy: Smoke detectors

Fire services across Australia are agreed: smoke detectors save lives. But smoke detectors must be placed in appropriate positions and they must be properly maintained. That means making sure there is a good supply of air around the detector. The battery must be able to supply the correct voltage.

There are two sorts of smoke detector. One uses a metal detector that responds to light to produce electron flow in a circuit. The current in photoelectric detectors is changed when smoke particles reduce the effect of light on the detector. This change triggers the alarm siren.

The most effective smoke detectors contain a small alpha-decay source, usually americium-241. Americium-241 has a half-life of 432 years; however, the amount in any one detector is about 0.000 02 g. The rate of decay from 1 g of americium-241 is about 37 billion per second. This means that there are really very few alpha particles per second coming from the source in a smoke detector.

The americium-241 is placed in an aluminium can, called the ionisation chamber. The can has holes in it to allow air to flow through it. Air, which is mainly nitrogen (78%) and oxygen (21%), flows into the ionisation chamber. The alpha particles ionise the nitrogen and oxygen atoms. This produces electrons and positive ions within the chamber. The can is negatively charged, so positive ions move towards it. Above the can is a positively charged plate. The free electrons move towards this plate. The result is a current.

When smoke enters the ionisation chamber the current changes. Smoke particles are mostly charged, so the electrons and ionised air particles tend to discharge on the smoke particles. The result is a smaller current. This change is used to trigger the alarm siren.

Questions

1. Draw a diagram to show how the battery, ionisation chamber and plate are connected in an ionisation smoke detector.
2. Use correct nuclide symbols to write the decay equation for americium-241.
3. How many alpha particles per second are produced by the americium-241 source in a smoke detector?
4. Give two reasons why aluminium is used to make the ionisation chamber.
5. Smoke detectors with radioactive sources are used in houses and disposed of in the ordinary garbage. Explain why smoke detectors are not considered dangerous.

QUESTION SET 3.5

Remembering

1. For α, β and γ rays, which is:
   - a most penetrating?
   - b most ionising?
   - c least likely to cause damage?
2. In a magnetic field, which of α, β^-, β^+ and γ rays are:
   - a deflected in the same direction?
   - b not deflected at all?

Understanding

3. Which of α, β and γ rays are the least deflected in a cloud chamber? Why?
4. Describe the operation of a G-M tube.

Applying

5. Three identical radioactive sources, A, B and C, are placed in a vacuum in front of identical G-M tubes. Their count rates are measured simultaneously. Sources A, B and C are at effective distances of 25 cm, 50 cm and 1.50 m respectively from the G-M tube. Relative to B, is the count rate likely to be greater than, the same as or less than the count rate at B at:
   - a C?
   - b A?

9780170242103

6 What is the uncertainty associated with the following count numbers in radiation-counting experiments?
   a 100
   b 400
   c 10 000

Analysing

7 1600 counts are recorded in a G-M tube in 20.00 s.
   a What is the count rate?
   b What is the uncertainty in the count rate? By how much would your answer differ if the background radiation was 24 min^{-1}?

Reflecting

8 What have you learnt about the comparisons between different forms of radiation?

Half-life and decay series

Radioactive decay is a random event. For any given nucleus of a radioactive (unstable) isotope it is impossible to predict when it will decay, or even if it will decay at all in some time period. This is very much like throwing dice or tossing coins. The result of a single coin toss or throw of a die cannot be predicted.

However, if we have a large number of radioactive nuclei, we can say that some fraction of them will decay in a given time. The fraction that decays in any given time period depends upon the **half-life** of the particular nuclide. The shorter the half-life is, the greater the fraction of nuclei that will decay in a given time period. The half-life is the average time taken for half the nuclei in a sample to decay.

In a typical sample of radioactive material, say 1 gram of uranium, we may have 10^{20} unstable nuclei. In one half-life, half of these nuclei will decay. In two half-lives, $\frac{3}{4}$ of the original nuclei will have decayed, and so on.

This is very much like throwing a large number of dice. We cannot predict on which side any particular die will land, but for a large number of throws we can confidently say that $\frac{1}{6}$ will show 1, $\frac{1}{6}$ will show 2, $\frac{1}{6}$ will show 3 and so on. For coins, we cannot say which will show heads and which will show tails, but for a large sample we will get 50% heads and 50% tails.

In general, for a sample of N_0 particles, the number, N, remaining after n half-lives is given by the equation:

$$N = N_0\left(\frac{1}{2}\right)^n$$

The time required for the decay of half the original sample of atoms is called the half-life ($t_{1/2}$) of the radioactive material. Each radioactive isotope has a unique half-life.

> **Radioactive decay is a random event. Half-life is the time taken for half the nuclei to decay.**
>
> $N = N_0(\frac{1}{2})^n$
>
> **where n = whole number of half-lives.**

The letter N is used to represent the number of nuclei at any time when radioactive decay is being measured. As we have seen, N is also used to represent the number of neutrons in a nucleus. You need to use the context to distinguish which N is meant.

Let's look at an example. Polonium-218 has a half-life of 3 minutes and decays to Pb-214. If we start with 1.00×10^{20} Po-218 nuclei at some time, then after 3 minutes 0.50×10^{20} Po-218 nuclei will have decayed to Pb-214, and we will have 0.50×10^{20} remaining Po-218 nuclei. After 6 minutes, two half-lives have passed, and $\frac{3}{4}$ of the original Po-218 nuclei will have decayed, leaving only 0.25×10^{20} Po-218 nuclei. Table 3.6 shows the number of nuclei remaining as a function of time.

Table 3.6 Decay of a particular sample of radioactive nuclides

Time (minutes)	0	3	6	9	12	15
Number of half-lives elapsed	0	1	2	3	4	5
Number of Po-218 nuclei remaining ($\times 10^{20}$)	1.00	0.50	0.25	0.125	0.0625	0.03125
Number of Po-218 nuclei decayed into Pb-214 ($\times 10^{20}$)	0	0.50	0.75	0.875	0.9375	0.96875

Our example shows us that after 5 half-lives, the fraction of radioactive nuclei has dropped to less than 5% of the original number, and hence more than 95% of the nuclei have decayed.

EXPERIMENT 3.2

RANDOM DECAY AND HALF-LIFE: A SIMULATION

Radioactive decay occurs in a random way. A nuclide either decays or it does not decay – two states.

Aim

To simulate random decay and half-life of a radioactive nuclide

Materials

Per group:

- 1 bag or cup
- 80 small counters with an obvious way to determine the Up side and the Down side.

In your write-up, add any risks you can think of, as well as ways to manage them.

Procedure

1. As a class, decide which side of the counters represents decay (Up) and which side represents not-decay (Down).
2. Shake the bag and pour counters onto the table.
3. Remove all counters that represent decay.
4. Count and record the counters that represent the remaining nuclides that are yet to decay.
5. Replace these in the bag.
6. Repeat the process until there is one or no counters remaining.

Results

Combine the results from all groups into a single data table for the class.

Analysis of results

1. Use whole class data to plot a graph of the number of counters remaining versus the number of trials.
2. From the graph, determine the half-life of the counters.

Discussion

1. For radioactivity, how did this experiment model:
   a. the randomness of decay?
   b. half-life?
2. A radioactive nuclide is an unstable nucleus that could decay at any moment. The decay occurs because the daughter nuclide is more stable than the parent. Discuss.

Extension

In this experiment the half-life was one 'throw' of the counters. Try modelling a half-life that is less than one 'throw' by repeating this experiment using a large number of dice. What is the half-life when you remove all dice with a one showing at each throw? What if you remove all those with a one or a two showing?

Decay series

Many products of radioactive decay are themselves radioactive. Eventually, a stable end product is reached. Three naturally occurring **decay series** have been identified.

1 Radium or uranium series from $^{238}_{92}U$ to $^{206}_{82}Pb$

2 Actinium series from $^{235}_{92}U$ to $^{207}_{82}Pb$

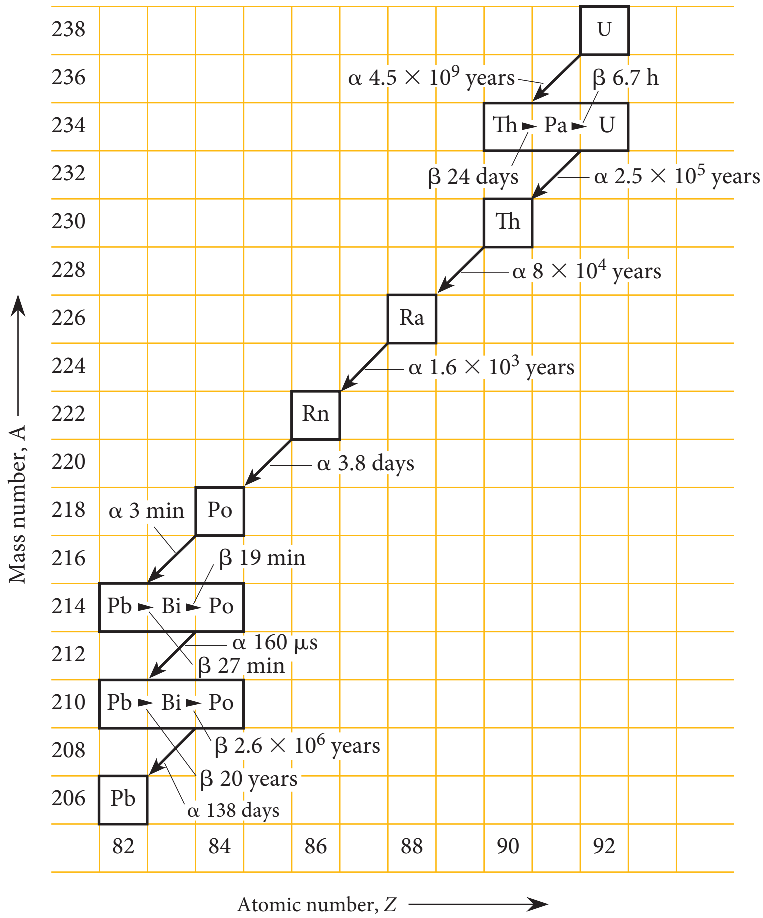
3 Thorium series from $^{232}_{90}Th$ to $^{208}_{82}Pb$

The end product of each of these series is a stable lead nuclide, $Z = 82$.

A fourth series, the neptunium series, starts at neptunium-237 and finishes at the stable nuclide thallium-205. Neptunium can only be produced artificially. Only two of the decay chain daughters occur naturally.

Radium series

The radium series can be represented graphically.



◀ **Figure 3.20** The radium decay series starts with ^{238}U and ends with stable ^{206}Pb. The half-life of each radioactive nuclide is also given. Note that γ rays are associated with some of these decays.

The nuclides in the radium series are:

parent: $^{238}_{92}U \rightarrow$

daughters: $^{234}_{90}Th \rightarrow ^{234}_{91}Pa \rightarrow ^{234}_{92}U \rightarrow ^{230}_{90}Th \rightarrow ^{226}_{88}Ra \rightarrow ^{222}_{86}Rn \rightarrow ^{218}_{84}Po$

$\rightarrow ^{214}_{82}Pb \rightarrow ^{214}_{83}Bi \rightarrow ^{214}_{84}Po \rightarrow ^{210}_{82}Pb \rightarrow ^{210}_{83}Bi \rightarrow ^{210}_{84}Po \rightarrow ^{206}_{82}Pb$

WORKED EXAMPLE 3.4

The neptunium series begins with an alpha decay, followed by a beta-minus decay.

a Show the first two decays in the series in correct symbol form. (2 marks)

b Write in words the names of the two daughter nuclides. (1 mark)

Answers	Logic	
a $^{237}_{93}Np \rightarrow {}^{233}_{91}Pa + {}^{4}_{2}He$	Use correct symbols for daughter and alpha particles.	1 mark
$^{233}_{91}Pa \rightarrow {}^{233}_{92}U + {}^{0}_{-1}e + \overline{\nu}$	Use correct symbols for daughter and beta-minus particles.	1 mark
b Daughter nuclides are protactinium-233 and uranium-233.	Use correct name for each daughter nuclide.	1 mark

Try these yourself

1 The thorium series begins with an alpha decay, followed by a beta-minus decay.

a Show the first two decays in the series in correct symbol form. (2 marks)

b Write in words the names of the two daughter nuclides. (1 mark)

2 The actinium series begins with an alpha decay, followed by a beta-minus decay.

a Show the first two decays in the series in correct symbol form. (2 marks)

b Write in words the names of the two daughter nuclides. (1 mark)

Table 3.7 Decay series: cascade of decays from a radioactive nuclide until a stable nuclide is reached

Name	Start nuclide	End nuclide
Radium	$^{238}_{92}U$	$^{206}_{82}Pb$
Actinium	$^{235}_{92}U$	$^{207}_{82}Pb$
Thorium	$^{232}_{90}Th$	$^{208}_{82}Pb$
Neptunium	$^{237}_{93}Np$	$^{205}_{81}Tl$

Artificial transmutation

Rutherford was the first to use radioactivity to produce new nuclides. He was able to make one element into another. He bombarded nitrogen-14 with alpha particles and analysed the result. Oxygen and hydrogen were formed. The reaction proceeded as follows.

Nitrogen nuclei absorb helium nuclei and form a composite, unstable nuclide, denoted by an asterisk:

$$^{14}_{7}N + {}^{4}_{2}He \rightarrow {}^{18}_{9}F^*$$

The composite nuclide decays to a more stable state:

$$^{18}_{9}F \rightarrow {}^{17}_{8}O + {}^{1}_{1}H$$

The discovery of the neutron enabled scientists to explore the behaviour of larger atomic nuclei. As it is neutral, a neutron is not repelled by the target nucleus. It can be absorbed into the nucleus of the target atom. This makes it very useful as a form of bombarding radiation. It is used in many experiments to transmute a number of nuclides artificially.

When a nucleus takes in a neutron it becomes less stable. Frequently, the nuclide becomes a beta-emitter.

Bombarding uranium nuclei with neutrons delivered unexpected results. Capture of a neutron by a uranium nuclide can lead to two results: formation of a transuranic element or a split into two nuclei of intermediate mass.

WOW

Neutron transmutation doping of silicon crystals

Neutrons are used at the OPAL reactor at ANSTO in Sydney to produce very high quality semiconductors. Pure silicon crystals are placed in a beam of neutrons from the reactor. The width, intensity and energy of the beam can be precisely controlled. Silicon nuclei that absorb a neutron are converted into phosphorus nuclei. The pure silicon has been 'doped' with phosphorus, improving its conductivity. The density of these phosphorus nuclei can be precisely controlled, and varied over the volume of the silicon crystal as desired. Hence a semiconductor with the desired properties can be very precisely manufactured. ANSTO performs neutron transmutation doping on semiconductors that are then used in many devices including digital cameras and video cameras. You may have even used one yourself.

Transuranic elements

Each element beyond uranium (atomic number >92) is a **transuranic element**. They do not exist naturally. All are produced artificially. All are radioactive. There are no known stable isotopes of any transuranic element. Some, like plutonium and neptunium, have very long half-lives (more than 4 million years). Others, such as americium, berkelium and californium are reasonably long-lived (800–34 000 years). Elements 109–118 have half-lives from minutes to milliseconds or less.

In 1934 Enrico Fermi (1901–54), an Italian-born scientist working in the United States, found several nuclides with different half-lives. These were new elements with atomic numbers greater than 92. For example, consider the process of forming plutonium-239 by neutron bombardment of uranium-238. The process goes through neptunium-239 to plutonium-239:

$$^{1}_{0}\text{n} + ^{238}_{92}\text{U} \rightarrow ^{239}_{92}\text{U} + \gamma$$

$$^{239}_{92}\text{U} \rightarrow ^{239}_{93}\text{Np}^{*} + ^{0}_{-1}\text{e} + \overline{\nu}$$

$$^{239}_{93}\text{Np}^{*} \rightarrow ^{239}_{94}\text{Pu} + ^{0}_{-1}\text{e} + \overline{\nu}$$

QUESTION SET 3.6

Remembering

1 Define 'half-life'. Write the equation that links number of nuclides and whole numbers of half-lives.

2 There are four decay series.

   a Name them.

   b What is the starting nuclide of each series?

   c What is the ending nuclide of each series?

   d Which of these arises from artificial transmutation?

3 What is a transuranic element? How were the first two transuranic elements produced?

Understanding

4 The half-life of polonium-218 is 3.0 minutes. A particular nuclide of polonium-218 has not decayed after 9.0 minutes. What are the chances that it will decay sometime before 12 minutes?

5 Consider the radium decay series shown in Figure 3.20.

  a Write the decay reaction for radium-226 to radon-222.

  b Write the decay reactions for lead-210 to bismuth-210 to polonium-210.

Applying

6 The half-life for thallium-200 is 1.0×10^4 s. A kilogram of thallium-200 contains close to 3.0×10^{24} nuclides. After 1.0×10^4 s, about how many nuclides in the original sample:

  a have decayed?

  b are still to decay?

7 In neutron bombardment, the production of a composite nuclide is followed by beta decay. Write the nuclear reaction for the formation of the composite nuclide and the decay equation for the neutron bombardment of:

  a boron-11.

  b mercury-191.

Analysing

8 Plot a graph to show the decay of 1.6×10^{26} particles of protactinium-225, which has a half-life of 2.0 s. Show at least 4 half-lives on your graph. From the graph, determine the number of nuclides remaining after 3.5 s.

Nuclear medicine

Nuclear medicine uses radiopharmaceuticals for medical diagnosis and treatment. In diagnosis, an external detector records the passage or localisation of a radioactive nuclide. Nuclear medicine treatment destroys cells and/or promotes healing. Treatment radiopharmaceuticals are inserted into body cavities, swallowed or injected. They must be localised to the target organ or tissue to be treated. Diagnostic radiopharmaceuticals must be readily available, cheap, and able to produce high-quality images for low doses. But, they should not cause unacceptable radiation or chemical risks. The best diagnostic radionuclides produce single energy γ rays of between 30 and 300 keV – preferably 150 keV.

Table 3.8 Some γ-emitting radionuclides used in diagnostic nuclear medicine

Isotope	Half-life	Decay mode	γ energy (keV)	Diagnostic uses
$^{99m}_{43}Tc$	6 hours	γ	140	Scans: bone, bone marrow, heart, liver, brain, testicular, kidney Flow studies: cerebrospinal fluid (CSF), heart, lymph glands
$^{123}_{53}I$	13 hours	γ	159	Adrenal gland tumour
$^{133}_{53}I$	8 days	β^-, γ	364	Thyroid cancer
$^{201}_{81}Tl$	73 hours	γ	6980, 167	Heart disease, angina

WHAT ANSTO DOES

Find out what isotopes are made at ANSTO, and what they are used for.

As you can see in Table 3.8, many of these isotopes have short half-lives. It is not practical to import these short-lived isotopes from overseas into Australia. It is necessary to have a local source of production if we wish to be able to use these isotopes for medical therapy and diagnosis. The ^{123}I used in Australia is produced at the OPAL reactor in Sydney and sent to hospitals all around the countries for use. ANSTO also produces ^{99}Mo sources. These sources are contained in sealed generators that are supplied to hospitals and diagnostic centres in Australia and other countries. The ^{99}Mo decays to ^{99m}Tc, which is extensively used in medical imaging but has too short a half-life to be shipped in from off-site for most facilities. Most of the isotopes listed in Tables 3.8 and 3.9 are produced at ANSTO.

Isotope exchange increases the concentration of radioactive nuclides by replacing some of the non-radioactive nuclides normally present. For example, radioactive iodine, ^{125}I, replaces natural iodine, ^{127}I, in the thyroid and emits a 150 keV γ-ray for detection. Foreign label nuclides such as ^{99m}Tc and ^{18}F are attached to chemicals in cerebrospinal fluid (CSF), whose pathway through the body is well-known. Biosynthesis radiopharmaceuticals such as ^{60}Co or ^{14}C can be introduced into the body, metabolised and then removed by excretion or some other pathway (natural or more invasive). The choice of radiopharmaceutical always takes account of the effective half-life, which is a combination of the biological half-life (time for half the dose to be removed from the body) and the physical half-life.

The choice of treatment radionuclides is governed by the problem and its location. For example, skin melanoma, bladder cancer and bony secondary cancers require different treatments. The type and energy of emissions from a radionuclide and their consequent range in tissue needs to be considered. The metabolic pathway followed by the radionuclide also needs to be considered. The chemical should become concentrated at the particular location where it can be most effective. It should also stay in the location for a sufficient time to achieve what is intended. The radiochemical should leave the body in a reasonable time (appropriate biological half-life).

Other factors that affect the choice of radiopharmaceutical include cost and whether the impurities introduced as a result of the preparation of the sample are likely to be dangerous or reduce the effectiveness of the treatment.

Table 3.9 Properties of some β^--emitting radionuclides used in treatment

Radionuclide	Half-life (days)	Energy of β^- particle (MeV)	Range in tissue (mm)	Application
$^{32}_{15}P$	14.3	1.71	8.7	Bone pain relief
$^{89}_{38}Sr$	50.5	1.49	8.0	Bone pain relief
$^{131}_{53}I$	8.0	0.61	2.4	Thyroid problems
$^{188}_{75}Re$	0.71	2.11	10.8	Artery clogging
$^{198}_{79}Au$	2.7	0.96	4.4	Liver scan

Radiation exposure in a nuclear medicine centre can be quite high. Radiochemicals are often contained in vials, which are kept in lead 'castles' or 'pigs'. Workers watch through lead glass windows and in a mirror as they draw up radiochemicals into syringes.

They wear disposable or washable clothing, gloves and face shields. Thermoluminescent dosimeters (TLDs) register the dose received, usually near worker's gonads and fingers.

Figure 3.21 ▶
A syringe shield is made of lead or tungsten, with a window of lead glass so that the markings on the syringe can be seen. Note the worker's disposable gloves.

QUESTION SET 3.7

Understanding

1 Distinguish between diagnosis and treatment.

Remembering

2 What two factors are taken into account when determining the effective half-life of a radiopharmaceutical?

Applying

3 Identify and describe three biological mechanisms by which diagnostic radionuclides are used. Why must the effective half-life be considered when deciding on a diagnostic radiopharmaceutical?

4 For diagnostic radiopharmaceuticals, explain why:
   a it is important to have a local source or radiopharmaceuticals.
   b it is important to have quite high rates of gamma ray emission.
   c technetium-99m is a nearly perfect diagnostic radionuclide.

Analysing

5 Would you be more likely to use phosphorus-32 or strontium-89 for pain relief in a palliative care patient with bone cancers, who has only a few days to live? Explain your reasoning.

6 a What is the purpose of keeping vials of radiochemicals in 'pigs'?
   b Why is a mirror used while drawing up a syringe of radioactive chemical?
   c Give reasons why workers wear disposable outer garments when handling radiochemicals.

CHAPTER SUMMARY

- Radiation is of two forms: non-ionising and ionising.
- The Rutherford–Bohr nuclear model of the atom consists of a nucleus containing protons and neutrons, with electrons in orbit about the nucleus.
- There are two ways to represent the nuclear model of the atom:
  - the planetary system (similar to the solar system)
  - the energy level diagram (similar to a ladder with uneven steps)
- A proton is positively charged, an electron is negatively charged and a neutron is neutrally charged.
- Protons and neutrons are collectively called nucleons.
- An atom is electrically neutral.
- The positive charge on the proton has the same magnitude as the negative charge on the electron.
- The atomic number (Z) of an atom is the number of protons in its nucleus.
- The mass number (A) of an atom is the number of nucleons in its nucleus.
- The number of neutrons in the nucleus (N) is $A - Z$.
- Isotopes are nuclides of the same element that have the same atomic number but different mass number. The most common, naturally occurring isotope is called 'the element'.
- Radioactivity is the uncontrollable, random decay of a nuclide.
- Alpha particles, beta particles and gamma rays are emitted from a nucleus during radioactive decay. (These radiations are invisible and can only be detected by their ionising properties.)
- An alpha particle is the nucleus of a helium atom (${}^{4}_{2}\text{He}$).
- A beta particle can be either an electron (${}^{0}_{-1}\text{e}$) or a positron (${}^{0}_{+1}\text{e}$).
- Alpha and beta particles are charged particles and so are deflected by electric and magnetic fields.
- A gamma ray exhibits electromagnetic wave properties and is not deflected by electric or magnetic fields.
- The daughter nuclide of a radioactive parent nuclide that undergoes alpha or beta decay is different from the parent.

Decay	Nucleon number change	Proton number change
Alpha-decay	4 less	2 less
Beta minus-decay	0	1 more
Positron-decay	0	1 less

- Penetrating power: Alpha particles are highly ionising but weakly penetrating. Beta particles, with half the charge and lower mass, are less ionising but more penetrating. Gamma rays penetrate most materials highly.
- Many naturally occurring radioactive materials decay through a series of nuclides until a stable isotope of lead is reached.
- The activity of a radioactive sample is the number of atoms in a sample that decay in unit time.
- The time required for one-half of an original sample of radioactive material to decay, or for its activity to decrease by one-half, is called the half-life of the material: $N = N_0(\frac{1}{2})^n$

CHAPTER GLOSSARY

alpha particle/alpha ray first known radioactive particle; it is a helium-4 nuclide with two positive charges

antineutrino a weakly interacting particle involved in energy transformations, especially beta-minus emission; thought to have little or no mass and zero charge

atom particle; originally an indivisible particle; now known to comprise several smaller particles

atomic mass mass of a nuclide compared to a single carbon-12 nuclide

atomic mass number (*A*) total number of nucleons in a nuclide

atomic number (*Z*) number of protons in a nucleus

atomic weight weighted average of all naturally occurring nuclides of an element in a sample

background radiation radiation that is naturally present at a location

beta particle/beta ray electron (beta-minus) or positron (beta-plus)

classical electrodynamics unified understanding of the mutual interactions of electricity and magnetism

cosmic radiation radiation whose origins are in space

count rate number of counts per second

daughter nuclide nuclide resulting from radioactive decay

decay series cascade of decays from a radioactive nuclide until a stable nuclide is reached

deflection difference between a straight path and the actual path

discrete quanta particular value; values between this and another value are not permitted

disintegration nuclear process leading to a new nuclide and emission of alpha, beta and/or gamma rays

electromagnetic spectrum spectrum of electromagnetic waves from gamma rays, through visible light to radio waves

element a substance that only has nuclides with the same number of protons

gamma rays high-energy electromagnetic radiation

half-life time taken for half a sample to decay

ionising power ability to ionise materials; inversely proportional to penetrating power

ionising radiation energy in particle or electromagnetic form that can affect the number of electrons surrounding a nucleus

isobar nuclides with the same mass (nucleon) number

isomer nuclides with the same atomic mass and nucleon number, but different energy states

isotone nuclides with the same number of neutrons

isotope any nuclide of the same atomic number (from Greek meaning 'equal type'); nuclide of an element that differs only with respect to the number of neutrons

metastable nuclide a nuclide that persists in an energy state above ground state for more than 10^{-12} s

neutrino a weakly interacting particle involved in energy transformations, especially positron emission; thought to have little or no mass and zero charge

neutron neutral nuclear particle with mass slightly greater than that of a proton

non-ionising radiation energy in electromagnetic form that does not affect atomic electrons

nuclear transformation nuclear process leading to a new nuclide and emission of alpha, beta and/or gamma rays

nucleons nuclear particles; protons and neutrons

nuclide species of atom

parent nuclide original nuclide before radioactive emission

penetrating power ability of ionising radiation to move into or through materials

proton positively charged subatomic particle

quantum physics physics based on discrete states of matter and energy at the smallest levels

radiation energy transfer across space; the process by which heat is transferred without the need for a medium; energy from radioactive atoms

radioactive decay particles or rays that come from energy re-arrangements in a nucleus

radioactivity emission of energy in particle or electromagnetic form from the nucleus of an atom

ratemeter counter that records counts per second

relative atomic mass atomic weight

scaler counter that records total counts

scintillation spark of light

terrestrial radiation radiation from materials in Earth's crust and core

transmutation nuclear process leading to a new nuclide and emission of alpha, beta and/or gamma rays

transuranic elements element beyond uranium; artificially produced nuclide with more than 92 protons

CHAPTER REVIEW QUESTIONS

Remembering

1 a Define 'non-ionising' and 'ionising' radiation.
  b What form of radiation is electromagnetic radiation?
  c Provide the name, symbol and charge for each type of radioactive emission.

2 Explain the difference between stable, meta-stable and unstable nuclei. What sort of particles can be emitted in the decay of:
  a meta-stable nuclei?
  b unstable nuclei?

3 Define these terms.
  a Atomic number
  b Atomic mass number
  c Unified mass unit
  d Atomic weight

4 Write down the general equation for:
  a alpha decay.
  b beta-minus decay.
  c positron decay.
  d gamma radiation.

5 How did the experiment by Geiger and Marsden help to refine the nuclear model of the atom?

Understanding

6 For radioactive emissions, make a table to show:
  a penetrating power.
  b ionising ability.
  c deflection in a magnetic region.
  d deflection in an electric region.

7 How do chemical and nuclear reactions differ?

8 Consider the equation $N = N_0\left(\frac{1}{2}\right)^n$.
  a Define each term.
  b Explain how the equation represents spontaneous radioactive decay.

9 How did the work of Henry Moseley improve the periodic table?

10 Identify two radiopharmaceuticals that are used in medical diagnosis.

Applying

11 Naturally occurring thallium comes in two forms, $^{203}_{81}\text{Tl}$ (29.50%) and $^{205}_{81}\text{Tl}$ (70.50%).
  a Find the number of neutrons in $^{205}_{81}\text{Tl}$.
  b Find the atomic weight of thallium.

12 Platinum-190 decays by emitting an α particle and changes to a completely different element.
  a Write a complete nuclear reaction equation that includes the symbol for the daughter nuclide.
  b What is the name of the daughter nuclide?

13 Iodine-119 decays to an isotope of tellurium, which later decays to an isotope of antinomy, which then gamma decays. Write the three separate decay equations.

14 Complete the following table by writing the decay equation using correct nuclide symbols.

Parent nuclide	Daughter nuclide	Decay equation
Nobelium-254	Fermium-250	
Iridium-182	Osmium-182	
Barium-137m	Barium-137	
Oxygen-15	Nitrogen-15	

15 When an isotope of molybdenum ($Z = 42$) is bombarded by neutrons, it produces a composite nuclide, which then decays to the metastable state of technetium-99, $^{99m}_{43}Tc$.

a Write the equation for the decay of the isotope of molybdenum to technetium-99m.

b Hence, write the nuclear reaction equation that leads from molybdenum to technetium.

16 Technetium-99m decays isomerically. It has a half-life of 6 h.

a Give three reasons why technetium-99m is such a useful radiopharmaceutical.

b Write the decay equation.

c Determine the time it takes for a 2.0 mg sample of technetium-99m to reduce to 62.5 µg of technetium-99m.

17 A sample of an alpha-emitting radiochemical comprises 1.6×10^{20} nuclides. Over a period of 7.0 minutes, the number of the original nuclides is reduced to 5.0×10^{18} nuclides. How long is the half-life?

18 A patient in a hospital has been injected with a radiochemical. What radiation safety precautions must be taken during the process and afterwards?

Analysing

19 Irid and Bism undertook identical radiation counting experiments using a G-M tube and timer. Irid decided to measure the time for 100 counts. This took 4.1 s. Bism decided on the time taken for 10 000 counts. This took 398 s.

a For each person, what was the:

i count rate?

ii uncertainty in the count rate?

b Give a reason for the sense or otherwise of their respective decisions about the measurement of count rate.

20 Plot a graph to show the decay of 6.4×10^{31} particles of polonium-193, which has a half-life of 4.0 s. From the graph, determine the number of nuclides remaining after 10.0 s.

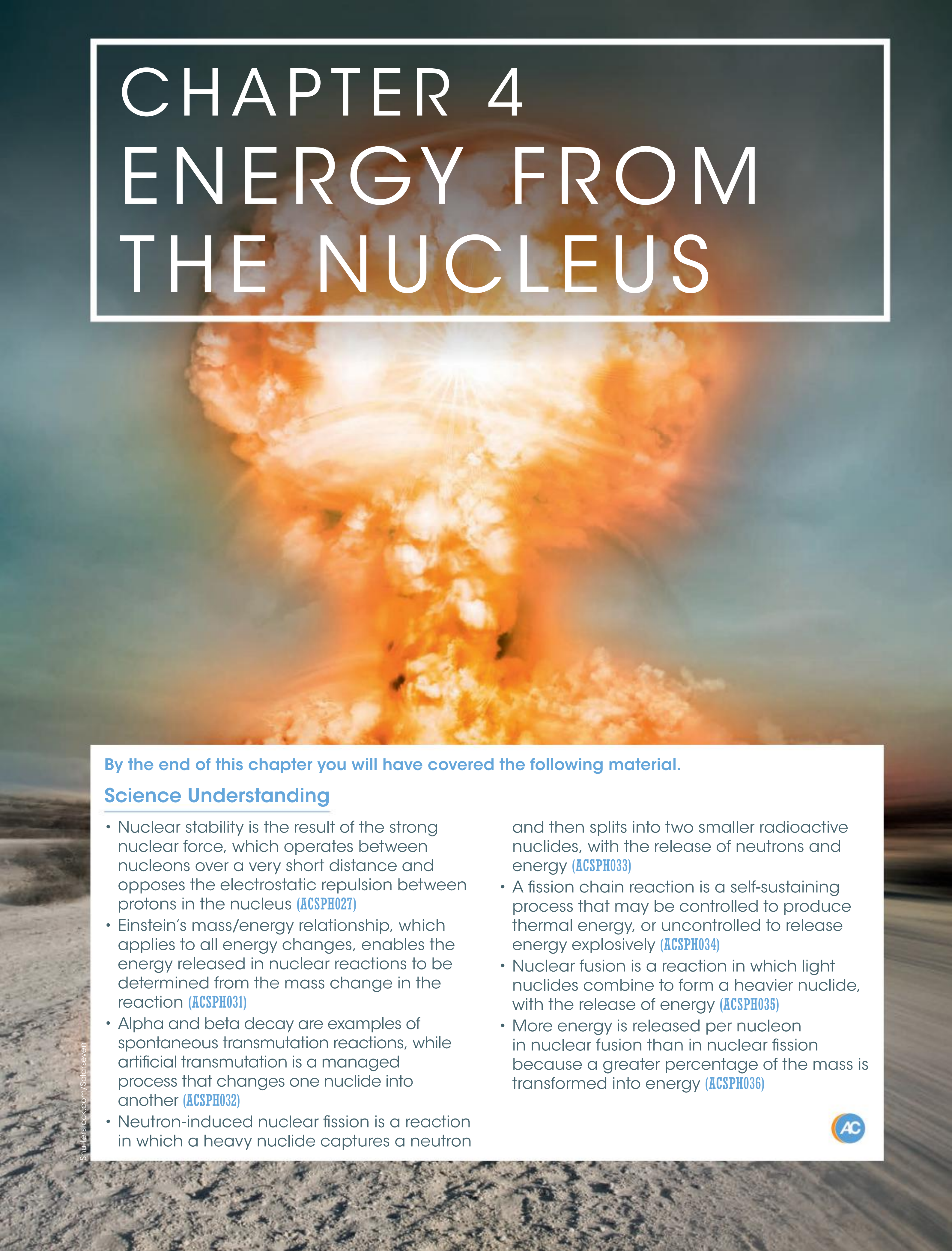
21 How did international cooperation lead to the development of, and improvements in, the understanding of the particle nature of matter, especially the nuclear atom?

Reflecting

22 What have you learnt about models and their development?

23 How has your knowledge of, and opinion about, nuclear processes changed?

24 What personal qualities are most evident in the scientists discussed in this chapter?



Shutterstock.com/Solarseven

CHAPTER 4 ENERGY FROM THE NUCLEUS

By the end of this chapter you will have covered the following material.

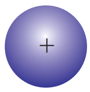
Science Understanding

- Nuclear stability is the result of the strong nuclear force, which operates between nucleons over a very short distance and opposes the electrostatic repulsion between protons in the nucleus (ACSPH027)
- Einstein's mass/energy relationship, which applies to all energy changes, enables the energy released in nuclear reactions to be determined from the mass change in the reaction (ACSPH031)
- Alpha and beta decay are examples of spontaneous transmutation reactions, while artificial transmutation is a managed process that changes one nuclide into another (ACSPH032)
- Neutron-induced nuclear fission is a reaction in which a heavy nuclide captures a neutron and then splits into two smaller radioactive nuclides, with the release of neutrons and energy (ACSPH033)
- A fission chain reaction is a self-sustaining process that may be controlled to produce thermal energy, or uncontrolled to release energy explosively (ACSPH034)
- Nuclear fusion is a reaction in which light nuclides combine to form a heavier nuclide, with the release of energy (ACSPH035)
- More energy is released per nucleon in nuclear fusion than in nuclear fission because a greater percentage of the mass is transformed into energy (ACSPH036)

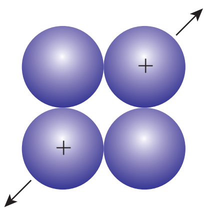
Introduction

Enormous amounts of energy can be produced from atomic nuclei. Radioactive decay products are more energetic than emissions from the atomic-level electron transitions. We have seen, for example, how they are used to enhance the scope of medical diagnosis and treatment. Nucleons are very strongly connected together in the nucleus. Re-arrangements of these nucleons, by 'splitting the atom' (fission) or by adding nucleons together (fusion) can release energy. Each fission or fusion event releases tiny amounts of energy. These amounts per event are, however, far greater than the energy per atom released in a chemical reaction such as the burning of fossil fuels. Fission and fusion release enormous amounts of energy because the enormous number of nucleons involved multiplies the energy released per event.

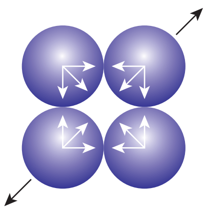
The use of nuclear energy is controversial, because human health can be compromised when things go wrong.



The nucleus of the hydrogen-1 atom consists of a single proton.



In a helium-4 nucleus electrostatic forces push the protons apart.



There must be forces between nucleons pulling them together. Gravitational forces are far too small.

Figure 4.1 ▲ The strong nuclear force acts to hold nucleons together against the force of electrostatic repulsion. Gravitational force is negligible.

What holds a nuclide together?

Protons have a positive charge. Positive charges repel each other. This repulsion is caused by the electrostatic force. The electrostatic force becomes relatively large when the protons are close together. In a nucleus, protons come within about 2×10^{-15} m (2 femtometres, 2 fm). The electrostatic force of repulsion by one proton on another proton is about 60 N when they are at their closest in a nucleus.

Protons have mass. Masses attract each other. This is called gravitational force. Perhaps the gravitational force of attraction keeps the protons together? But no! Gravitational force by one proton on another proton is miniscule even when they are 2 fm apart. It is too small by a factor of 10^{36}!

Enter the **strong nuclear force**. The strong nuclear force is required to keep nucleons together. In a helium-4 nucleus, for example, the protons and neutrons use the strong nuclear force to pull inwards on each other in order to overcome the electrostatic force of repulsion.

Four fundamental forces

Physicists have identified four fundamental forces that appear to hold the key to understanding the universe. We have indicated that, at the nuclear level, the strong nuclear force has the greatest effect on keeping nucleons in the nucleus, overcoming the smaller electromagnetic force. The gravitational force is a distant last. There is a fourth force, the weak force, that acts within nucleons. If the strength of the gravitational force is taken as the basis upon which to compare the effect of these forces at the nuclear level, the others are far greater.

Table 4.1 Comparison of the effect of the four fundamental forces within a nucleus

	Gravitational	Weak	Electromagnetic	Strong
Relative magnitude	1	10^{32}	10^{36}	10^{40}
Range (m)	Infinite	10^{-18} or 1 attometre, 1 am	Infinite	10^{-15} or 1 fm

FOUR FUNDAMENTAL FORCES

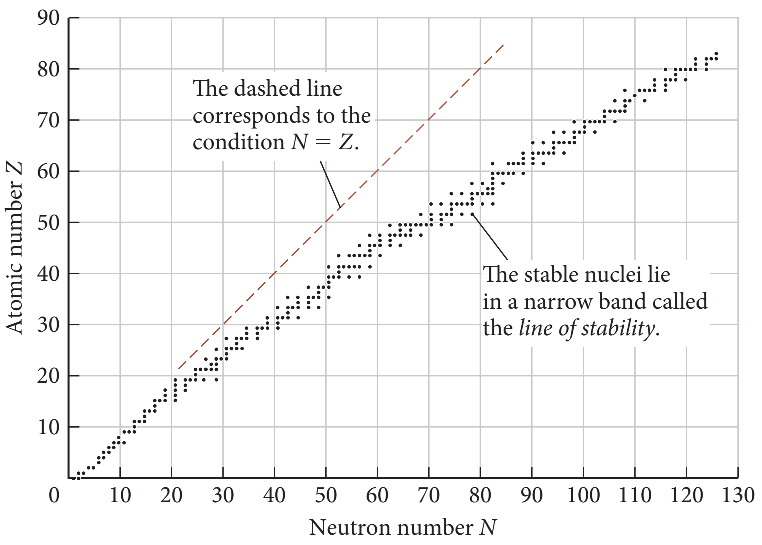
Find out more about the four fundamental forces and the way they act.

Stability of nuclides

All nuclides, except hydrogen-1, comprise protons and neutrons. The neutrons assist to reduce the effect of the electrostatic force of repulsion between protons. Within the nucleus the strong nuclear force overcomes the electrostatic force. When the effect of the strong nuclear force is sufficiently strong, the nuclide is stable. Otherwise the nuclide is unstable and will emit radiation: α, β and/or γ rays.

Nuclides comprising up to about 40 nucleons are stable when they have equal, or nearly equal, numbers of protons and neutrons. Heavier nuclides are more stable when the balance is in favour of neutrons.

In the **stability curve** shown in Figure 4.2, the unstable nuclide shown is most likely to decay by β^- decay. This decay increases the proton number, Z, by +1, while reducing the neutron number, N, by one. The daughter nuclide will then be on or closer to the stability line.



▲ Figure 4.2
The stability curve for nuclides

Energy stored in, and released from, nuclei

Nuclei are made from protons and neutrons. The energy that would be needed to disassemble a nucleus into its component nucleons is the **binding energy**. Each nucleon, on its own, has a mass. But when nucleons are brought together to form a nucleus, the mass of the nucleus is less than the sum of all the individual nucleons. The difference, Δm, between the sum of the individual masses and the mass of the nucleus into which they are combined is called the **mass defect**. The mass defect is a measure of the energy, ΔE, needed to bring all the parts of a nucleus together. **Einstein's mass–energy equation** is a quantitative statement of this effect:

$$\Delta E = (\Delta m)c^2$$

where c = the speed of light, $3.00 \times 10^8\ \mathrm{m\,s^{-1}}$.

Some of the mass of the individual nucleons appears as the binding energy of the nucleus. In this sense, it is best to consider mass and energy as interchangeable:

$$\text{mass} \Leftrightarrow \text{energy}$$

> Mass and energy are equivalent. Mass is a manifestation of energy.
>
> $$\Delta E = (\Delta m)c^2$$

THE PENINSULA OF NUCLEAR STABILITY

Download, open the file and follow the instructions to gain a better understanding of the role of neutrons in the stability of the first ten elements.

A new unit of energy: the electron-volt, eV

The energy equivalent of nucleon mass is very small. It is so small that a different energy unit is used, so that the numbers are simpler to play with. The new energy unit is called **electron-volt, eV**.

One electron-volt is an energy unit. It is equivalent to the energy that an electron gains when it moves through a potential difference of one volt. Thus, 1 eV is equivalent to 1.602×10^{-19} J. Nuclear energies are often in the range of thousands of eV (keV), millions of eV (MeV) or billions of eV (GeV).

Table 4.2 shows that individual nucleons have slightly greater masses than the mass of 1 unified mass unit. The energy equivalent of 1 u is 931.5 MeV.

STABILITY CURVE

This link provides nuclide information and allows you to predict decay of nuclides.

Table 4.2 Mass of nucleons in kilogram and unified mass units

	Mass (kg × 10^{-27})	Mass (u)
Proton	1.673	1.0073
Neutron	1.675	1.0086

WORKED EXAMPLE 4.1

Find the energy equivalent of a free proton in:

a joule (J) (3 marks)

b electron-volt (eV). (2 marks)

Answers		Logic	
a	$\Delta E = (\Delta m)c^2$	Use the correct formula.	1 mark
	$\Rightarrow \Delta E = (1.673 \times 10^{-27}\ \text{kg}) \times (3.000 \times 10^8\ \text{m s}^{-1})^2$	Substitute the correct values.	1 mark
	$\Rightarrow \Delta E = 1.506 \times 10^{-10}\ \text{J}$	Calculate the answer.	1 mark
b	$\Delta E = \dfrac{1.506 \times 10^{-10}\ \text{J}}{1.602 \times 10^{-19}\ \text{J eV}^{-1}}$	Substitute the correct values.	1 mark
	$\Delta E = 9.40 \times 10^8\ \text{eV}$		
	$\Delta E = 940\ \text{MeV}$	Calculate the answer.	1 mark

Try these yourself

Find the energy equivalent of a free neutron in:

a joule (J). (3 marks)

b electron-volt (eV). (2 marks)

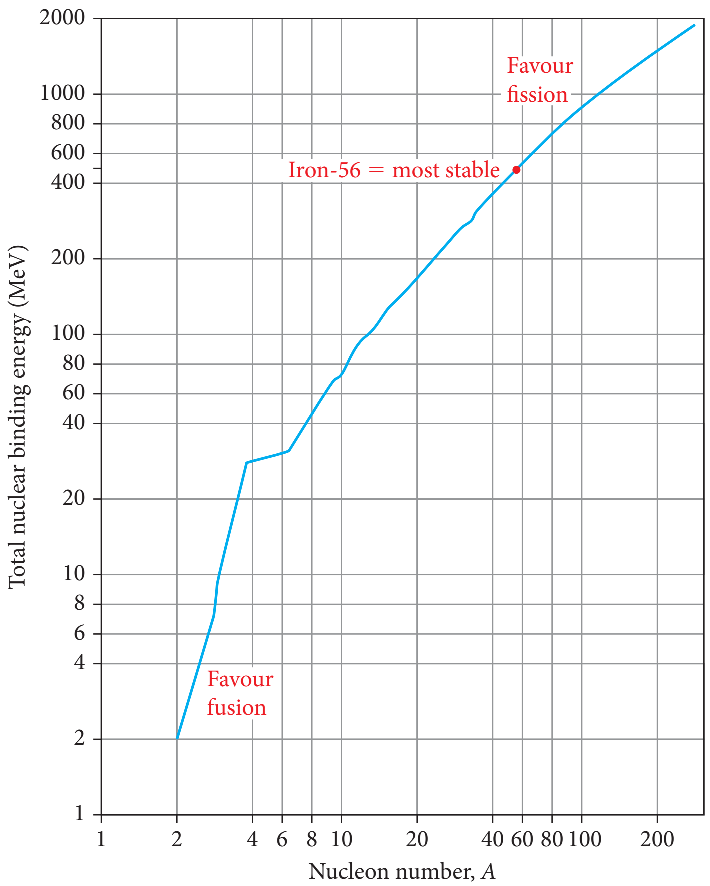


Figure 4.3 ▲ Total binding energy versus nucleon number. The scale is logarithmic (base 2).

Nuclear binding energy

The total energy needed to hold a nucleus together is called the **nuclear binding energy**. This is shown graphically in Figure 4.3 for nuclides up to uranium, which is the heaviest naturally occurring element.

The strong nuclear force acts most strongly between nearby nucleons. The binding energy per nucleon is the significant quantity when determining the stability of nuclides. The greater the binding energy per nucleon, the harder it is to pull the nucleus apart. Therefore, a nuclide is more stable when its binding energy is greater. Iron-56 is at the very top of the curve. It is the most stable of the nuclides.

Table 4.3 shows the total binding energy for nuclides as well as the binding energy per nucleon. Consider helium-4, which has a total binding energy of 28.29 MeV. It has 4 nucleons, so its binding energy per nucleon $= \dfrac{28.29\ \text{MeV}}{4} = 7.07$ MeV per nucleon.

Fusion is the coming together of two nuclides to form a new nucleus with a greater atomic number. The new composite nucleus is more stable because its binding energy per nucleon is greater. This is more likely to occur for nuclides that have an atomic number $Z < 56$.

Table 4.3 Binding energy and binding energy per nucleon for some nuclides

Element	Binding energy (MeV)	Binding energy per nucleon (MeV)
Deuterium (Hydrogen-2)	2.23	1.12
Helium-4	28.29	7.07
Lithium-7	40.15	5.74
Beryllium-9	58.13	6.46
Iron-56	492.24	8.79
Silver-107	915.23	8.55
Iodine-127	1072.53	8.45
Lead-206	1622.27	7.88
Polonium-210	1645.16	7.83
Uranium-235	1783.80	7.59
Uranium-238	1801.63	7.57

Binding energy per nucleon governs stability. The higher the binding energy per nucleon the more stable the nuclide.

- **Fusion is favoured for light nuclides ($Z < 56$).**
- **Fission is favoured for heavy nuclides ($Z > 56$).**

Fission is the splitting of a heavy nucleus ($Z > 56$) into fragments with lower atomic numbers. The new nuclei, called fission fragments, are more stable than the original nuclide because of their lower binding energies.

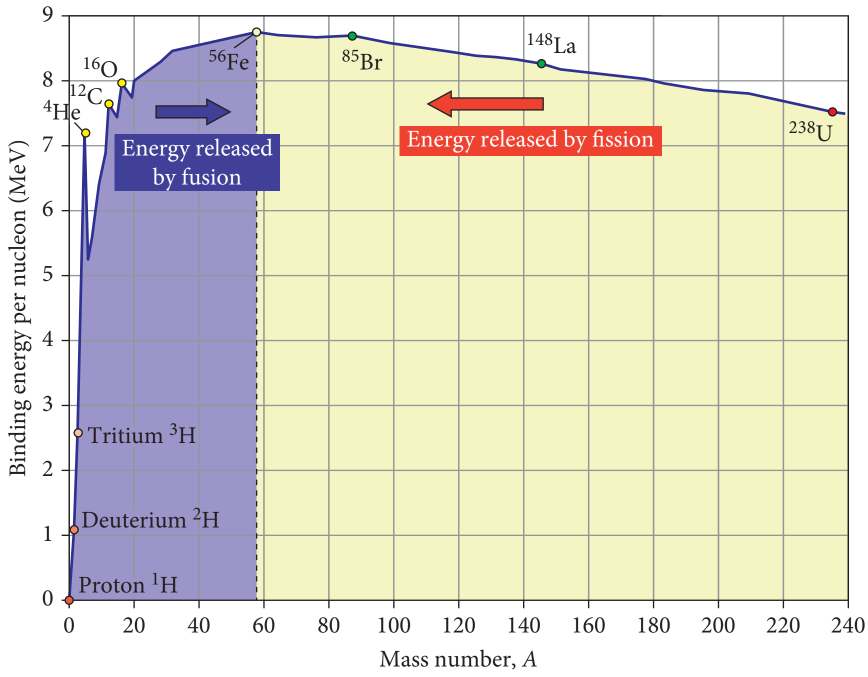
Fusion versus fission

Fusion reactions release much more energy than fission reactions. We have seen that fusion is favoured for elements up to Fe-56. For light elements, the curve in Figure 4.4 is quite steep, which means that any fusion reaction will release a relatively large amount of energy when the new nuclide is formed. For example, Figure 4.4 shows that, for tritium, the binding energy per nucleon is about 2.9 MeV. This is higher than the binding energy per nucleon for the proton (0 MeV) and deuterium (1.1 MeV). Fusion of a proton with deuterium to produce tritium releases about 1.8 MeV of energy per nucleon. This amounts to the release of approximately 62% of the original binding energy per nucleon.

At the other end of the graph, fission is favoured over fusion. The binding energy per nucleon of uranium-235 is about 7.6 MeV. For the two most common fission fragments the binding energy per nucleon is about 8.6 MeV. Taking account of both fission fragments, the difference is about 2.0 MeV. For fission, the release of energy per nucleon is about 26% of the original binding energy per nucleon.

Fusion reactions therefore release a greater proportion of the mass–energy available than do fission reactions.

Figure 4.4 ▶ Binding energy per nucleon as a function of nucleon number, A.



QUESTION SET 4.1

Remembering

1 List, in order of importance, the four forces that act within the nucleus.

2 Write down Einstein's mass–energy equation. Define each term.

3 How does nuclear binding energy per nucleon affect stability?

Understanding

4 Sketch the stability curve in terms of numbers of neutrons and protons. Show the regions where it is probable that these radioactive emissions will occur.

a Beta minus

b Positron

c Alpha

5 Define 'electron-volt'. Why are units of electron-volt used?

Applying

6 Sketch the binding energy per nucleon versus atomic mass number. On the sketch, show:

a the most stable nuclide.

b why fusion is preferred for light elements and fission is preferred for heavy elements.

c the energy per nucleon difference between lithium-6 and lithium-7.

d the energy per nucleon difference between uranium-235 and xenon-130.

7 A positron has a rest mass of 0.0005 u. Find its energy equivalent in:

a electron volt, eV.

b joules, J.

Analysing

8 All caesium is found naturally as the stable nuclide $^{133}_{55}Cs$. Use the stability curve to explain why caesium-123 is a positron emitter and caesium-141 is a beta minus emitter.

Nuclear energy: fission

The energy released in nuclear reactions, including fission and fusion, is enormous compared to the energy released in the most explosive chemical reaction. The differences are typically of the order of 1–10 billion times more energetic.

Nuclear fission is the process by which a nucleus splits into two fragments. In general, the fragments are rarely the same size, so it is incorrect to say that the atom 'splits in half'. In the process, neutrons are released and energy stored as binding energy is released.

Nuclear fission is triggered by the absorption of a neutron. Irène Joliot-Curie, daughter of Marie Curie, was the first to identify the products of nuclear fission. The process was first suggested to the German chemist Otto Hahn by Ida Noddack, a Hungarian chemist. Lise Meitner, who worked with Hahn, developed a model to explain the process, which was written up by her nephew Otto Frisch, the first person to use the term 'fission' in the literature. At Meitner's suggestion, Hahn and Fritz Strassmann undertook similar experiments to those of Joliot-Curie. The two men received the Nobel Prize in 1944. Meitner's decisive contributions were not recognised.

Enrico Fermi was the first to control nuclear fission. On 2 December 1942, in a squash court under the stadium at the University of Chicago, the first self-sustaining nuclear reactor began operation. Five days later, the Japanese attacked Pearl Harbour, which led to the atomic bomb being dropped on Hiroshima in 1945.

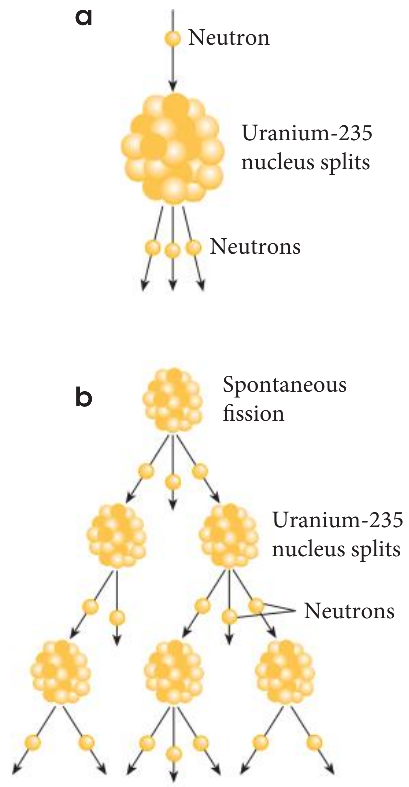


Figure 4.5
Nuclear fission. a) A slow neutron causes a uranium-235 nucleus to split, releasing three fast neutrons. b) A chain reaction occurs, if, for example, two of the released neutrons cause further nuclear fission in other uranium nuclei. Vast amounts of energy can be released.

Fission chain reaction

Uranium-235 can absorb a **'slow' neutron** or **'thermal' neutron** that has about 5–10 keV of energy. The neutron is absorbed in the uranium nucleus and forms a composite nucleus. This composite nucleus then splits into two fragments, each with lower atomic mass than the uranium-235. On average, between two and three neutrons are also released. The neutrons released from the fission of uranium-235 are **'fast' neutrons** and do not usually get absorbed in uranium-235 nuclei. This is a natural process that rarely amounts to anything substantial. However, if conditions are right, then the process can be used to harness the energy of the mass defect, which is about 200 MeV per fission event.

A chain reaction occurs when more than one of the neutrons released from the initial fission event cause new events to occur. In Figure 4.5, each fission event produces two or three neutrons, some of which go on to cause new fission events. This rapidly multiplies to huge numbers of fission events. Unless this is carefully controlled, a runaway explosion will occur.

Controlled fission is used in nuclear power stations. Thermal nuclear power stations use thermal (slow) neutrons to release energy. This energy ends up as heat to produce steam that powers the turbines and generators of the electricity production plant. **Fast breeder reactors** use the capture of fast neutrons in uranium-238 to breed the transuranic, fissile plutonium-239 nuclide via two beta decays (see page 99).

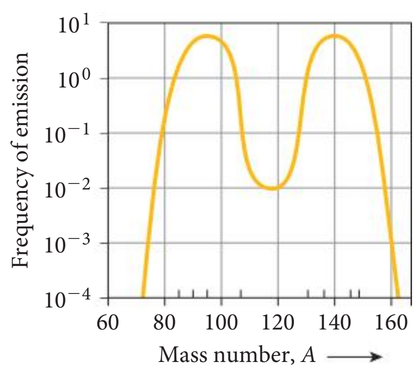


Figure 4.6
The most likely fission fragments from U-235 have atomic mass numbers around 95 and 140. The vertical scale is logarithmic – equal distances along the axis represent equal ratios, in this case × 10.

Fission fragments

Nuclides that are capable of undergoing nuclear fission after absorbing a neutron are said to be **fissile**. Fissile nuclides are very uncommon. Uranium-235 and plutonium-239 are readily fissile and undergo nuclear fission with low-energy 'slow' neutrons (in the range 0.02 eV to several keV). Uranium-238 is only slightly fissile; it requires a very high-energy neutron (≥1 MeV) to induce fission. Thorium-232 can absorb a neutron and beta decay to become uranium-233, which is also fissile.

A uranium-235 nucleus may split in many different ways. Radiochemical analysis shows that most fission fragments have an atomic mass number between 72 and 158 and an atomic number between 30 and 63. The splitting of a fissile nucleus into two equal parts is rare – about 0.01%. The most likely mass numbers are around 95 and 140 (10%), as shown in Figure 4.6. The graph also shows that a break-up of a radioactive nucleus into two equal parts is quite rare.

More than 40 different pairs of fission fragments of uranium-235 have been found. A typical fission reaction is given below. Note there is always a composite nucleus formed before the fission occurs:

$$ {}^{1}_{0}\text{n} + {}^{235}_{92}\text{U} \rightarrow {}^{236}_{92}\text{U}^{*} $$
$$ {}^{236}_{92}\text{U}^{*} \rightarrow {}^{92}_{36}\text{Kr} + {}^{141}_{56}\text{Ba} + 3\,{}^{1}_{0}\text{n} $$

Krypton-92 and barium-141 are known as the **fission products** or **fission fragments**. In this fission event, three neutrons are freed from the uranium nucleus when it splits. The fission fragments have relatively large kinetic energies. On average, about 2.2 MeV of energy is released per fission event (thermal energies are a thousand times smaller). It is the kinetic energies of the fission fragments that is transferred to neighbouring particles, causing heat transfer to the surrounds.

The fragment nuclides ${}^{92}_{36}\text{Kr}$ and ${}^{141}_{56}\text{Ba}$ are both unstable. The ${}^{141}_{56}\text{Ba}$ nuclide undergoes β^- decay:

$$ {}^{141}_{56}\text{Ba} \rightarrow {}^{141}_{57}\text{La} + {}^{0}_{-1}\beta + \bar{\nu} \text{ (half-life 18 minutes)} $$

Lanthanum-141 also undergoes β^- -decay:

$$ {}^{141}_{57}\text{La} \rightarrow {}^{141}_{58}\text{Ce} + {}^{0}_{-1}\beta + \bar{\nu} \text{ (half-life 3.6 hours)} $$

Cerium-141 also undergoes β^- decay to praseodymium-141:

$$ {}^{141}_{58}\text{Ce} \rightarrow {}^{141}_{59}\text{Pr} + {}^{0}_{-1}\beta + \bar{\nu} \text{ (half-life 32 days)} $$

Other common fission fragment pairs produced in the fission of uranium-235 include xenon-140 and strontium-94, and tin-132 and molybdenum-101. Some of these fission fragment nuclei can themselves be struck by, and absorb, a neutron, forming a different radionuclide, which may then radioactively decay. Fission fragments that absorb neutrons are called **neutron poisons**.

Liquid drop model of fission

A sphere is the least energetic way in which nucleons can be arranged in a nucleus. When the nuclear sphere is distorted by an extra neutron, the nucleus can split into two. This is similar to the way a drop of water divides when extra droplets are added.

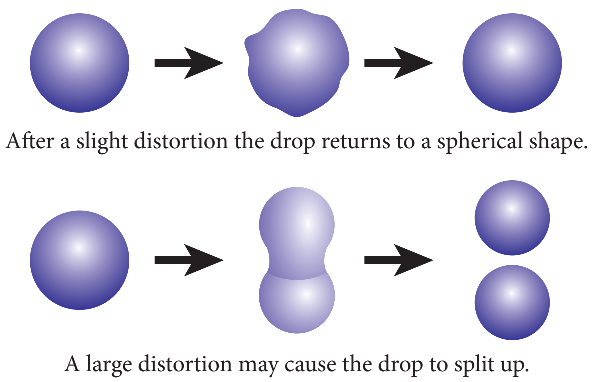


Figure 4.7 ▲ Liquid-drop model. Large distortions of the nucleus cause it to split up.

Fission occurs when neutron absorption in uranium and plutonium nuclei causes splitting into two, usually unequal, fragments. Neutrons are released and can be controlled (to sustain) or deliberately uncontrolled (to magnify) the original effect.

Mass defect in fission

In nuclear fission of uranium-235, the total mass before fission is greater than the total mass after the fission event. The mass difference is what is converted into energy. This energy is transferred via the fission products. The large daughter nuclei carry most of this energy as kinetic energy. The released neutrons also have kinetic energy.

For example, when uranium-235 undergoes fission to produce krypton-92 and barium-141, three neutrons are released:

$$ {}^{1}_{0}\text{n} + {}^{235}_{92}\text{U} \rightarrow {}^{92}_{36}\text{Kr} + {}^{141}_{56}\text{Ba} + 3\,{}^{1}_{0}\text{n} $$

There are 236 nucleons before and after this fission event, yet the mass of the products is less than the mass of the original neutron and uranium-235 nuclide. This is the mass defect.

Rather than use the mass of a nucleon, 1.660×10^{-27} kg, we shall use its equivalent in unified atomic mass units, u (Table 4.4). 1 u is defined as the mass of $\frac{1}{12}$ of the mass of a neutral carbon-12 atom. This is almost the same value as the mass of one nucleon. Why? Because the mass is a common factor for each calculation. If we want to convert back to units of kilogram at the end, we need only multiply our answer by the conversion factor:

$$1\text{ u} = 1.660 \times 10^{-27}\text{ kg}$$

One mole of particles is the same number as the number of atoms in 12 g exactly of carbon-12. This is Avogadro's number, $N_A = 6.0 \times 10^{23}$ particles. Tying Avogadro's number to carbon-12 means that, to a good approximation, the mass number (number of nucleons) of a nuclide, expressed in grams, contains 6.0×10^{23} particles.

Table 4.4 Particles in a fission event involving uranium-235

Particle	$^{1}_{0}$n	$^{235}_{92}$U	$^{92}_{36}$Kr	$^{141}_{56}$Ba
Mass (÷ 1.660×10^{-27} kg)	1.009	235.044	91.926	140.914

In atomic mass units:

Initial mass = 1.009 u + 235.044 u = 236.053 u
Final mass = 91.926 u + 140.914 u + 3 × 1.009 u = 235.867 u
Mass defect = 236.053 u − 235.867 u = 0.186 u

It is this mass defect, Δm, that is converted into energy, ΔE:

$$\Delta E = (\Delta m)c^2$$

At this point, it is necessary to convert back to SI units:

$\Delta m = 0.186\text{ u} \times 1.66 \times 10^{-27}\text{ kg u}^{-1} = 3.090 \times 10^{-28}\text{ kg}$
$\Delta E = 3.090 \times 10^{-28}\text{ kg} \times (3.0 \times 10^{8}\text{ m s}^{-1})^2$
$\Delta E = 2.78 \times 10^{-11}\text{ J}$

This is a very small amount of energy; however, 250 gram of pure uranium contains more than 10^{23} nuclides. If all of these were to undergo fission, the effect would be enormous:

$$\Delta E = 2.78 \times 10^{-11} \times 10^{23}\text{ J} = 2.78 \times 10^{12}\text{ J (2.78 TJ)}$$

WORKED EXAMPLE 4.2

A thermal neutron, mass 1.01 u, causes fission of U-235 (235.04 u). The fission fragments and their masses, in unified mass units, are Rb-93 (92.92 u) and Cs-141(140.92 u).

a How many fast neutrons are released in this fission event? (3 marks)

b Write the nuclear fission reaction using correct nuclide and nucleon symbols. (2 marks)

c What is the mass defect in this event in:

  i unified mass units? (1 mark)

  ii kilogram? (1 mark)

d How much energy is released? Give your answer in joule. (2 marks)

Answers	Logic	
a Nucleon number is conserved.	Use the correct conservation rule.	1 mark
Total nucleons before = 1 + 235 = 236	Calculate correct sum.	$\frac{1}{2}$ mark
Total nucleons after = 93 + 141 + x	Calculate correct sum.	$\frac{1}{2}$ mark
⇒ $x + 234 = 236$		
⇒ $x = 2$		
⇒ Two neutrons were released.	Calculate the answer.	1 mark

b	${}^{1}_{0}n + {}^{235}_{92}U \rightarrow {}^{93}_{37}Rb + {}^{141}_{55}Cs + 2\,{}^{1}_{0}n$	Use the correct atomic numbers.	1 mark
		Place fission reactants and products in correct order.	1 mark
c	Mass defect = (1.01 + 235.04) − (92.92 + 140.92 + 2 × 1.01) u	Substitute the correct values and add to get answer.	1 mark
	⇒ **i** mass defect = 0.19 u		
	⇒ **ii** mass defect = $0.19\,u \times 1.66 \times 10^{-27}\,kg\,u^{-1}$	Convert u to J.	1 mark
	$= 3.15 \times 10^{-28}\,kg$		
d	Energy released:		
	$\Delta E = (\Delta m)c^2$		
	$\Rightarrow \Delta E = 3.15 \times 10^{-28}\,kg \times (3.00 \times 10^{8}\,m\,s^{-1})^2$	Substitute the correct values.	1 mark
	$\Rightarrow \Delta E = 2.84 \times 10^{-11}\,J$	Calculate the answer.	1 mark

Try these yourself

1 In a nuclear reaction that starts with thorium, a thermal neutron, mass 1.01 u, causes fission of U-233 (233.044). The fission fragments and their masses, in unified mass units, are Mo-104 (103.91 u) and Sn-126 (125.91 u).

- **a** How many neutrons are released in this fission event? (3 marks)
- **b** Write the nuclear fission reaction using correct nuclide and nucleon symbols. (2 marks)
- **c** What is the mass defect in this event in:
  - **i** unified mass units? (1 mark)
  - **ii** kilogram? (1 mark)
- **d** How much energy is released? Give your answer in joule. (2 marks)

2 In a fast breeder reactor, a fast neutron, mass 1.01 u, causes fission of Pu-239 (239.05 u). One of the two fission fragments is Tc-104 (103.91 u). Three neutrons are released. The mass defect in this event is 0.19 u.

- **a** What is the nuclide symbol for the second fission fragment? (1 mark)
- **b** Write the nuclear fission reaction using correct nuclide and nucleon symbols. (2 marks)
- **c** What is the atomic weight of antimony-133? (1 mark)
- **d** How much energy is released? Give your answer in joule. (1 mark)

QUESTION SET 4.2

Remembering

1. What is nuclear fission? How does it occur?
2. Draw a series of diagrams and label them to show the liquid-drop model of nuclear fission.

Understanding

3. What does Figure 4.6 show? It is not correct to say, 'In fission, nuclides mostly split in half.' Why?
4. Why is fission favoured over fusion for uranium?
5. Why are some fission fragments called neutron poisons?
6. What is 'mass defect' when applied to fission? Why is it important for fission?

Applying

7. Each fission event produces very little energy. Why then is nuclear energy used for electrical power stations and explosive devices?
8. Boron-10 is a neutron poison. Boron-11 is an alpha emitter. Write a sequence comprising two nuclear equations to represent these two statements correctly.

Analysing

9 A thermal neutron, mass 1.01 u, causes fission of U-233. The fission fragments and their masses, in unified mass units, are Rh-105 (104.906 u) and Cs-125 (125.909 u).

a How many neutrons are released in this fission event?

b What is the mass defect in this event in:

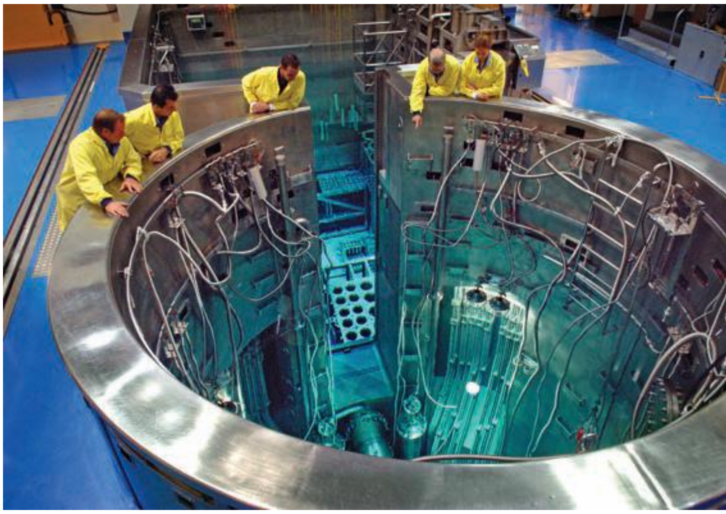
i unified mass units?

ii kilogram?

c How much energy is released? Give your answer in joule.

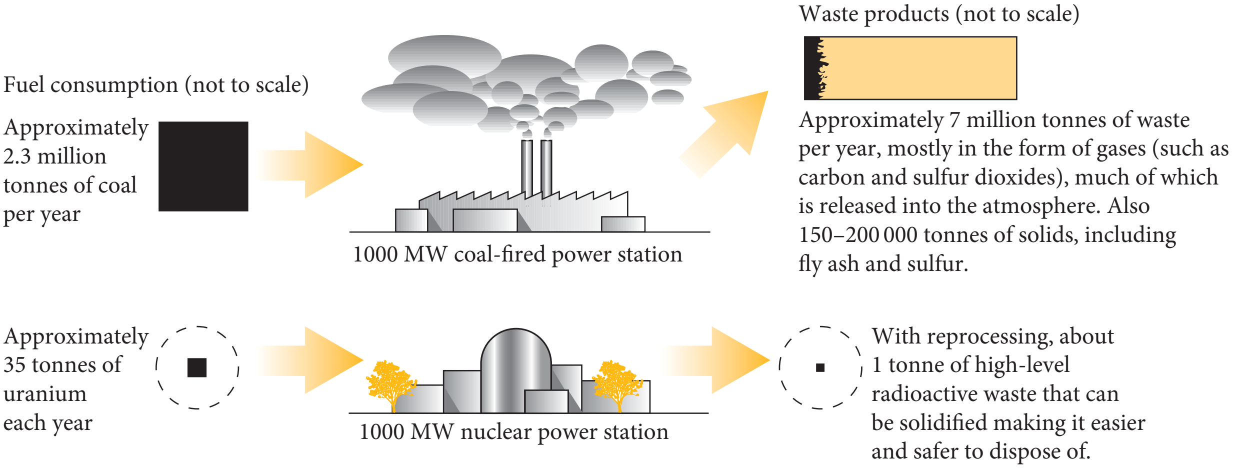
Nuclear reactors

A nuclear reactor is a device in which controlled fission is used to produce both new substances and release energy. Nuclear reactors produce far more energy per kilogram than the burning of fossil fuels. In normal operation, nuclear reactors release no radioactive materials or other chemicals, such as carbon dioxide. Coal-fired power stations release pollutants such as fly ash, oxides of carbon, sulfur and nitrogen along with radioactive chemicals. In both fuel consumption and waste production a nuclear power station is far more fuel efficient and far less wasteful, including unintended emissions, than a coal-fired power station.



Newspix/Alan Pryke

▲ Figure 4.8
Australia's Open Pool Australian Lightwater reactor (OPAL) at Lucas Heights, NSW



▲ Figure 4.9
Comparison between the efficiency of coal-fired and nuclear-powered electricity production

Chain reaction in a nuclear reactor

The chain reaction can be controlled for nuclear power generation. One neutron produces one fission neutron in a **controlled chain reaction**. If more than one neutron is produced on average, then a runaway reaction occurs. If the average number of neutrons produced is less than one, then the reaction dies away.

It is quite difficult to establish and maintain a chain reaction. The proportion of uranium-235 in naturally occurring uranium ore is very small (0.65%) compared to uranium-238 (99.3%).

The proportion of the other natural uranium isotope, uranium-234, is negligible (0.05%). Uranium-238 acts almost exclusively as a neutron poison. Uranium-238 beta decays, leading to the production of transuranic elements, the most important of which is the highly fissile plutonium.

Enrichment

For fission to occur in uranium-235 the initial neutron must be a 'slow' neutron. The neutrons produced in fission events are 'fast' neutrons. The probability of these neutrons causing new fission events in uranium-235 is very small. It is more likely that they will be captured by uranium-238 or some other neutron poison already produced by earlier fission events. These neutron poisons hold onto the neutrons and emit α, β or γ radiation.

In order to ensure sustained fission, the proportion of uranium-235 must be increased. This **enrichment** process is achieved by separating uranium-235 out of naturally occurring uranium ore. This separated amount is then put back into a quantity of naturally occurring uranium. This increases, or enriches, the proportion of uranium-235 in the quantity. It is a complex and expensive process. In order to guarantee an **uncontrolled chain reaction**, enrichment may be as high as 97% uranium-235. For a controlled reaction, 1–4% enrichment is sufficient.

Moderator

In order to reduce the energy of the neutrons produced by fission events, a moderator is used. The **moderator** is a material with nuclides that have slightly larger masses than the neutron – hydrogen ($^{1}_{1}H$), deuterium ($^{2}_{1}H$) and tritium ($^{3}_{1}H$), for example. Neutrons share their energy with these nuclides, through multiple collisions. They rapidly lose energy, which increases the probability of neutrons entering a uranium-235 nucleus and causing fission.

Reactor vessel

In a reactor, a controlled chain reaction will not proceed unless most of the neutrons available can be used. Apart from absorption in neutron poisons, some neutrons will escape from the fuel. This is because the absorption of a neutron and subsequent fission occurs only when there is a head-on collision between nucleus and neutron. The reactor vessel is designed to have the right surface area-to-volume ratio, and is made of a high nucleon number material, so that the neutrons are reflected back into the sample.

Control rods

On average, the number of neutrons produced by fission events must be equal to the number of fission-producing neutrons in order to sustain a nuclear reaction. Sometimes the reaction threatens to run away, so **control rods** containing neutron poisons, such as boron-10, are moved into the fuel to lower the number of neutrons in the sample. The control rods are removed when the chain reaction starts to produce too few neutrons.

Australian nuclear reactors

Australia does not have any nuclear power reactors. We rely mainly on the burning of coal for our electricity production, even though we have large deposits of uranium, which we mine and export.

Australia does have one small nuclear reactor for research and for radiopharmaceutical production. This is the Open Pool Australian Lightwater reactor (OPAL), which is administered by the Australian Nuclear Science and Technology Organisation (ANSTO). OPAL is located to the south of Sydney at Lucas Heights. As we saw in the previous chapter, many radiopharmaceuticals are short-lived, so it is important to have a local source. Maintaining even a small nuclear reactor also gives Australia membership of several important international organisations and committees that oversee the development and use of nuclear technology internationally.

VIRTUAL TOUR OF OPAL RESEARCH REACTOR

Take a virtual tour of OPAL here.

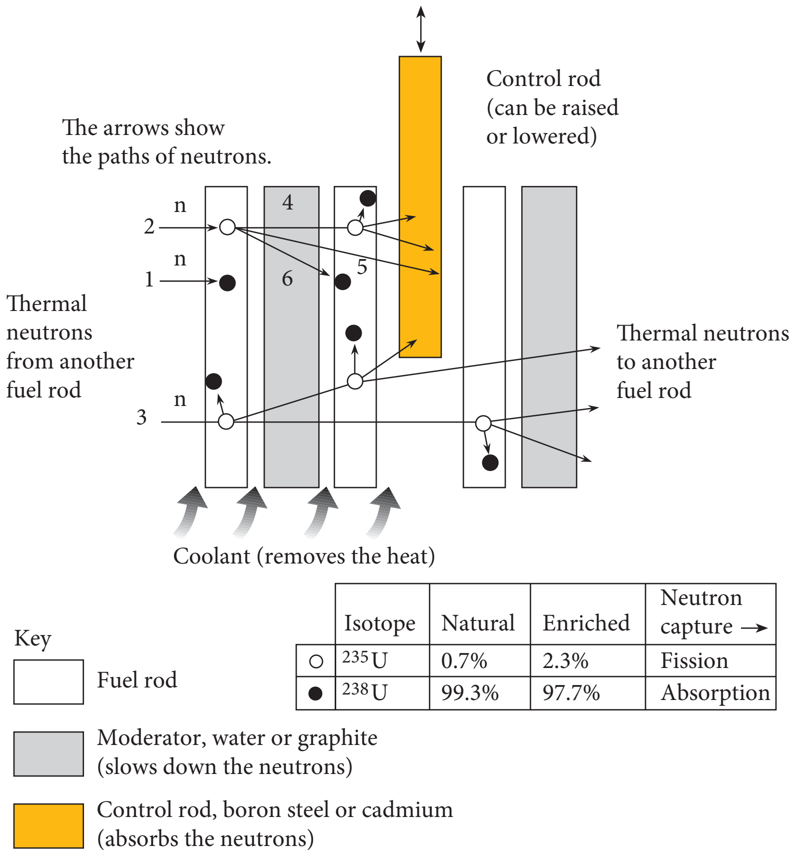


Figure 4.10 Controlling a chain reaction in a nuclear reactor. Some neutrons are numbered to explain the process. Neutron 1 is captured by a fuel rod, neutrons 2 and 3 cause nuclear fission, releasing more neutrons, neutron 4 causes further fission, neutron 5 is absorbed by a control rod and neutron 6 is absorbed by a fuel rod.

Table 4.5 Components of a thermal nuclear reactor and their functions

Component	Function
Nuclear fuel	Usually 1–4% enriched uranium-235 is used as the nuclear fuel, which is placed in airtight containers and located in the reactor core.
Moderator	It is used to slow down the high-energy, fast-moving neutrons produced by the splitting of the uranium atom to continuously cause fission in uranium-235. It must have a low mass number, readily gain kinetic energy from collisions with neutrons, and be a weak absorber of neutrons.
Control rods	In order to control the chain reaction, some neutrons must be absorbed. This is done by means of control rods that are inserted in the reactor. These rods are made of substances that readily absorb neutrons (neutron poisons), such as cadmium or boron steel.
Coolant	As a large amount of energy is produced in each fission reaction, and this is carried as kinetic energy by the fission products, the temperature of the reactor would increase unless a coolant was used. The coolant material is used to absorb this heat so that it can be used for energy purposes outside the reactor.
Radiation shield	A nuclear reactor is enclosed in a thick concrete vessel. The nuclear reactor must be shielded to protect workers and nearby communities from the large quantities of γ-radiation and penetrating neutrons that are released.

Case study

Dr Ross Whitfield

▲ Figure 4.11
Dr Ross Whitfield points to the building he works in on a picture of the Spallation Neutron Source site. In the background you can see the long, straight line of the linear accelerator and a circular synchrotron ring.

Dr Whitfield grew up in Mildura, a small city located in the north-west of Victoria, along the Murray River. Mildura is known for its grape and citrus production. He enjoyed growing up with what a regional town has to offer, in particular freedom and an outdoor lifestyle.

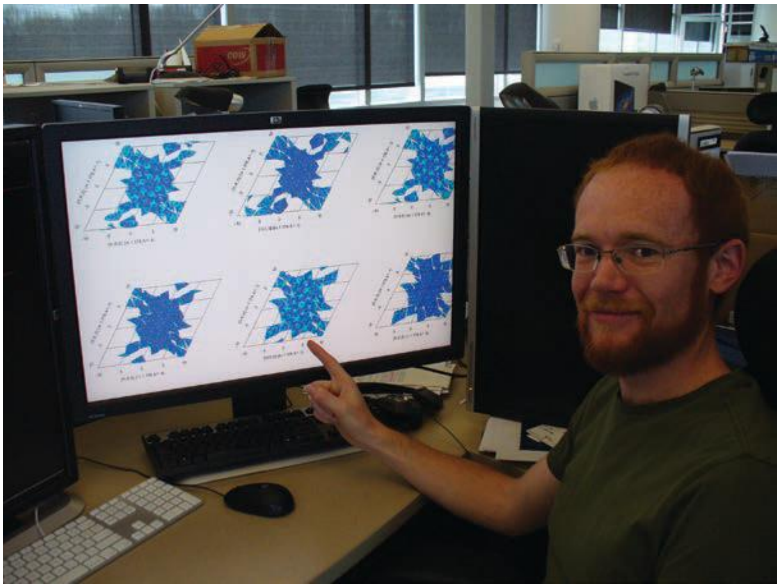
He moved to Canberra to study at the Australian National University (ANU). Dr Whitfield is interested in all areas of science, but chose to study physics because he thought it offered the most exciting future prospects.

After the first three years of his Bachelor of Science (BSc) with honours, Dr Whitfield spent a year working at the Australian Nuclear Science and Technology Organisation (ANSTO), in Sydney, as part of a year-in-industry program. During his time at ANSTO he studied a range of aluminium and titanium metal alloy powders and participated in experiments at the European Synchrotron Radiation Facility in Grenoble, France, where X-ray diffraction data were collected. The analysis performed on the diffraction data allowed the behaviour of the material at different temperatures to be understood. Meeting and collaborating with a wide range of scientists at ANSTO was an important part of his experience. Dr Whitfield enjoyed his time ANSTO, which influenced him to undertake a PhD in science and pursue a career as a scientist.

After completing his year in industry, Dr Whitfield returned to the ANU to finish the final year of his BSc and do a PhD. For his PhD research he studied the local structure of a ferroelectric ceramic material. A ferroelectric material is one that has electric dipoles (positive and negative charges on molecules) that naturally line up in the same direction. Ferroelectric materials are used in electronic components such as capacitors and transistors.

Dr Whitfield's research involved experiments in which he collected both X-ray and neutron-diffraction data from the ferroelectric material. He used facilities such as the neutron source at the Los Alamos National Laboratory and the X-ray facility at the Advanced Photon Source in the US to perform experiments. He also did many experiments at ANSTO, where he used the WOMBAT high-intensity diffractometer. The WOMBAT diffractometer uses a beam of neutrons from the OPAL reactor to create a neutron-diffraction pattern. The pattern is due to interference of neutrons scattered from different atoms or layers of atoms in the sample material. He also used the ECHIDNA high-resolution neutron diffractometer and an X-ray beamline at the Australian Synchrotron in Melbourne.

The analysis of the diffraction patterns was done using computer modelling of the arrangement of, and interactions between, atoms in the material. This modelling allowed a greater understanding of the physical and chemical properties of the material on the nano-scale.



▲ **Figure 4.12**
One of the diffraction patterns

Dr Whitfield completed his PhD in 2013 and is now working at the Oak Ridge National Research Laboratory in Tennessee, USA. The research programs at Oak Ridge focus on materials, energy and neutron science, as well as high-performance computing and national security. The facility is home to several of the world's fastest supercomputers, a spallation neutron source (SNS) and the High Flux Isotope Reactor.

This is a historically significant facility because it was first built as part of the Manhattan Project – the American atomic bomb project in World War II. The first reactor built at this site, the X-10 graphite reactor, was used to produce plutonium for nuclear weapons made at Los Alamos. Since then, the nuclear reactors at the facility, including the X-10, have mainly been used for scientific research and the production of radiopharmaceuticals.

Dr Whitfield is part of the Neutron Data Analysis and Visualization Division at the SNS. His responsibilities involve the development of data-analysis and data-reduction software with the other members in his group, to support the science conducted at the SNS by the instrument scientists and users of the instruments. Scientists come from all over the world to use the neutron instruments at the SNS. Dr Whitfield works on advanced data-analysis techniques for these experiments. The development of these techniques is done in collaboration with scientists at other neutron facilities from around the world, including the neutron spallation source (ISIS) in the UK and the Institut Laue-Langevin research reactor in Grenoble, France.

The development of advanced data-analysis techniques is important as it allows visiting scientists to the facility to obtain as much information as possible from the data they collect. Dr Whitfield works closely with the instrument scientists and with international collaborators from Germany, USA, Australia and other parts of the world.

Dr Whitfield has found the opportunity to participate in science on an international level and at a world-leading facility to be an exciting and fulfilling career.

Questions

1 Identify and briefly describe the following from their acronyms: ANSTO, ANU, ECHIDNA, OPAL, WOMBAT, SNS, ISIS.

2 Produce a timeline of Ross's academic development up until he received his PhD.

3 What are the materials Ross has studied? What are they used for?

4 List the ways in which Ross has collaborated with Australian and international scientists and related research facilities.

5 Compare and contrast Ross's current work with his early research interests.

Thermal nuclear reactors

TYPES OF NUCLEAR REACTOR

Find out about several types of nuclear reactor.

The purpose of a nuclear power reactor is to release nuclear energy at a controllable rate. The reactor is where the energy from the nuclei of atoms is released during fission. The reactor does not generate electricity directly. It is a heat source that is used to produce steam as the working fluid to rotate a turbine connected to the electrical energy generator. There are several types of **thermal reactors**, but they all operate on the same principle of heat exchange to produce steam (see page 63).

Advanced gas-cooled reactor (AGR)

One type of thermal nuclear reactor, the advanced gas-cooled reactor (AGR) is shown in Figure 4.13.

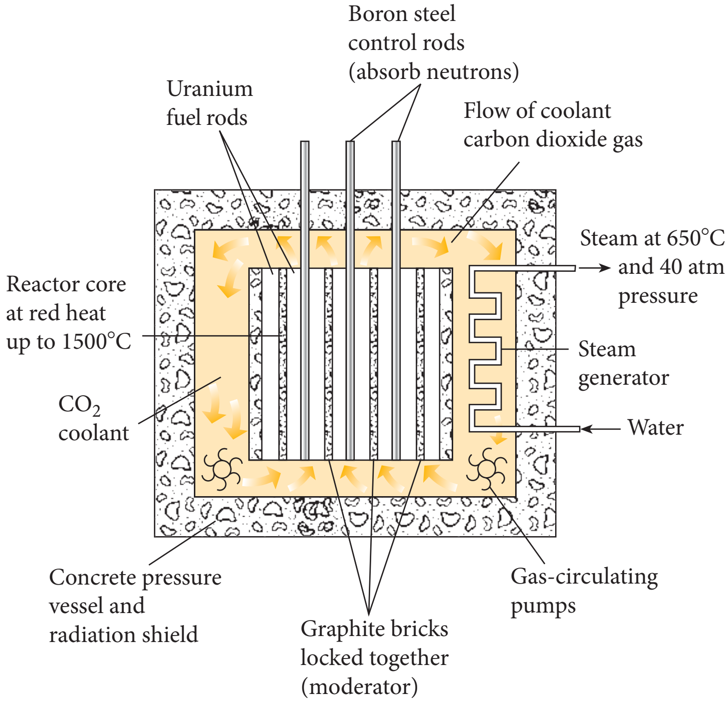


Figure 4.13 ▶ Advanced gas-cooled reactor (AGR) is one type of thermal nuclear reactor.

In the AGR system, the pile, or core, is surrounded by carbon dioxide. The steam system is independent of the nuclear reactor pile. Carbon dioxide acts as a coolant. The heat exchanger takes the heat from the reactor and transfers it, via the carbon dioxide, to pipes containing water. This water is kept at high pressure so it boils into steam at very high temperatures. This steam is then directed at the turbines in the power station. The power station is physically separated from the reactor. The moderator is carbon in the form of graphite, and the control rods are made of boron steel, boron being the effective neutron poison.

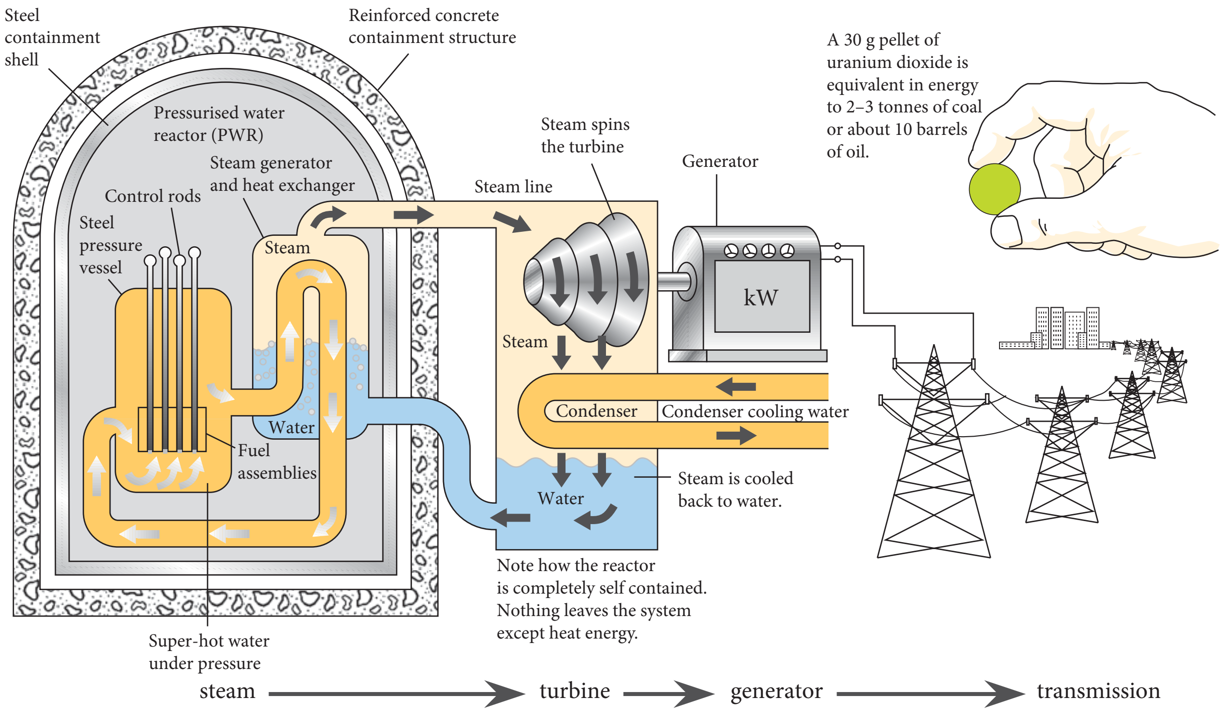
Pressurised water reactor

In the pressurised water reactor (PWR), the moderator is water that is kept under very high pressure within the reactor pile. The heat from this water is transferred to another water system, which is isolated from the reactor water. Through this heat exchange process, the steam is produced that turns the turbines in the electrical energy generation plant.

Fast breeder reactor

Fast breeder reactors use the fast neutrons produced during fission of uranium-235. These fast neutrons are captured by the 96–99% of the nuclear fuel that is uranium-238. The new nuclides then beta decay through neptunium to plutonium. Plutonium is fissile, so this is then chemically removed and used as the fuel for a plutonium reactor and in weapons. Reactors that 'breed' plutonium do so with 'fast' neutrons, hence 'fast breeder'.

a



Risks of using nuclear energy

Nuclear energy production starts with the mining of radioactive materials that are usually left underground and out of the way of human interactions. Nuclear power stations concentrate energy production in a few locations and are likely targets for militant individuals, organisations and nations. It can take thousands of years before much of the waste from spent fuels is considered safe. And let's not forget the use of nuclear weapons and nuclear deterrents by nations prepared to advance their causes or protect their dominant positions in the world.

Nuclear waste

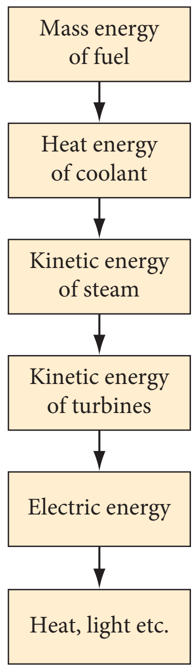
Nuclear energy production by fission reactors produces radioactive waste. The disposal of radioactive waste is one of the major problems faced by the nuclear power industry. Table 4.6 is a comparison of waste produced by nuclear reactors and coal-fired power stations.

Table 4.6 Waste produced by power plants

Type of power	Capacity	Waste
Nuclear power station	1000 MW	25 tonnes of radioactive material
Coal-fired power station	1000 MW	Millions of tonnes of carbon dioxide and sulfur dioxide

Radioactive waste products are classified into three categories: high, medium and low level. High-level wastes are radioactive. They continue to release energy during the decay process. This energy may be radioactivity or heat. Systems must be put in place to reduce harmful effects. Table 4.7 shows a classification of radioactive waste.

b



▲ **Figure 4.14** The PWR transfers heat energy to the power station, which generates and transmits electrical energy. a) System from reactor to consumer b) Flowchart showing energy transformations from reactor to consumer

Table 4.7 Classification of radioactive waste

Waste category	Description	Storage
High level	Used fuel rods, highly radioactive, take about 1000 years to return to the same radioactive level as the uranium ore that was originally used. It takes about 5 million years to become harmless	Must be stored in shielded containers to prevent radiation and must also be cooled to stop overheating
Medium or intermediate level	Other reactor components in the reactor core such as fuel containers, gauges, pipes	Requires shielding but not cooling
Low level	Used protective clothing, water from showers and water from cleaning protective gear	Can be released to the environment after being diluted

Scientific literacy: Nuclear disasters

Chernobyl, 26 April 1986

At Chernobyl, in Ukraine, the core of Reactor #4 exploded at about 1:23 a.m. on the morning of 26 April 1986. The reactor was being shut down for routine maintenance. During the shutdown period, tests of the effectiveness of recent changes to the reactor's control and maintenance power systems were planned. These were to occur when the reactor reached 20–30% output. Below 20% the reactor can become unstable. Safety procedures required the reactor to be completely shut down if it went below 20%.

During 25 April, the reactor was carefully brought down to 50%. At 11:10 p.m., the power was further reduced. An hour and eight minutes later, the operator correctly switched the control system to one that more effectively managed the reactor in this lowered power output state. However, the operator failed to enter an instruction to hold the power automatically between 20 and 30%. The power fell rapidly to a very dangerous 1% or less.

The operator took charge of the control rods, withdrawing them and stabilising the power at 7%. Despite this being a dangerous and unstable situation the operator proceeded to set up for the tests. The result was that very serious instabilities occurred, which should have meant immediate shutdown. The operator knew this at 1:22:30 a.m., but continued with the tests. Thirty-four seconds later the operator blocked the last but one automatic shutdown signal, so that the test could continue. Twenty-seven seconds later the reactor went into overdrive and the output raced upwards, smashing through the final automatic block about four seconds later. The result was a massive explosion in the core.

Science Photo Library/RIA NOVOSTI

Figure 4.15 ▲
The inside of the Chernobyl nuclear power plant after the disaster. The lid of the reactor was blown off.

The explosion caused radioactive material to be released into the building and driven into the ground beneath. The temperature generated, about 5000 K, meant that a fire started. This produced large amounts of gaseous, radioactive materials that were able to escape into the atmosphere. These were carried over, and deposited on, nearby villages, towns and cities, and were blown to the west, particularly over Scandinavia.

Fukushima, 11 March 2011

Getty Images/DigitalGlobe

▲ **Figure 4.16**
DigitalGlobe's WorldView-2 satellite image of Fukushima nuclear power station. The roof and wall panels around reactor Unit 1 have exploded and the containment structure around Unit 3 has been badly damaged. The water intake facility has been damaged, as well as the fuel oil tanks that probably power the diesel generators that pump water through the cooling apparatus during a power outage. The facility's coolant system was completely without function at the time the image was taken.

At 2:46 p.m. on 11 March 2011, there was a major earthquake 180 km off the coast of Honshu, Japan's largest island. The island was shifted a few metres to the east. The shoreline dropped half a metre. The earthquake produced a 23 m high tsunami. This had reduced to 15 m when it came ashore at the Fukushima nuclear power facility about 40 minutes later. A second tsunami arrived eight minutes later.

The Tepco nuclear power stations at Fukushima shut down automatically when the earthquake occurred. The earthquake did no direct damage to the power stations. There are enormous amounts of heat to be removed from the central core even after it has stopped acting as a nuclear reactor. Cold sea water is pumped into heat exchangers, then released into the ocean. When the tsunamis struck, they flooded the pumps and made them useless. The overheated radioactive material in the core started to react with water to produce hydrogen, which later exploded (Figure 4.16). This released radioactive materials into the air. The core itself could not be cooled, so melted the floor and moved down into the ground.

Japan is well-known for its seismic activity and related tsunamis. At Fukushima, the protective tsunami wall was 3 m high. The decision to build it this high was made with data that was more than 40 years old. No changes had been made, despite new evidence from more recent earthquakes in the region. Earthquakes of magnitude 7.5 or above occur on average every 12 years in this part of Japan. Magnitude 7.7 earthquakes in 1983 and 1993 caused tsunamis that started out at 14.5 m and 31 m respectively. A tsunami with a run-up height of 38 m struck the shoreline in Japan from a magnitude 7.6 earthquake in 1986.

These experiences should have informed the operators and Government regulators of the need for higher, tsunami-proof walls. A report by the Japanese Government Earthquake Research Committee recommended walls be built higher. It was due for release in April 2011, but the magnitude 9.0 earthquake occurred before it could be released and acted upon.

Questions

1 Make separate timelines of events that led to the nuclear accidents at:
  a Chernobyl.
  b Fukushima.
2 Use the timeline to describe and explain the cause of the Chernobyl disaster. Consider, for example, the effect of decisions by the operator on the stability of the reactor, the reaction rate, and the use of control rods.
3 What could have been done by Tepco and the Japanese government to reduce or avoid the disaster at Fukushima?
4 Heat was a major factor in both disasters. In what ways were the effects of heat the same or different at Chernobyl and Fukushima?
5 Compare and contrast the Chernobyl section and the Fukushima section with respect to:
  a general and relevant information.
  b adequacy of the scientific details.
  c implicit and explicit commentary by the author.

SAFETY AT WORK

Examine the safety record of different industries involved in primary energy production. Find other resources that come from a different position in the debate about energy production and safety.

Some nuclear waste is buried deep at sea, in underground storage containers or in geological formations that are unlikely to be disturbed. But how safe are these methods? Containers may decompose or be damaged by external forces, releasing the waste into the general environment. There is also the possibility that the waste could be stolen and converted into nuclear weapons.

Apart from the radioactive waste, the industry has other risks. The mining of uranium exposes people to radioactivity. The uranium ore has to be processed before being used in a power station, exposing another group of workers to the risks of radiation. Nuclear power stations produce large amounts of heated water from their cooling systems. If this is allowed to flow into surrounding lakes and streams, the ecological balance is upset. Very few accidents have occurred in the nuclear industry, with very small loss of life – all loss of life at work is to be deplored – compared to the coal-mining industry. In any one year, about 600 coal industry workers die worldwide. After 50 years of operation, the nuclear industry has yet to reach 100 worker deaths. Accidents in the hydroelectric industry can be catastrophic on populations downstream when dams burst – hydroelectric energy production has a record of deaths per year about five times that of the coal industry. These figures do not take into account the medium- and long-term effects of accidents and unintended emissions from all these industries.

Nuclear weapons

Humans can use (or misuse) the enormous amount of energy available when nuclear fuels are used. When the first atomic bomb was tested in the New Mexico desert in 1945, even those working on the project were surprised by its destructive power. The first U-235 bomb used in warfare was dropped on the civilians of the Japanese city of Hiroshima less than a month later, on 6 August. A Pu-239 bomb was dropped three days later on the city of Nagasaki. The blasts destroyed almost every building, killed nearly 45 000 people immediately and injured more than 150 000 more. Many more died as a result of the after-effects. The area was heavily radioactively contaminated for about 3 months, but the human misery lasted much longer. These bombs brought the Pacific war to an end.

The effects of the blast can still be seen nearly 70 years later. The long-term effects of the bombs in terms of cancers, genetic problems and other medical conditions are still being studied. Survivors of the blasts are called 'hibakusha'. Mutations of children *in utero* at the time of the bombing has led to widespread fears about genetic mutations from hibakusha. Statistically, this is an unsupported fear. There appears to be no difference in rates of abnormalities in the descendants of hibakusha and the general population. Nevertheless, hibakusha descendants continue to experience discrimination.

EINSTEIN

Read about Einstein's views and actions on pacifism.

The Manhattan Project

War in Europe was declared on 3 September 1939. It was only after the bombing of Pearl Harbor on 7 December 1941 that the United States of America declared war, first on Japan (8 December) and then Germany (11 December).

Albert Einstein was a scientist who believed passionately in the value of each person; consequently he was a pacifist. But he recognised the importance of opposing excessive military force, such as that of German Nazism.

On 2 August 1939, one month before war in Europe was declared and more than two years before America entered the war, Einstein wrote to US President Roosevelt about the possibility of producing a fission bomb within a reasonably short time. He cited the work of Enrico Fermi and Leó Szilárd in America and Irène and Frederic Joliot-Curie in France. He indicated concern about the capacity of German scientists, with access to uranium in Europe, to produce a fission bomb first. In October 1939, Roosevelt established a secret committee to investigate uranium research. This led to the Manhattan Project, which was dedicated to the production of atomic bombs.

Bomb-making principles are relatively simple. However, it took 4 years and billions of dollars to solve all the scientific and engineering problems.

Lise Meitner, the first to explain fission, was invited but declined to join the Manhattan Project. Some Project scientists, including Australia's Sir Marcus Oliphant, later expressed regret that they were involved.

HIROSHIMA AND NAGASAKI

Find out about casualties, radiation fallout and effects on humans of these atomic blasts.

Neutron radiation

Neutrons can be emitted spontaneously or as a result of the bombardment of nuclei with energetic particles. These neutrons are uncharged and can penetrate atoms and interact with other nuclei. These interactions can cause them to emit α and β particles and γ rays. The recoil nuclei or the charged particles emitted by the nuclei are very heavily ionising and so biologically damaging.

Because neutrons are so penetrating, and because the human body contains so much hydrogen, an external beam of neutrons constitutes an important radiation hazard. The damage done depends on the energy of the neutrons – fast neutrons being more of a hazard than slow ones.

ATOMIC ARCHIVE

Explore the complex history surrounding the invention of the atomic bomb.

QUESTION SET 4.3

Remembering

1 To what do the letters AGR, PWR and OPAL refer?

2 a What is a controlled chain reaction?
  b Define 'thermal nuclear power'.

Understanding

3 Compare the mass of coal to the mass of uranium needed for a 1 GW power station.

4 Why is uranium enriched?

5 In a nuclear reactor, why is it necessary to use:
  a a moderator?
  b control rods?

Applying

6 How are neutron poisons used positively in:
  a thermal nuclear reactors?
  b fast breeder reactors?

7 Plutonium has a half-life of 24 100 years. How would you store it?

Analysing

8 Use Figure 4.14(b) and further research to draw a Sankey diagram from energy input to energy output for a nuclear power station.

Reflecting

9 Compare three reasons for and three reasons against the involvement of scientists in the development and use of the nuclear bombs that destroyed Hiroshima and Nagasaki in 1945.

10 Would you recommend the development and use of nuclear power to help solve human-induced effects on climate? Explain your position.

Nuclear energy: fusion

Nuclear fusion is the process in which two nuclei come together to form a larger nucleus. In the process, energy becomes available. This can occur for elements up to atomic number 56 (iron). Making new nuclei by fusion requires enormous energy.

According to the **Big Bang theory**, fusion was possible anywhere in the universe, between the first 1 second and 3 minutes of the universe's history. During this period, particles began to form through fusion in a process known as **nucleosynthesis**, which we can still see occurring in stars. Over millions of years a variety of elements can be produced within stars by the fusion-generated process of nucleosynthesis.

You will study the Big Bang theory in more detail in Unit 4.

Early research into fusion was hampered by a **negative energy problem**. The energy needed to produce fusion reactions was greater than the energy derived from the reactions. But no longer. Fusion research has now reached the point where, in the foreseeable future, a fusion reactor will be used to generate electricity, which will feed directly into the electricity grid in Europe.

Nucleosynthesis: the proton–proton cycle

Protons are hygrogen-1 nuclei $^{1}_{1}p$ or $^{1}_{1}H$. Other hydrogen isotopes are deuterium, $^{2}_{1}H$ or D, and tritium, $^{3}_{1}H$ or T. How do these form?

It is very unlikely that two protons could stick together without the mediation of a neutron. Two protons can, however, form deuterium by fusion, with the release of a positron and a neutrino. The reaction is shown in Figure 4.17:

$$^{1}_{1}H + ^{1}_{1}H \rightarrow ^{2}_{1}H + ^{0}_{+1}e + \nu$$

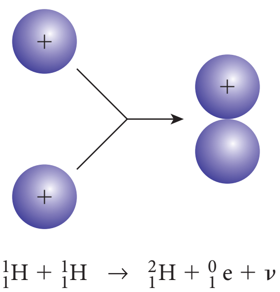


Figure 4.17 ▲ Fusion of two protons. Positron decay ensures a neutron is part of the new nuclide.

In a proton rich star, further fusion events can occur with protons. A neutron is available from within the deuterium nucleus. The resultant nuclide is helium-3, because it has two protons and one neutron.

$$^{2}_{1}H + ^{1}_{1}H \rightarrow ^{3}_{2}He + \gamma$$

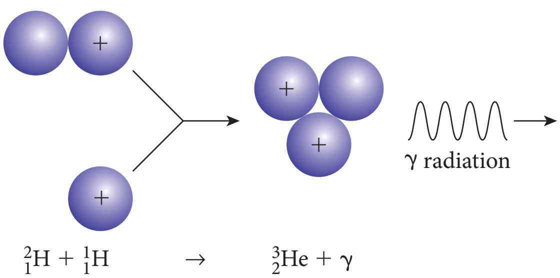


Figure 4.18 ▶ Fusion of deuterium and a proton

Helium-3 can fuse with another proton to produce lithium-4, but it is more likely – although the probability is very low – of two helium-3 nuclides colliding. The fusion products are helium-4 (alpha particle) and two protons:

$$^{3}_{2}He + ^{3}_{2}He \rightarrow ^{4}_{2}He + 2^{1}_{1}H + \gamma$$

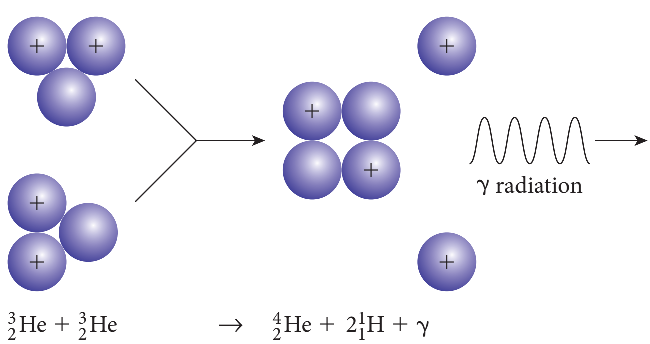


Figure 4.19 ▶ Fusion of two helium-3 nuclides

This set of fusion reactions starts and finishes with protons. It is called the hydrogen cycle or the proton–proton cycle.

Fusion occurs when two light nuclei join to produce a nucleus with a lower binding energy per nucleon.

WORKED EXAMPLE 4.3

Consider the data in Table 4.8.

Table 4.8 Mass of particles, in unified mass units, involved in the hydrogen cycle

Particle	Proton	Neutron	Deuterium	Helium-3	Helium-4	Positron
Mass (u)	1.0078	1.0086	2.0141	3.0160	4.0026	0.0005

1 u is equivalent to 1.661×10^{-27} kg

Speed of light = $3.00 \times 10^{8}\ \text{m s}^{-1}$

Consider the fusion reaction $^{1}_{1}\text{H} + {}^{1}_{1}\text{H} \rightarrow {}^{2}_{1}\text{H} + {}^{0}_{+1}\text{e} + \upsilon$.

a What is the mass defect in:

i unified mass units? (1 mark)

ii kilogram? (1 mark)

b How much energy is released in this reaction? (2 marks)

Answers	Logic	
a i $2 \times 1.0078\,\text{u} = 2.0141\,\text{u} + 0.0005\,\text{u} + \Delta m$	Use the correct equation.	1 mark
$\Rightarrow \Delta m = 0.001\,\text{u}$	Substitute the correct values.	1 mark
ii $\Delta m = 0.001\,\text{u} \times 1.667 \times 10^{-27}\,\text{kg u}^{-1}$		
$\Rightarrow \Delta m = 1.667 \times 10^{-30}\,\text{kg}$		
b $\Delta E = (\Delta m)c^2$	Use the correct equation.	1 mark
$\Rightarrow \Delta E = 1.667 \times 10^{-30}\,\text{kg} \times (3.00 \times 10^{8}\,\text{m s}^{-1})^2$	Substitute the correct values.	1 mark
$\Rightarrow \Delta E = 1.5003 \times 10^{-13}\,\text{J}$		

Try these yourself

1 Consider the fusion reaction $^{2}_{1}\text{H} + {}^{1}_{1}\text{H} \rightarrow {}^{3}_{2}\text{He} + \gamma$.

a What is the mass defect in:

i unified mass units? (1 mark)

ii kilogram? (1 mark)

b How much energy is released in this reaction? (2 marks)

2 Consider the fusion reaction $^{3}_{2}\text{He} + {}^{3}_{2}\text{He} \rightarrow {}^{4}_{2}\text{He} + 2\,{}^{1}_{1}\text{H} + \gamma$.

a What is the mass defect in:

i unified mass units? (1 mark)

ii kilogram? (1 mark)

b How much energy is released in this reaction? (2 marks)

Mass defect in fusion

For **light nuclides** the energy per nucleon is greater when there are more particles in their nucleus. The larger nucleus is more stable.

Other fusion reactions can and do occur. For example, the production of nitrogen-14 from carbon-12 proceeds by way of two proton additions and positron emission:

$$^{12}_{6}C + ^{1}_{1}H \rightarrow ^{13}_{7}N + \gamma$$

$$^{13}_{7}N \rightarrow ^{13}_{6}C + ^{0}_{+1}e + v$$

$$^{13}_{6}C + ^{1}_{1}H \rightarrow ^{14}_{7}N$$

Fusion enables the synthesis of many other nuclides.

JOINT EUROPEAN TORUS

Watch as fusion of 5 mg of deuterium takes place in the 'magnetic bottle' at the Culham Centre for Fusion Energy facility in England.

Controlling fusion for energy production

Fusion for energy production and distribution is still in the development phase.

Fusion reactions occur at very high temperatures, about 100 million degrees, and at high density. At high temperature, small nuclides have enough energy to crash through the electrostatic repulsion and get close enough to use the strong nuclear force to stick together. At high enough density, it is possible to ensure enough fusion events to produce about 1 GW of energy. This is comparable to a fission reactor.

At high temperatures the electrons are stripped from the atoms, leaving the nuclei exposed. This creates a neutral plasma that is exceedingly difficult to contain. Containment is achieved magnetically in a toroid (doughnut) shaped vessel.

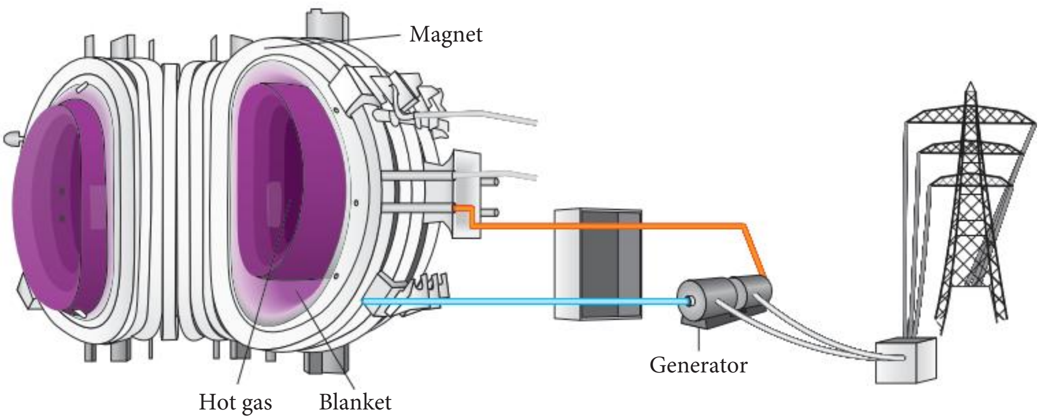


Figure 4.20 ▶ Fusion reactions produce heat in the toroidal plasma 'bottle'. The heat exchanger takes heat to produce steam, which is used to generate electrical energy.

The most promising fusion reaction is the deuterium plus tritium reaction, or D–T reaction:

$^{2}_{1}H$	+	$^{3}_{1}H$	$\rightarrow$	$^{4}_{2}He$	+	$^{1}_{0}n$
2.0141 u		3.0160 u		4.0026 u		1.0086 u

In nuclear fission reactions, the majority of the energy released is carried with the fission fragments. These are able to transfer their kinetic energies directly to the heat exchanger. Unfortunately, in the D–T fusion reaction, about 80% of the energy is carried by the neutrons. However, the neutral charge of neutrons means they tend to penetrate to the nuclei of the atoms from which the heat exchanger is made. This causes unwanted nuclear reactions. The heat exchanger does not get the energy produced by fusion directly. It is therefore necessary to add an extra process to capture the neutrons, and cause an appropriate nuclear reaction. This is done with liquid lithium. The nuclides produced in this process then transfer their kinetic energy to the heat exchanger.

Typically, the neutrons from the D–T fusion reaction encounter a liquid lithium 'bath' that surrounds the containment chamber. Lithium has two main isotopes, lithium-6 and lithium-7. Their reactions with neutrons produce alpha particles and tritium:

$$^{1}_{0}n + ^{6}_{3}Li \rightarrow ^{4}_{2}He + ^{3}_{1}H$$

$$^{1}_{0}n + ^{7}_{3}Li \rightarrow ^{4}_{2}He + ^{3}_{1}H + ^{1}_{0}n$$

It is these nuclei that then transfer their kinetic energy to the heat exchanger.

QUESTION SET 4.4

Remembering

1 Define 'fusion'.

2 Draw a diagram to show the fusion of two protons.

3 Describe the main features of a fusion reactor which could be used to produce electrical energy.

Understanding

4 Why is fusion favoured over fission for light elements?

5 How likely is it for fusion to occur in stars?

6 Why is a lithium 'bath' needed in a fusion reactor designed to produce electrical energy?

Applying

7 During a fusion reaction in a star, lithium-6 takes in a neutron. A new composite nucleus is formed that decays by alpha emission. Write the nuclear equations for the:

a production of the composite nucleus.

b subsequent alpha decay.

8 How is the kinetic energy from the D–T reaction transferred to the heat exchanger? Use nuclear reactions to support your explanation.

Analysing

9 For the fusion reaction $^{12}_{6}C + ^{1}_{1}H \rightarrow ^{13}_{7}N + \gamma$ what is:

a the mass defect?

b the energy released?

Particle	Proton	Carbon-12	Nitrogen-13
Mass (u)	1.0078	12.0000	13.0057

Reflecting

10 Why do you think fusion is 'cleaner' than fission?

Effect of radiation on humans

Neutrons, X-rays, alpha particles, beta particles and gamma rays are ionising radiations which interact with electrons, forming either ions or free radicals. Free radicals are atoms or molecules with unpaired valence electrons. They are highly reactive. In living matter, these ions and free radicals can start unwanted chemical reactions in cells, damaging or even killing living cells. Cell division may also be affected. This can lead to genetic mutation and uncontrolled cell division causing tumours.

The effect of radiation on the human body depends on the quantity of radiation applied to living tissue (the dose) and its absorption in that tissue (absorbed dose).

The **dose** is the amount of energy, E, carried from the source to a body of mass, m. The **absorbed dose**, D, is the energy arriving at the body per unit of mass:

$$D = \frac{E}{m} \text{ (Unit: joule per kilogram or gray, Gy)}$$

Different types of ionising radiation cause different effects in living tissue. Radiation damage depends not only on the radiation dose but also on the type of radiation and the nature of organs, tissues and cells exposed. Bone marrow, the digestive system, the nervous system, the reproductive organs and the eyes are all seriously affected by ionising radiation.

The **equivalent dose**, H, is a measure of the biological effects of different radiations. Each type of ionising radiation is given a **radiation weighting factor**, W_R. Beta particles and gamma rays are similar in their biological effects, but slow neutrons and alpha particles are respectively 3 times and 20 times more damaging.

Table 4.9 Radiation weighting factors for different ionising radiations

Ionising radiation	Radiation weighting factor (no unit)
Beta particles	1
Gamma rays	1
Slow (thermal) neutrons	3
Alpha particles	20

The equivalent dose is the product of both absorbed dose and the radiation weighting factor:

$$H = D \times W_R \text{ (Unit: joule per kilogram; sievert, Sv)}$$

Both gray and sievert refer to the energy per mass, but each refers to a different thing: absorbed dose (Gy) and equivalent dose (Sv).

Dose, E: energy arriving at a body

Absorbed dose, D: $D = \frac{E}{m}$ (joule per kilogram; Gy)

Radiation weighting factor, W_R: relative biological effect of different ionising radiations

Equivalent dose, H: $H = D \times W_R$ (joule per kilogram; Sv)

Biological effects of radiation

The effects of radiation exposure can be classified as somatic (bodily effects; for example, causing cancers, tumours) or genetic (for example, mutations of genes in reproductive cells, which can result in birth deformities).

Table 4.10 Some effects of acute ionising radiation on the human body

Blood system	Gastrointestinal tract	Central nervous system	Other related effects
• Lowered bone marrow activity • Anaemia	• Severe nausea • Vomiting • Diarrhoea	• Confusion and nervousness • Loss of vision • Seizures • Coma	• Severe weakness and tiredness • Fluctuating blood pressure • Lymphatic system failure • Fever • Shock • Death (>2 Sv)

Signs of radiation sickness include nausea, vomiting and diarrhoea (NVD) and headaches. Delayed effects of radiation doses, such as cancerous growths and leukaemia, may not appear for many years after the initial irradiation. An equivalent dose above about 2 Sv would give you only about a 50% chance of survival. Equivalent doses above 2 Sv kill too many cells within the bone marrow and intestines for you to live or live normally.

The rate at which a person receives the ionising radiation is also significant. Acute irradiation occurs over seconds, minutes and hours. Chronic irradiation occurs over days, weeks and months. Table 4.11 shows the effects of acute and chronic whole body irradiation for different equivalent doses. Notice the unit is millisievert, mSv, which is used more often than sievert, Sv.

Table 4.11 Effects of a range of whole-body dose equivalents

Dose equivalent (mSv)	Period of delivery of dose equivalent	Effect
15 000	Acute and chronic	Fatal effects in all cases
10 000	Acute and chronic	Likely immediate fatal illness
2000	Acute and chronic	Threshold for immediate, fatal radiation sickness
1000	Chronic Acute	Increased likelihood of later development of cancer Threshold for immediate, non-fatal radiation sickness
100	Acute and chronic	Increasing likelihood of cancer
50	Acute and chronic	Maximum annual dose for any worker Lowest dose at which cancers may be caused in adults
20	Chronic	Average limit per year for workers in the nuclear industry, uranium and mineral sands miners, and medical workers

Effects can be immediate and continuous; however, for lower doses, there is usually an acute phase, then a period of relative calm for days or even months. Then, more severe symptoms return. There is an increased risk of survivors developing problems, such as cancer, at a later stage. Radiation sickness also lowers the immune system. With aggressive treatment, such as antibiotics, people can survive a 10 Sv dose equivalent, but >15 Sv is definitely fatal.

Radiation doses below about 10 mSv have very little effect on humans. Radiation doses below about 1 Sv would not kill you immediately. Your body may be able to overcome damage caused to living cells by such radiation. However, there may be an increased possibility of getting cancer some years after receiving the dose. Doses of about 1 Sv can cause radiation sickness within a few days. A large acute dose, in excess of 10 Sv, would give no chance of survival.

A typical annual dose due to natural background radiation is about 3 mSv. This varies greatly from place to place.

For many applications, the passage of 5 half-lives brings the radiation within acceptable long-term limits. For humans, however, a rule of thumb is that the elapse of 10 half-lives effectively means no radiation is present.

WOW

Misuse of radiation for murder

In November 2006, ex-Russian spy Alexander Litvinenko died in a London hospital from symptoms of radiation poisoning. He was most likely poisoned three weeks earlier. Litvinenko was part of a Russian group in London that included critics of the Russian Presidency. During the investigation, police experts detected a large dose of alpha radiation in his urine, and polonium-210 in his body. They found traces of polonium-210 in his home, a sushi bar where he ate shortly before falling ill and a hotel where he had met two Russians. Litvinenko's drink appeared to have been spiked with a few grains of white powder. Within minutes, the ionising radiation began killing gastrointestinal cells. The energy also raised his internal temperature significantly. Polonium-210 has a half-life of 138 days and a biological half-life of 30–50 days. In 2013, the British Government admitted that the original inquest into Litvinenko's death was affected by the need to maintain good 'international relations'.

QUESTION SET 4.5

Remembering

1 Define the following terms.
   a Dose
   b Absorbed dose
   c Radiation weighting factor
   d Equivalent dose

2 Write the equation for equivalent dose.

3 Name the two categories of effects of exposure to ionising radiation. Give a specific example for each.

Understanding

4 In what ways does ionising radiation affect the:
   a blood system?
   b gastrointestinal tract?

5 Exposure to ionising radiation may be acute or chronic. Define both categories and give a specific example for each.

Applying

6 Twenty joules of ionising radiation is absorbed in a 100 g cancerous tumour. What is the absorbed dose?

7 A 90 kg person receives an equivalent dose of 1000 mSv. If the radiation were slow neutrons, what was the original dose?

Analysing

8 The effective half-life, $t_{\frac{1}{2}}$(effective), of radioactive materials in a human can be estimated by:

$$t_{\frac{1}{2}}(\text{effective}) = \frac{t_{\frac{1}{2}}(\text{physical}) \times t_{\frac{1}{2}}(\text{biological})}{t_{\frac{1}{2}}(\text{physical}) + t_{\frac{1}{2}}(\text{biological})}$$

   a Estimate the effective half-life for polonium-210.
   b What does this figure mean for a person irradiated with polonium-210?

Reflecting

9 Identify and describe four new things you have learnt about the effects on humans of ionising radiation.

CHAPTER SUMMARY

- The strong nuclear force is required to hold nucleons together, especially to overcome the electrostatic force of repulsion between protons.
- Nuclei are stable when the strong nuclear force and the electrostatic force are well-balanced. Neutrons are important in maintaining this balance.
- The stability curve shows stable nuclides. It can be used to predict the type of radioactivity most likely from a non-stable nuclide.
- A loss of mass occurs when nucleons combine to form a nucleus. This is the mass defect, Δm, expressed in Einstein's mass–energy equation: $\Delta E = (\Delta m)c^2$ where c = the speed of light, $3.00 \times 10^8\ \mathrm{m\,s^{-1}}$.
- Nuclear binding energy is the energy needed to separate a nucleus into its component nucleons.
- Fusion is the process by which nucleons join to form a new nucleus. Nucleosynthesis is the set of fusion reactions that lead from nucleons to a variety of nuclides. This occurs for light elements ($Z < 56$). Energy is released.
- Fission is the process by which heavy nuclei separate into fragments, with the release of energy. The probability of equal fragments from a fission event is quite low – other possibilities are more probable.
- Fission chain reaction occurs when the neutrons released from a fission event go on to produce more fission events.
- Controlled fission reactions are used in nuclear reactors. This requires the use of a moderator and control rods.
- Uncontrolled fission reactions are used in weapons.
- The use of fusion and fission raises ethical and environmental questions.
- The management of radioactive nuclear waste from fission reactors depends on whether it is high-, medium- or low-level waste.
- Ionising radiation causes somatic and genetic effects; which occurs depends on the dose and its absorption in living tissue. Different radiations affect tissue differently.
- The dose is the amount of energy, E, received. The absorbed dose, D, is the energy absorbed by the body per unit of mass:

  $D = \frac{E}{m}$ (Unit: Gy)
- The equivalent dose, H, takes account of the absorbed dose and the radiation weighting factor, W_R.

  $H = D \times W_R$ (Unit: Sv)

  For beta particles and gamma rays, $W_R = 1$; for slow neutrons, $W_R = 3$; and for alpha particles, $W_R = 20$.
- Ionising radiation can have acute and chronic effects on humans. They are most obvious in the blood system, gastrointestinal tract and central nervous system.
- The limit of equivalent dose for workers in radiation industries is 20 mSv per year. Death is the most likely outcome above about 10 Sv, whether received acutely or chronically.

CHAPTER GLOSSARY

absorbed dose energy per mass absorbed in a body (Gy)

Big Bang theory theory about the history of the physical universe

binding energy energy needed to disassemble a nucleus into its component nucleons; measure of stability of a nuclide

control rod rod made from a neutron poison, used to absorb excess neutrons in a nuclear reactor

controlled chain reaction a chain of nuclear reactions that are controlled to limit the rate at which reactions occur. In steady state (reaction rate held constant), an average of one neutron from each reaction goes on to cause another reaction. This is the case for a nuclear power reactor running at constant power output

dose energy received from a radiation source (J)

Einstein's mass–energy equation $\Delta E = (\Delta m)c^2$

electron-volt, eV small energy unit; 1.602×10^{-19} J

enrichment process of separating out U-235 from a sample and adding it to another sample, increasing the proportion of U-235 in natural uranium

equivalent dose combination of absorbed dose and relative effect in biological systems of particular ionising radiations

fast breeder reactor nuclear reactor that uses neutrons with high energies (fast neutrons) to cause fission events

fast neutron neutron with kinetic energy of 100 keV or more

fissile able to undergo fission; U-235, U-238, U-233, Pu-239

fission the splitting of a heavy nucleus ($Z > 56$) into fragments with lower atomic numbers and neutrons; energy stored in the nucleus becomes available

fission fragment nucleus produced as a result of fission; fission product

fission product nucleus produced in a fission event; fission fragment

fusion the coming together of two nuclei to form a new nucleus with greater atomic number ($Z < 56$)

light nuclide lightweight nuclide with small atomic mass (small nucleon number)

mass defect mass difference between the constituents of a nucleus and the mass of the nucleus; measure of the energy needed to hold a nucleus together

moderator light atoms in a nuclear reactor that slow down fast neutrons to thermal speeds in order to increase the likelihood of further fission events

negative energy problem a problem whereby the energy output from a fusion reactor is less than the energy input required to produce fusion

neutron poisons nuclei including fission fragments that absorb a neutron, thus making the neutron unable to cause further fission events

nuclear binding energy total energy needed to hold a nucleus together; the greater the binding energy, the greater the stability

nucleosynthesis period in the history of the physical universe when the majority of nuclides formed

radiation weighting factor relative effect of types of ionising radiation on biological tissue

slow neutron neutron with kinetic energy around 0.1–20 keV; thermal neutron

stability curve plot of nuclides showing stable isotopes

strong nuclear force force that overcomes the electrostatic repulsion force at nuclear distances (approx. 15 fm)

thermal neutron neutron with kinetic energy around 0.1–20 keV; slow neutron

thermal reactor nuclear reactor that uses neutrons with thermal energies (slow neutrons) to cause fission events

uncontrolled chain reaction a chain of nuclear reactions that is not controlled. Usually this means a reaction rate that increases rapidly. For this to occur the average number of neutrons from each reaction that go on to cause more reactions is greater than one

CHAPTER REVIEW QUESTIONS

Remembering

1 Define 'nuclear binding energy per nucleon'. How does this quantity affect nuclear stability?

2 Describe the following processes.
   a Fission
   b Fusion

3 Use a diagram to show the meaning of 'chain reaction' as applied to a nuclear reactor.

4 Define these terms.
   a Thermal nuclear reactor
   b Neutron poison

5 Write the equation for equivalent dose. Define all terms.

Understanding

6 On a sketch of the stability curve show the region where:
   a the proton and neutron numbers are about equal.
   b beta-minus emission is most likely.

7 What is mass defect when applied to the nuclear binding energy per nucleon in a fission event?

8 Why is fission favoured over fusion for heavy nuclides?

9 In a fission reactor what is the purpose of the:
   a moderator?
   b control rods?

10 Describe the symptoms of ionising radiation on the gastrointestinal tract.

Applying

11 Explain why xenon-140 is likely to be involved in fission, but carbon-12 in fusion.

12 Graphite is mainly carbon-12. Carbon-13 is a beta-minus emitter. Explain how graphite could be used in control rods in a nuclear pile. Support your answer with nuclear reaction equations.

13 Why does a nuclear fission reactor vessel have to be carefully shaped?

14 Calculate the mass defect in the D–T fusion reaction.

15 Twenty-five joules of ionising radiation is absorbed in a 250 g cancerous tumour. What is the absorbed dose?

16 A 65 kg person receives an equivalent dose of 1.2 Sv. If the radiation were slow neutrons, what was the original dose?

Analysing

17 Use the stability curve to decide:
   a which of tantalum-186 and tantalum-172 is a positron emitter, and which a beta-minus emitter.
   b whether thorium-224 is likely to be a beta or alpha emitter.

18 A thermal neutron, mass 1.0086 u, causes fission of U-235 (235.044 u). The fission fragments and their masses, in unified mass units, are I-131 (130.906 u) and Y-103 (102.945 u).
   a How many neutrons are released in this fission event?
   b What is the mass defect in this event in:
      i unified mass units
      ii kilogram?
   c How much energy is released? Give your answer in joule.

19 Both fusion and fission systems require heat exchangers to generate electricity.

a What is the purpose of the heat exchange system?

b How does the heat exchange process differ for a fusion system compared to a fission system?

20 A person is exposed to an equivalent dose of 1.5 Sv of ionising radiation. Describe the likely symptoms and likely outcome for the person.

Reflecting

21 Review what you have learned about stability, fission and fusion.

22 How has your understanding of nuclear safety improved? Give examples.

23 Will fission or fusion save the planet from the negative effects of climate change? Discuss.

CHAPTER 5
ELECTRICITY

Shutterstock.com/Ase

By the end of this chapter you will have covered the following material.

Science Understanding

- Electrical circuits enable electrical energy to be transferred efficiently over large distances and transformed into a range of other useful forms of energy including thermal and kinetic energy, and light (ACSPH037)
- Electric current is carried by discrete charge carriers; charge is conserved at all points in an electrical circuit (ACSPH038)
- Energy is conserved in the energy transfers and transformations that occur in an electrical circuit (ACSPH039)
- The energy available to charges moving in an electrical circuit is measured using electrical potential difference, which is defined as the change in potential energy per unit charge between two defined points in the circuit (ACSPH040)
- Energy is required to separate positive and negative charge carriers; charge separation produces an electrical potential difference that can be used to drive current in circuits (ACSPH041)
- Power is the rate at which energy is transformed by a circuit component; power enables quantitative analysis of energy transformations in the circuit (ACSPH042)
- Resistance for ohmic and non-ohmic components is defined as the ratio of potential difference across the component to the current in the component (ACSPH043)

Introduction

Electricity is a very convenient form of energy. It is available from many sources, such as batteries, alternators and solar panels. It can be generated by power stations that use coal, water (hydroelectricity), natural gas, wind or nuclear fuel. Electrical power is transmitted over large distances for domestic, commercial and industrial use.

Electrical energy is easily transformed into other types of energy, such as heat energy in toasters, ovens and heaters; sound energy in televisions and iPods; light energy in incandescent, fluorescent and LED lights; and mechanical energy in refrigerator motors, electric drills, hair dryers and electric shavers. It can also be used in other appliances to operate logic circuits in alarms, computers, robots and control devices.

In Chapter 1 you were introduced to the kinetic particle model. Heat is transferred by particles bumping into one another. These particles were atoms or molecules. In nuclear physics you learnt about subatomic particles – the particles inside atoms. We used the idea of delocalised electrons to help explain the heat conducting properties of metals. In our study of electricity we will look more deeply into the location and role of electrons, one type of subatomic particle.

Electricity is a form of energy. Energy can be transferred from place to place. It can be transformed from one form to another. Energy in an isolated system is conserved. These are important concepts that you have already come across and which will be developed further in this chapter.

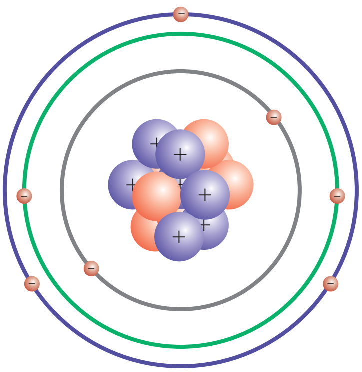


Figure 5.1 ▲ The Rutherford–Bohr model of an atom

Electrons

The kinetic particle model of matter, which we will use in this chapter, involves understanding the nuclear atom. For the Rutherford–Bohr model of an atom, which was discussed in Chapters 3 and 4, electrons circulate in the region of space around the nucleus in energy level shells (Figure 5.1). The nuclear atom includes a positively charged nucleus comprising positively charged protons and uncharged neutrons. Electrons that orbit the nucleus have a negative charge.

Metals are good conductors of heat and electricity. The arrangement of atoms in metals means that some electrons are relatively free to move. These delocalised or free electrons can move around within the metal. This means that they can conduct electricity.

ACTIVITY 5.1

POSITIVE AND NEGATIVE CHARGES

Aim

To investigate the effect of positive and negative charges on objects

You will need

- plastic straws
- paper serviettes
- glass rod
- wool or fur for charging the glass rod
- thread

▶

What to do

	What do you think will happen?	What did you observe?
Rub a straw with a paper serviette. Place it on different vertical surfaces. Is it different on a metal surface than on a non-metal surface?		
Rub a second straw with a serviette. Hold the two charged straws near each other.		
Cut a 3 cm section of straw and attach it to the thread. Hold a charged straw near it. Hold a charged glass rod near the straw.		

What did you discover?

How does what you found compare with what others in the class found?

Charge

You are probably familiar with some common examples of the effects of electrostatic charge: a plastic comb run through your hair, or a plastic ruler or pen rubbed on woollen material, attract small pieces of tissue paper. You sometimes hear a crackling noise and observe flashes of light if you take off a polyester or nylon top in the dark as the static electricity discharges. Static electricity is the build up of charge on an object.

If you rub a balloon on your hair and then hold it a few centimetres away, your hair will stick to the balloon. This is the caused by the interactions between the net charges on the balloon and on your hair.

Objects are made up of positive and negative charges. The overall sum of charges on a neutral object is zero – they have the same number of positive and negative charges. A positively charged object has more positive charges than negative charges. A negatively charged object has more negative charges than positive charges.

If two charged objects of the same net charge come near each other they push away from each other. Like charges repel. If two charged objects have opposite net charges, they come towards each other. Opposite charges attract.

Like charges repel. Opposite charges attract.

A neutral object has the same number of positive charges as negative charges.

Neutral objects do not attract or repel other neutral objects.

Positive charges are attracted to negative charges.

Let us look at this concept more closely with everyday examples. Blow up two balloons and hold them near each other. Do they attract or repel or stay still?

The balloons are both neutral because they have the same number of protons and electrons. Because they are both neutral, there is no attraction or repulsion (Figure 5.2a).

If you rub one of the balloons in your hair, electrons move from your hair onto the balloon. The balloon now has more electrons than protons which means it has an overall negative charge. You hair has lost some of its negative charge, so its overall charge is positive. The positive charge on the hair is attracted to the negative charge on the balloon (Figure 5.2b).

If you rub both of the balloons on your hair and hold them near each other what happens? Do they attract or repel or stay still?

The balloons are both negatively charged; they have more electrons than protons. Like charges repel, so the balloons move away from each other (Figure 5.2c).

Two objects that are both negatively charged repel.

A charged object is attracted to neutral objects.

What happens when you hold a charged object near a neutral object? Recharge one of the balloons on your hair and hold it near a neutral balloon or a scrap of paper. What happens?

The negatively charged balloon attracts the neutral balloon or scrap of paper. A charged object attracts neutral objects because the charges in the neutral object are attracted to the charged object and rearrange themselves. This rearrangement means that the local area nearest the charged object attracts the charged object to the neutral object (Figure 5.2d).

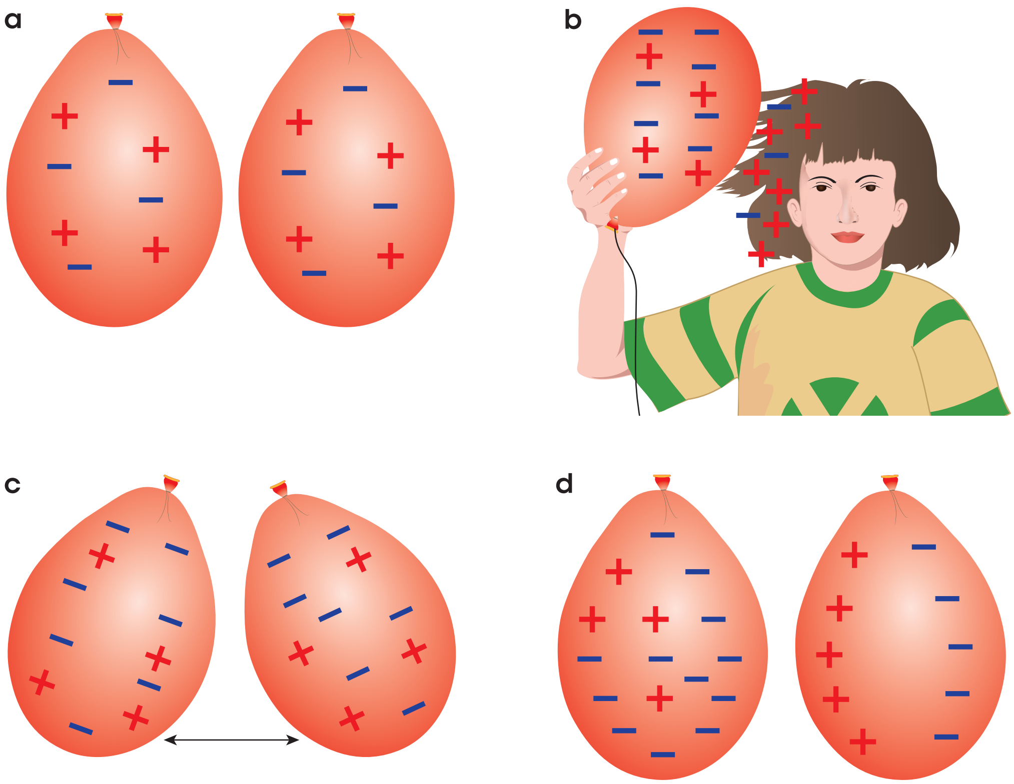


Figure 5.2 ▶ a) Two neutral balloons do not attract or repel. b) Electrons move from your hair onto the balloon. There are more electrons than protons so the balloon has an overall negative charge. c) Both balloons have an overall negative charge which causes them to repel. d) The large negative charge on the balloon causes a local re-arrangement of charges on a neutral balloon. The neutral object is attracted to the negatively charged balloon.

Measuring charge

If an object has the same number of electrons and protons then it will be neutral. There will be the same amount of positive charge as negative charge.

If an object has more electrons than protons, it will have an overall negative charge. If an object has more protons than electrons, it will have an overall positive charge.

The quantity of charge is represented by the symbol q. Charge (q) is measured in coulomb, C. One electron has a charge of 1.602×10^{-19} C. How many electrons equal one coulomb of charge? Let x equal the number of electrons.

$$1\text{ C} = x \times 1.602 \times 10^{-19}$$

$$x = \frac{1}{1.602 \times 10^{-19}}$$

$$x = 6.24 \times 10^{18}$$

To get one coulomb of charge you need 6.24×10^{18} electrons.

If one electron has a charge of 1.602×10^{-19} C, then is it possible for an object to have a charge of 0.8×10^{-19} C? You can't have half an electron, so it is impossible to have a charge of 0.8×10^{-19} C. Any amount of charge has to be a multiple of 1.602×10^{-19} C.

The amount of charge depends on the difference between the number of protons and the number of electrons.

Charge cannot be created or destroyed, but it can move from one object to another.

WOW

What is charge?

Charge is an intrinsic property of an object.

In everyday language, charge is used to mean a form of energy. We say we 'charge' our phone, but this is not scientifically accurate. When you charge your phone you are providing energy to the battery in order to separate charge. The battery separates the charge using chemical reactions. This separated charge is stored as electrical potential energy until it is needed.

Sometimes people think that charge is energy, but an electron isn't energy, just like a bouncy ball isn't energy. A bouncy ball is not energy, but it can be given energy by throwing it, or it can be made to store energy by lifting it up off the ground ready to be dropped. The charge itself is not energy, but the charged object can be given energy or it can store energy.

Insulators, conductors and semiconductors

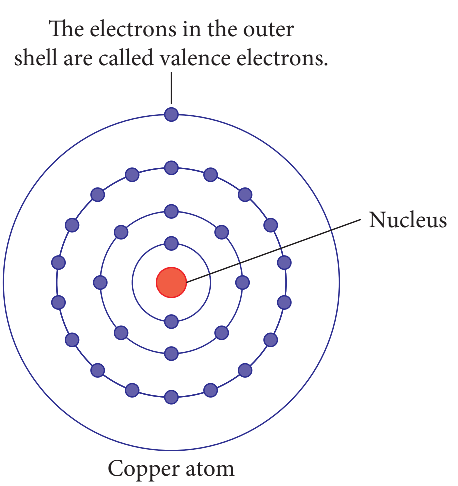
A material that allows current to flow through it easily is a **conductor**. Conductors are materials that have a large number of free or conduction electrons. These free electrons move from one area to another when there is a potential difference (see page 148). This flow of free electrons makes up the current. Metals are an example of a good conductor because they have many free electrons.

An **insulator** is a material that does not allow current to flow through it. Insulators do not have free electrons.

A **semiconductor** is a material with a very small number of free electrons at room temperature. A current can flow through a semiconductor, but not easily.

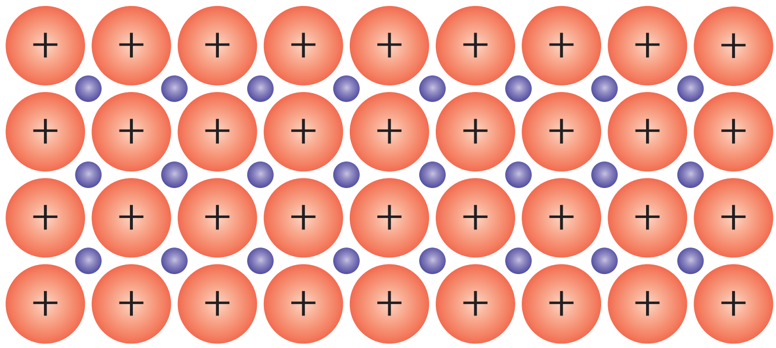
Whether a material is a conductor, an insulator or a semiconductor depends on what sort of atoms it is made of, and how those atoms are connected to each other.

In a metal, when the atoms bond together they do so in such a way that the outermost electrons are shared between a large number of atoms. Consider the single copper atom shown in Figure 5.3. There are many electrons surrounding the nucleus. When the copper atom is isolated, all the electrons are bound to the nucleus by the electrostatic force. However, when we bring a large number of copper atoms together, the nuclei and most of the electrons form a lattice of metal ions; that is, a **metal lattice**. A small number of outer-shell (valence) electrons, one in copper (one, two or three in other metals), become un-bound or free of their original nuclei. These are the free electrons that are involved in conduction when a current flows. Hence, they are also referred to as conduction electrons. These free electrons also have an important role in holding the metal together; they act as a negative 'sea' or 'glue' of electrons surrounding the positive ions, which consist of the nuclei and their bound electrons. This is shown in Figure 5.4. This is what chemists mean when they talk about metallic bonding – it is a bond between many atoms due to these free electrons.



▲ Figure 5.3
In a copper atom, inner-shell electrons shield outer-shell electrons from the full effect of the 29 positive charges (protons) in the nucleus.

Figure 5.4 ▶ The nuclei form a regular lattice with the delocalised electrons free to move about.



In insulators, the atoms are held together by bonds in which the bonding electrons are tightly held to individual atoms or a small number of atoms, most usually a pair. These electrons are therefore not free to move about the material. Chemists refer to these bonds as ionic when the bonding electron is held tightly to a single atom, and covalent when the bonding electron is held tightly between a pair or small number of atoms.

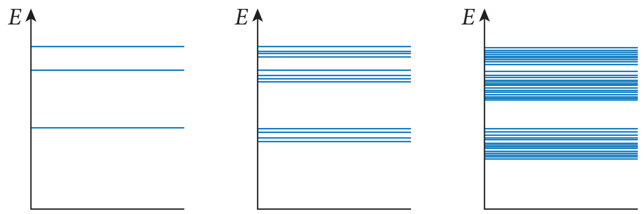


Figure 5.5 ▲ Energy levels of a) one atom and b) 4 bound atoms. c) The energy levels of many bound atoms form bands.

Semiconductors are also held together by covalent bonds in which the bonding electrons are shared between two or a few atoms; however, they are not held as tightly as in an insulator.

We can also understand conductivity in terms of energy. Recall from Chapter 3 that we can represent an atom either by drawing a picture like that shown in Figure 5.3, or by showing the energy levels of the electrons as in Figure 3.8(b). As we bring individual atoms together, the energy levels of the electrons change slightly. For a large number of atoms, each energy level for an individual atom splits into a large number of closely spaced energy levels. These groups of closely spaced levels are called bands. This is shown in Figure 5.5.

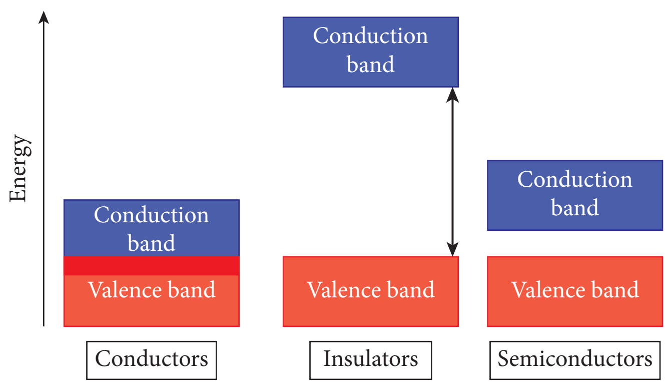
Free electrons occupy bands with relatively high energy. In a metal, there is no significant energy gap between low energy levels (called the **valence band**) and the **conduction band**. The energy levels are effectively continuous.

In an insulator, there is a large energy gap between the valence band and the conduction band, and all the electrons are in the valence band. It takes a large amount of energy to move an electron into the conduction band in an insulator. Typically, the material will melt or be physically damaged before it becomes a conductor.

In a semiconductor, the gap between the highest energy levels in the valence band and the lowest energy levels in the conduction band is small. Hence it is easy to move a few electrons into the conduction band, to allow a current to flow. The energy required for this typically comes from the internal energy of the material, which is related to its temperature. Hence semiconductors generally become better conductors as they get warmer.

The energy band structure of conductors, insulators and semiconductors is shown in Figure 5.6.

Figure 5.6 ▶ Energy band structure for conductors, insulators and semiconductors



Increasing temperature affects metals and semiconductors quite differently. As we have seen, increasing the temperature of a semiconductor generally increases its conductivity, at least up to a point. This is because the increased energy in the system results in a larger number of conduction electrons. A metal already has a very large number of conduction electrons at low temperature. When the temperature of a metal is increased this number does not change significantly. What does happen is that all the atoms 'jiggle about' more, and the free electrons move faster. The result is more collisions. (Imagine trying to cross a room full of running and bouncing children compared with one in which they are sitting still at their desks.) Hence increasing the temperature of a metal decreases its conductivity. At high temperatures this effect is also significant for semiconductors. Hence the conductivity of a semiconductor is low at both very low and very high temperatures, and has a maximum in between.

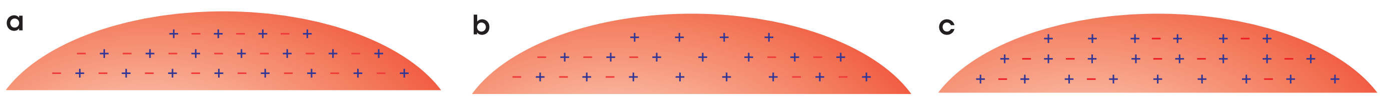
Table 5.1 Examples of conductors, insulators and semiconductors

Good conductors	Copper	Aluminium	Graphite
Poor conductors	Water	Human body	Sugar
Insulators	Glass	Rubber	Dry air
Semiconductors	Silicon	Germanium	Gallium

Adding and removing charge from conductors and insulators

If electrons are added to an object, the electrons repel each other and, given the right conditions, can spread out. If that object is an insulator, the electrons cannot spread out on that surface and the charge remains in one place. If the object is a conductor, the electrons can easily spread out on the surface to produce an even distribution.

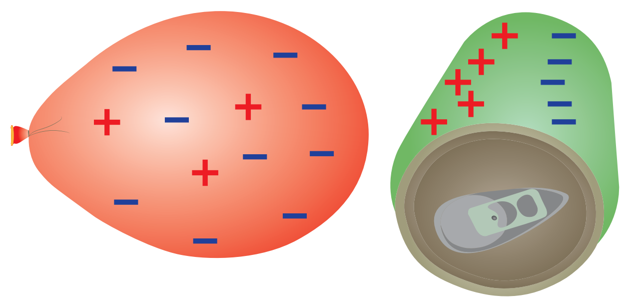
If electrons are removed from a conductor, then there are more positive charges in one area. The protons cannot move, but they attract electrons from nearby atoms, which can also distribute the charge (Figure 5.7).



▲ **Figure 5.7**
a) A conductor has an even spread of electrons and protons with no net charge. b) Some electrons are removed from the surface of the conductor. There is an excess of protons in one spot. c) The electrons are free to move and are attracted to the positive area. They spread out over the surface to distribute the charge evenly. There is an overall positive charge.

Let us do another experiment with a charged balloon. Lay an empty soft drink can on its side on your desk. Hold a charged balloon close to one side of the can. What do you think will happen? What actually happens? Can you make the can roll? Can you explain why this happens?

The can is neutral; it has the same number of protons and electrons. When the negatively charged balloon is held near the can, the electrons are repelled and move to the other side of the can. This means that the side closest to the balloon has a positive charge and is attracted to the balloon (Figure 5.8).



▲ **Figure 5.8**
The charges in the soft drink can rearrange themselves because of the negative charge of the balloon.

QUESTION SET 5.1

Remembering

1 Name the two types of charge and the subatomic particles associated with each.

2 Which of the following amounts of charge are possible? Why?

A 1.2×10^{-19} C

B 2.4×10^{-19} C

C 4.0×10^{-19} C

D 4.8×10^{-19} C

Understanding

3 Which is easier to remove from an atom, the protons or the electrons? Why?

4 Kristy and Adam are discussing an object with neutral charge. Adam says that a neutral object doesn't have any charge on it. Kristy says that a neutral object has lots of positive and negative charges on it, there are just equal numbers of each. Who is correct and why?

5 Electrons are removed from an area of a neutral conductor. What is most likely to happen?

A Protons in the conductor move towards the area.

B Electrons in the conductor move towards the area.

C Nothing moves.

6 Electrons are added to an area of an insulator. What is most likely to happen?

A Protons in the insulator move towards the area.

B Electrons in the insulator move away from the area.

C Nothing moves.

Applying

7 A positively charged conductor is moved near a neutral conductor.

a Do the two conductors attract or repel? Use a diagram to explain.

b The positively charged conductor touches the neutral conductor. What happens to the charges?

8 How many electrons make 0.5 coulomb of charge?

Reflecting

9 How has your understanding of the following changed?

a The interaction of positive and negative charges

b Electrons in a metal

c Differences between insulators, conductors and semiconductors

Energy in circuits

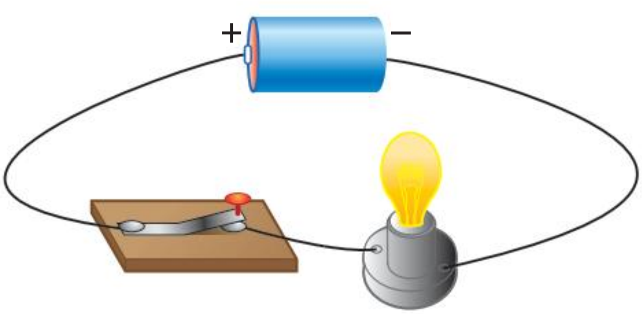


Figure 5.9 ▲ A light globe, switch and a battery connected by wires

Electrical circuits involve energy and the movement of charge. When charges move around the circuit they can lose or gain potential energy.

We have explored what happens when there is an excess of charge on a conductor – the electrons are repelled from each other and move so that the charge is distributed more evenly. We can now apply this concept to see how a battery provides energy to an **electrical circuit**. An electrical circuit is a complete loop through which charges can flow. If the loop is not complete then charge cannot flow.

Figure 5.9 shows a light globe, switch and a battery connected by wires. This is an example of a basic electrical circuit.

The battery is an example of an energy source; it provides energy to the circuit. The energy causes charge to flow through the circuit.

The light globe is an example of a load or sink and uses the energy provided by the source.

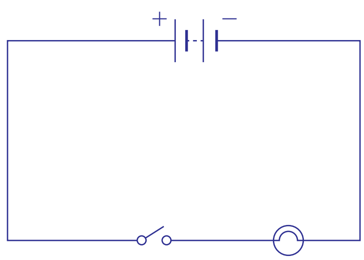
Physicists use symbols to represent different components in circuit diagrams. Some examples of common symbols are shown in Figure 5.10. The most important ones to focus on for now are the battery, filament lamp and resistor. Later we will look at the other electronic components.

If you used symbols to draw the circuit in Figure 5.9, it would look like Figure 5.11.

Device	Symbol	Device	Symbol
Wires crossed, not joined		Earth or ground	
Wires joined; junction of conductor		Switch	
Fixed resistor		Diode	
Variable resistor		Photodiode	
Light-dependent resistor		LED	
Rheostat or resistor with moving contact		AC supply	
Thermistor		Voltmeter	
Filament lamp		Galvanometer	
Battery of cells		Ammeter	
Alternative for battery		Signal lamp or indicator	
Cell			

◀ **Figure 5.10** Conventional symbols used in electrical circuits

See Appendix 5 for a complete list of circuit symbols.



▲ **Figure 5.11** A circuit drawn with symbols

Separation of charges

Energy is required to separate opposite charges. If a positive and negative charge are close together they are attracted strongly. If energy is provided then the charges can be pulled apart. Now the charges have energy ready to be released which is called **potential energy**. This is energy ready to be transferred or transformed. If let go, the positive and negative charges move towards each other. As they move towards each other they lose potential energy, and it takes work to separate them again. This is similar to the way potential energy increases when particles in a liquid move apart, eventually forming a gas (see Chapter 1, page 7).

Two positive charges will repel each other more when they are close together than when they are further apart. They have potential energy that enables them to move away from each other. As they move away from each other their potential energy decreases. It takes work to move them back towards each other. This work is stored as potential energy.

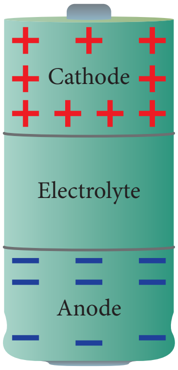


Figure 5.12 ▲
A simplified battery

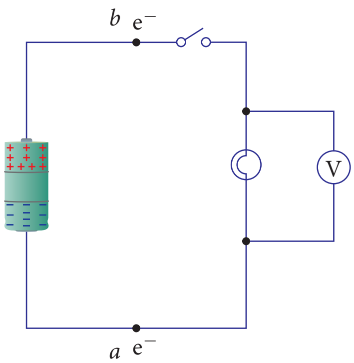


Figure 5.13 ▲
Electron *a* near the negative terminal has a high potential energy. Electron *b* near the positive terminal has a low potential energy. The voltmeter is measuring the potential energy difference between electron *b* and electron *a*.

Work is done on unlike charges to separate them. Potential energy increases the further apart they are. Unlike charges lose potential energy as they approach each other.

Work is done on like charges to bring them together. Potential energy increases the closer together they are. Like charges lose potential energy as they separate.

Batteries

A battery is a source of potential energy per charge, or ***emf*** (electromotive force). In a charged battery, a chemical reaction separates the positive and negative charges. There are lots of like charges close together in separate parts of the battery. This gives the charges a large amount of potential energy. In a circuit, the negative terminal of the battery is connected to the positive terminal by wires. The build-up of electrons in the negative terminal causes the electrons in the wire to move towards the positive terminal. The chemical reaction keeps separating charges as current flows through the circuit.

Figure 5.13 shows a simple circuit with a battery and a light globe. Electrons near the negative terminal (e.g. electron *a*) have a high potential energy – they are near other electrons. As the electrons move through the circuit their potential energy is transformed into light and heat in the light globe. Electrons near the positive terminal (e.g. electron *b*) have a low potential energy – they are near many protons.

The same logic can be applied to protons. Protons near the positive terminal have a high potential energy – they are near other protons. Protons near the negative terminal have a low potential energy because they are near electrons. This will be important later when we discuss conventional current.

Potential difference

Electric potential describes how much potential energy there is for each charge at different locations in a circuit. Electric potential is the electric potential energy per unit charge at a particular point. We write this as:

$$V = \frac{W}{q}$$

Electric potential is measured in units of joules per coulomb, J C^{-1}, which is given the name volt, V.

We measure the potential at a point relative to some reference point, often the ground. Hence what we generally measure is not potential but **potential difference**. We give this the same symbol, *V*. The potential difference between two points is simply the difference in potential energy per unit charge between those two points. It is sometimes referred to as 'voltage' because it is measured in volts, V.

When a charge q moves between two points it will lose or gain potential energy. The energy change, loss or gain is equal to:

$$W = qV$$

Potential difference is measured with a voltmeter. A voltmeter measures the difference in potential between two points. The voltmeter needs to be connected in parallel to two different parts of the circuit as shown in Figure 5.14.

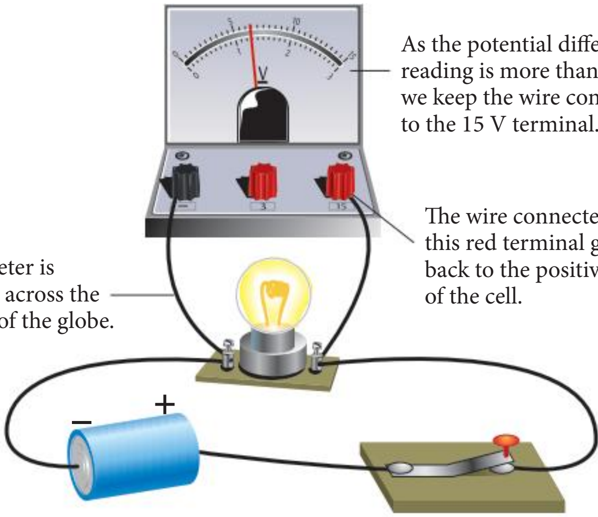


Figure 5.14 The voltmeter is measuring the difference in potential between two points. It is placed in parallel across an element.

QUESTION SET 5.2

Remembering

1 Is it easy or hard to separate like charges?

2 Is it easy or hard to separate unlike charges?

3 Which electron is at a higher potential, the electron at point a or the electron at point b? Why?

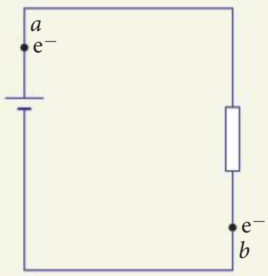


Figure 5.15 Circuit diagram

Applying

4 Calculate the potential difference of a battery that supplies 4.5 J of energy to every 3.0 C of charge that passes through the cell.

5 Calculate the potential difference of a battery that supplies 1.92×10^{-18} J of energy to every electron that passes through the battery.

Reflecting

6 People often 'charge' their phones when their battery is running low. Are they adding more charge to their phones, or is another process taking place? Describe what is happening.

Current

When there is an electrical potential difference between two points in a circuit, the electrons are attracted towards the positive terminal. This causes the electrons to move along the wire towards the positive terminal, which creates a flow of **current**.

The amount of current (I) depends on the amount of charge (Δq) that passes a point and how long it takes those charges to pass that point (t).

The unit of current is the ampere, A. One ampere is equal to one coulomb of charge (or 6.24×10^{18} electrons) passing a point in one second.

$$1 \text{ ampere} = 1 \text{ coulomb per second, } 1\,\text{A} = 1\,\text{Cs}^{-1}.$$

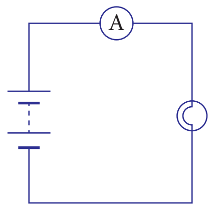


Figure 5.16 ▲ An ammeter is placed in series in the circuit.

The amount of charge that passes a point in a given time is:

$$\Delta q = It$$

Current is measured with an ammeter. It measures the amount of charge passing through a point each second. An ammeter is inserted in one part of a circuit in series as shown in Figure 5.16.

How fast does electricity move?

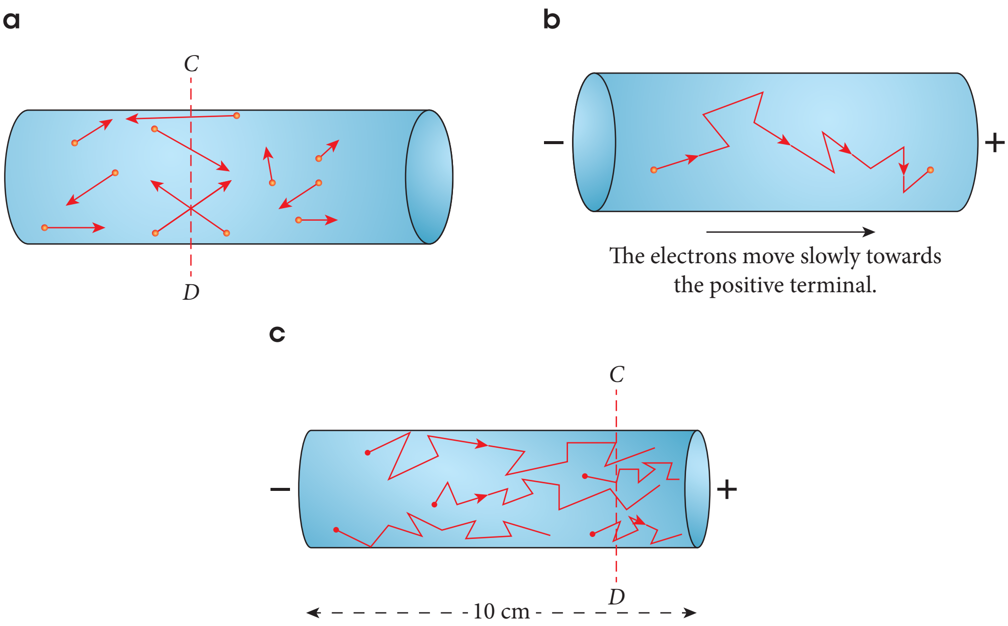
When you flick a light switch, a lamp comes on immediately. Actually, 'immediately' is not quite right. We can say that 'electricity travels' very quickly, but even that needs to be considered more closely.

When the switch connects the circuit, the electrical energy source – battery, solar panel, power station etc. – sets up a potential difference between the ends of the circuit. All of the electrons are encouraged to move along the wire towards the positive terminal. This electrical effect is established at about the speed of light, $3.00 \times 10^8\,\mathrm{m\,s^{-1}}$. In a 1.0 m long wire, the electrical effect takes about $\frac{1.0\ \mathrm{m}}{3.00 \times 10^8\ \mathrm{m\ s^{-1}}} = 3.3 \times 10^{-9}\,\mathrm{s}$ or about 3 nanoseconds to be established. That's *almost* immediate.

Bulk movement of electrons

Before the electrical effect is established, there is no current. The conduction band electrons move randomly in all directions. There is no net movement in a particular direction as shown in Figure 5.17(a).

Figure 5.17 ▶ a) Conduction electrons move randomly. The same number cross line *CD* from left to right as go the other way. b) The typical path for a conduction band electron after a potential difference is applied to a wire. c) Model of electron flow in a cylindrical section of metal wire.



When the circuit is closed, the potential difference is applied and the electrical effect is established. This imposes a general direction on the electrons – from the negative end towards the positive end. A typical path travelled by an electron is shown in Figure 5.17(b).

Let us now consider a simple model of electron current. Electrons are moving in a cylindrical length of wire at random speeds, sometimes left, sometimes right, sometimes up, sometimes down – in all directions. The electrical effect causes them, as a group, to move from left to right.

The 10 cm long section of wire shown has a current of 5.0 A in it. It has a cross-sectional area of $8.0 \times 10^{-2}\,\mathrm{cm^2}$. There are 6.0×10^{21} conduction electrons in this piece of wire. How long will it take for all these electrons to move through the wire and into the section to the right on the line *CD*? We will assume that the electrons are evenly distributed throughout the wire.

Remember that one coulomb of charge is 6.24×10^{18} electrons. A current of 5.0 A means that $5.0 \times 6.24 \times 10^{18} = 31 \times 10^{18}$ electrons must pass through this area each second. This is a very small proportion of all the electrons in the 10 cm section. Since the electrons are evenly distributed, all 31×10^{18} electrons that pass through the shaded area into the next part of the wire are contained in the right-hand end. This length is:

$$\frac{31 \times 10^{18}}{6.0 \times 10^{21}} \text{ of } 10 \text{ cm} = 5.2 \times 10^{-3} \text{ of } 10 \text{ cm}$$
$$= 5.2 \times 10^{-2} \text{ cm}$$

That is, when the current is 5.0 A, the required number of electrons that pass through the area in 1.0 s must travel 5.2×10^{-2} cm. Therefore, the speed, or **drift velocity**, of this group of electrons towards the positive end of the wire is:

$$5.2 \times 10^{-2} \text{ cm s}^{-1} \ (= 5.2 \times 10^{-4} \text{ m s}^{-1})$$

Drift velocity is the speed of electrons moving as a result of the electrical effect.

At this speed it will take a long time for all the electrons in the 10 cm section (6.0×10^{21}) to move past the shaded area. The time is:

$$t = \frac{1}{5.2 \times 10^{-2} \text{ cm s}^{-1}} \times 10 \text{ cm}$$
$$t = 192 \text{ s}$$
$$= \frac{192}{60} \text{ min}$$
$$= 3.2 \text{ min}$$

That is, it takes more than 3 minutes for a conduction band electron in this wire to travel 10 cm!

WORKED EXAMPLE 5.1

A current of 2.0 A flows in 20 cm of a conducting wire. There are 2.4×10^{23} conduction electrons in this section.

a How many electrons pass through the end of the section every second? (2 marks)

b What length of the wire do this number of electrons occupy? (2 marks)

c What is the magnitude of the drift velocity? (1 mark)

d How long, on average, will it take for a conduction electron to travel 1.0 m? (2 marks)

Answers	Logic	
a $I = \frac{\Delta q}{t}$	Use the correct formula.	1 mark
$\Delta q = It$		
$\Delta q = 2 \text{ A} \times 1 \text{ s}$		
$\Delta q = 2 \text{C}$		
$= 2 \times 6.24 \times 10^{18}$ electrons $\text{C}^{-1} = 1.25 \times 10^{19}$ electrons	Calculate the answer.	1 mark
b 2.4×10^{23} electrons in 20 cm		
$= 1.2 \times 10^{22}$ in 1 cm		
$\frac{1.25 \times 10^{19} \text{ electrons}}{1.2 \times 10^{22} \text{ electrons cm}^{-1}} = 0.001 \text{ cm}$		
$= 1 \times 10^{-5}$ m	Calculate the answer.	2 marks

c Electrons travel $1.0 \times 10^{-5}\,m\,s^{-1}$ | Calculate the answer. | 1 mark

d $v = \frac{x}{t}$ | Use the correct formula.

$t = \frac{x}{v}$

$t = \frac{1\,m}{1 \times 10^{-5}\,m\,s^{-1}}$ | Substitute the correct values. | 1 mark

$t = 100000\,s$

$t = 27.7\,h$ | Calculate the answer. | 1 mark

Try these yourself

1 A current of 120 mA flows in 2.0 cm of a conducting wire. There are 1.0×10^{21} conduction electrons in this section.
- a How many electrons pass through the end of the section every second? (2 marks)
- b What length of the wire do this number of electrons occupy? (2 marks)
- c What is the magnitude of the drift velocity? (1 mark)
- d How long, on average, will it take for a conduction band electron to travel 5.0 m? (2 marks)

2 A current of 50 mA flows in 14.5 cm of a conducting wire. There are 1.3×10^{22} conduction electrons in this section.
- a How many electrons pass through the end of the section every second? (2 marks)
- b What proportion of the wire do this number of electrons occupy? (2 marks)
- c What is the magnitude of the drift velocity? (1 mark)
- d How long, on average, will it take for a conduction band electron to travel 1.2 m? (2 marks)

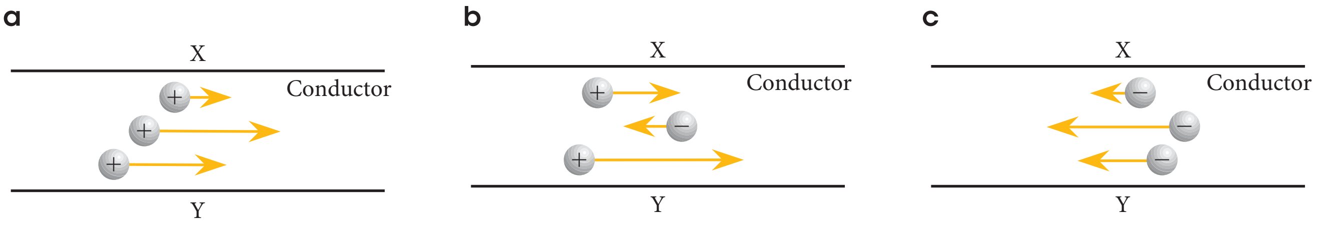
Conventional current

Up until this point we have talked about current as the flow of electrons because it is the movement of electrons that actually creates the current in a metal wire.

When scientists were discovering things about electricity they created the convention that electricity is the flow of positive charge.

Conventional current is the flow of positive charge. Physicists have kept the convention because it makes no difference to the laws and rules related to electricity, magnetism and electromagnetism.

Consider the following cases, in which charges are moving past a line XY, as shown in Figures 5.18. In all three cases, the right side becomes more positive by three units.



▲ **Figure 5.18**
a) Three positive charges moving to the right makes a current of +3 units. b) Two positive charges moving to the right and one negative charge moving to the left, making the right more positive, makes a current of +3 units. c) Three negative charges moving to the left, that is leaving the right, makes a current of +3 units.

Conventional current describes the direction of the current as the flow of positive charges flowing from the positive terminal to the negative terminal.

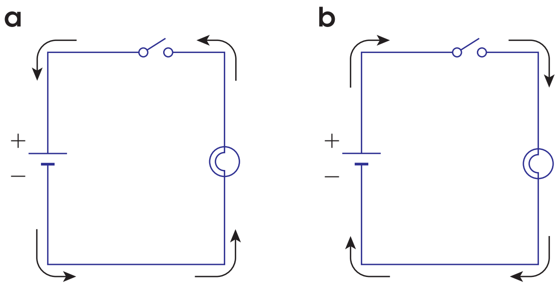


Figure 5.19
a) Electrons flow from the negative terminal to the positive terminal.
b) Conventional current is the flow of positive charge from the positive terminal to the negative terminal.

Direct current and alternating current

Electrical energy is supplied by either **direct current (DC)** or **alternating current (AC)** sources. In DC, the net charge flows in one direction; in AC, the charge flow alternates.

DC is used in mobile phones, torches and toys. The most common source of DC is a battery. DC is generally safer than AC because the rapid switching of AC current has serious effects on nerve responses.

AC is used in car alternators and motors. The electrical energy supplied by a power point in the wall is AC because it is simpler to produce and transmit, and, during transmission, power losses in the wires can be reduced. The standard AC power supply in Australia has a potential difference of 240 V and a frequency of 50 Hz. Different countries have different standards for voltage and frequency.

The operation of AC and DC power, and the transmission of AC power, will be studied in Unit 3.

QUESTION SET 5.3

Remembering

1 Which terminal (positive or negative) do electrons move towards when there is a potential difference applied across a wire?

2 Does conventional current describe the flow of positive charge or negative charge?

3 What is the difference between direct current and alternating current?

Understanding

4 Beth switches on a torch. Katie notices that the light shines almost instantly and says that the electrons travel very fast from the battery to the light bulb. Beth disagrees, saying that the electrons were already moving very fast before the switch was turned on. Explain who is correct and why.

Applying

5 Convert 0.63 A to milliamps.

6 How much charge passes through a load if a current of 4.0 A flows for 5.0 s?

7 3 coulombs of charge flow past in 30 s. How much current is this?

8 A current of 80 mA flows in 5.0 cm of a conducting wire. There are 3.0×10^{21} conduction electrons in this section.

- a How many electrons pass through the end of the section every second?
- b What length of the wire do this number of electrons occupy?
- c What is the magnitude of the drift velocity?
- d How long, on average, will it take for an electron to travel 1.0 m?

Reflecting

9 a What have you learnt about current?

- b Why do physicists use conventional current instead of electron current?

Types of circuits

There are two main types of circuits. **Series circuits** have only one path that the charge can flow through as in Figure 5.20(a). **Parallel circuits** have multiple paths that current can flow through as in Figure 5.20(b).

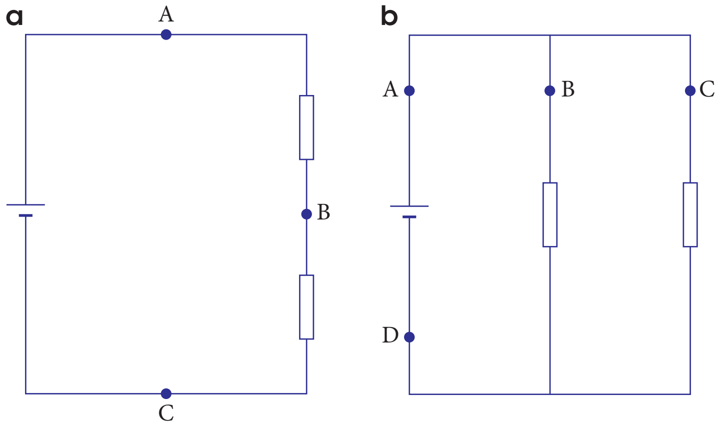


Figure 5.20 ▲ a) A series circuit with two resistors; b) a parallel circuit with two resistors

Series circuits

In a series circuit, the charged particles only have one pathway to go along. At each point in the path the flow of charge is the same. The current in Figure 5.20(a) is the same at points A, B and C. This is a consequence of the conservation of charge.

Parallel circuits

In a parallel circuit, there are at least two pathways for the charged particles to travel. At a junction they can go either way, but the total number of charged particles that arrives at the junction each second must be the same as the total number that leaves the junction each second.

Let us look at the current as it travels around the parallel circuit shown in Figure 5.20(b). The current travels from the positive terminal towards the first junction. Some of the current then travels towards point B, while the rest of the current travels toward point C. At the second junction, all of the current comes back to the one path and passes through point D.

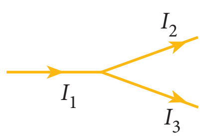


Figure 5.21 ▲ Kirchhoff's current law. The sum of the currents entering the junction equals the sum of the currents leaving the junction.

Kirchhoff's current law

Charge is conserved. The total number of charges entering a junction is the same as the number of charges leaving the junction. This means that the sum of all of the currents going into a junction is the same as the sum of all the currents out of that junction.

The total current into a junction = the total current out of the junction

$$\sum I_{in} = \sum I_{out} \quad \text{or} \quad I_1 = I_2 + I_3$$

WORKED EXAMPLE 5.2

In Figure 5.21 $I_1 = 3$ A and $I_2 = 1$ A. What is the value of I_3? (2 marks)

Answer	Logic	
$I_1 = I_2 + I_3$	Use Kirchhoff's current law.	1 mark
$I_3 = I_1 - I_2$		
$I_3 = 3\text{ A} - 1\text{ A}$		
$I_3 = 2$ A	Calculate the answer.	1 mark

Try these yourself

1 In Figure 5.21 $I_1 = 100$ mA and $I_2 = 10$ mA. What is the value of I_3? (2 marks)

2 In Figure 5.21 $I_2 = 0.75$ A and $I_3 = 1.5$ A. What is the value of I_1? (2 marks)

Kirchhoff's energy law

In any circuit, sources such as batteries provide potential energy to electrons, and loads such as resistors and globes convert that energy to other forms.

Figure 5.22 shows a simple circuit with a 12 V battery, a globe and resistor. Voltmeters are attached to the circuit as shown to measure the potential difference across each of these components.

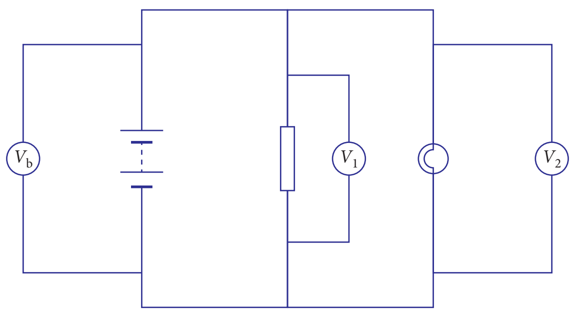
◀ **Figure 5.22** In a series circuit the potential difference provided by the source is shared between the loads.

The potential difference across the battery is 12 V. Remember that potential is potential energy per unit charge. Let's define the zero of potential energy as being at the negative terminal of the battery. (We can define our zero as anywhere that is convenient, as long as we then stick to it.) So that means that a charge q close to the positive terminal of the battery has potential energy $W = (12\text{ V}) \times q$. Think about moving around the circuit now. If you do a complete lap from the positive terminal of the battery back to the same point, you return to a point with the same potential energy again. Hence, whatever potential energy was gained by passing through the battery is lost as you pass through the globe and the resistor. We call components such as globes and resistors loads or sinks, because potential energy is lost (transformed into other forms) as a charge passes through these components.

The total change in potential around a complete loop in a circuit is zero. That means the total of all positive potential differences equals the total of all negative potential differences.

On our diagram we can see that $V_b = V_1 + V_2$. The potential difference provided by the battery is shared between the two loads.

Now consider Figure 5.23. We have the same components, but now each is connected directly across the battery terminals. We call this a parallel circuit. Apply the same idea as above. If a charge moves around one loop of this circuit, returning to its original position, the potential energy of the charge must be the same as when it started. It doesn't matter which loop we follow, this must be true. Hence, the potential difference measured at each voltmeter is the same in magnitude, although the sign of the potential difference is opposite for V_b to the two loads: $V_b = V_1 = V_2$.



◀ **Figure 5.23** In a parallel circuit the potential difference provided by the source is used by each load.

Kirchhoff's current law
The sum of the currents entering the junction equals the sum of the currents leaving the junction.

Kirchhoff's energy law
The sum of the potential difference from sources is equal to the sum of the potential difference used in the circuit.

WORKED EXAMPLE 5.3

In Figure 5.22 $V_1 = 2\,V$ and $V_2 = 4\,V$. What is the value of V_b? (2 marks)

Answer	Logic	
$V_3 = V_1 + V_2$	Use Kirchhoff's energy law.	1 mark
$V_3 = 2\,V + 4\,V$		
$V_3 = 6\,V$	Calculate the answer.	1 mark

Try these yourself

1 In Figure 5.22 $V_1 = 750\,mV$ and $V_2 = 1\,V$. What is the value of V_b? (2 marks)

2 In Figure 5.22 $V_2 = 8\,V$ and $V_b = 12\,V$. What is the value of V_1? (2 marks)

3 In Figure 5.23 $V_2 = 6\,V$. What is the value of V_b? (2 marks)

Properties of series and parallel circuits

When circuit elements are in series there is one pathway. The current is the same all the way around, but the potential differences are split and shared between each of the circuit elements.

In a parallel circuit, the current is split and shared in the different pathways. The potential difference is the same across parallel circuit elements.

QUESTION SET 5.4

Remembering

1 The amount of current changes at different points of a series circuit. True or false?

2 The potential energy changes at different points of a series circuit. True or false?

Applying

3 Figure 5.24 shows three ammeters A, B and C in three wires as part of a continuous circuit. If ammeter A reads 1.40 A and ammeter B reads 0.50 A, what is the reading on ammeter C?

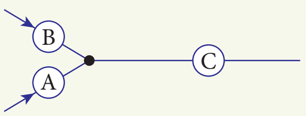


Figure 5.24

4 Figure 5.25 shows a series circuit with two resistors. If $V_1 = 4.2\,V$ what is the reading on V_2?

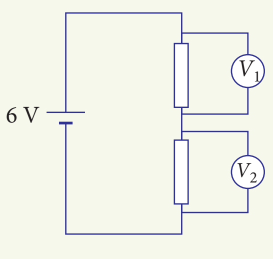
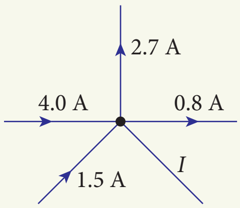


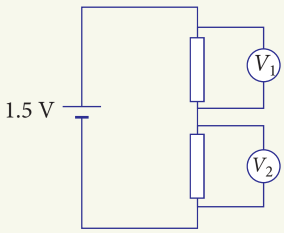
Figure 5.25

5 What is the size and direction of the current I in Figure 5.26?



◀ Figure 5.26

6 V_2 in Figure 5.27 has a reading of 750 mV. What is the reading of V_1?



◀ Figure 5.27

Reflecting

7 What is it about current that means it always has to be conserved at a junction?

Resistance

We have already looked at the movement of electrons through a wire. When a potential difference is applied across a wire, the electrons are encouraged to move from negative to positive. This movement is not in a straight line; the electrons zigzag as they collide with the atoms in the wire. The wire resists the flow of charge.

Some materials resist the movement of charge more than others. **Resistivity, ρ** refers to how much a material opposes the flow of charges, so $R \propto \rho$.

Table 5.2 Resistivity of some materials commonly used in electrical circuits

Material	Resistivity at 20°C (Ω m)
Copper	1.7×10^{-8}
Constantan	49×10^{-8}
Nichrome	1.1×10^{-6}
Carbon (graphite)	5.0×10^{-5}
Porcelain	3.0×10^{12}
Silicon	0.1–6.0
Germanium	0.001–0.5

When you measure the resistivity of a substance you need to state the temperature at which it was measured. As we have seen, temperature affects the conductivity of a material.

Recall the kinetic particle model of matter in Chapter 1.

The **resistance** of a wire is also affected by the size of the wire. As the cross-sectional area, A, or the thickness of the wire increases, the resistance decreases. There are more conduction electrons in any length of the wire, hence $R \propto \frac{1}{A}$. As the length of the wire increases, the resistance also increases. The charge has to collide with more atoms to get to the other end:

$$R \propto \ell$$

Altogether:

$$R = \rho \frac{\ell}{A}$$

where R = resistance, ρ = resistivity, ℓ = length and A = cross-sectional area.

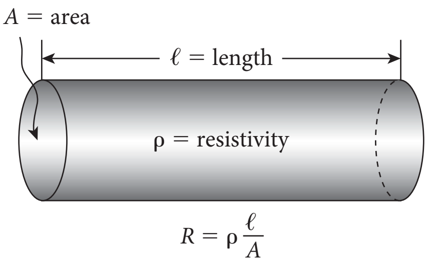


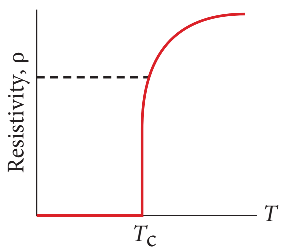
Figure 5.28 ▶ To calculate the resistance of a wire you need to know the resistivity of the material at a given temperature, the cross-sectional area of the wire and the length of the wire.

WOW

Transition temperature

Some metals and alloys have zero resistance once they are cooled to a temperature called the transition temperature.

This effect was first noticed in 1911, when Heike Kammerlingh Onnes cooled mercury to 4.2 K. Since then, other materials have been found to have much higher transition temperatures. The high temperature superconductors are ceramic materials.



◀ **Figure 5.29** Superconducting materials behave normally above a transition temperature (T_C) at which their resistivity falls to zero.

Table 5.3 Transition temperatures of some superconductors

Year	Material	Temperature (K)
1911	Mercury (Hg)	4.2
1941	Niobium nitride (NiN)	16
1986	Lanthanum barium copper oxide ($La_{2x}Ba_xCuO_4$)	30
1987	Yttrium barium copper oxide ($YBa_2Cu_3O_7$)	92
1994	Mercury thallium barium calcium copper oxide ($Hg_{12}Tl_3Ba_{30}Ca_{30}Cu_{45}O_{125}$)	138
2009	Thallium barium calcium copper oxide ($Tl_2Ba_2Ca_2Cu_3O_{10}$)	254

WORKED EXAMPLE 5.4

A wire has a cross-sectional area of 2.0 mm^2 and is 10.0 m long. If the resistance of the wire at room temperature is $60 \times 10^{-3}\,\Omega$, what is the resistivity of the wire? (3 marks)

Answer	Logic	
$R = \rho\frac{\ell}{A}$	Use the correct formula.	
$\Rightarrow \rho = \frac{RA}{\ell}$	Rearrange the formula to find ρ.	1 mark
$R = 60 \times 10^{-3}\,\Omega, \quad \ell = 10.0\,\text{m}$		
$A = 2.0\,\text{mm}^2 = 2.0\,(\text{m} \times 10^{-3})^2 = 2.0 \times 10^{-6}\,\text{m}^2$	Substitute the correct values.	1 mark
$\rho = \frac{60 \times 10^{-3}\Omega \times 2.0 \;\times 10 - 6\text{m}^2}{10\text{ m}}$		
$= 1.2 \times 10^{-8}\,\Omega\,\text{m}$	Calculate the answer.	1 mark

Try these yourself

1 A wire has a cross-sectional area of 0.5 mm^2 and is 2.0 m long. If the resistance of the wire is 50 mΩ, what is the resistivity of the wire? (2 marks)

2 A copper wire has a resistivity of 1.8×10^{-8}. It has a length of 50 cm and a cross-sectional area of 1 mm^2. What is the resistance? (2 marks)

EXPERIMENT 5.1

FINDING THE RESISTIVITY OF A WIRE

The resistance of a wire depends on length (ℓ), cross-sectional area (A) and the resistivity (ρ) of the wire. In this experiment we are going to calculate the resistance of a wire for different lengths. This will allow us to calculate the resistivity of the wire.

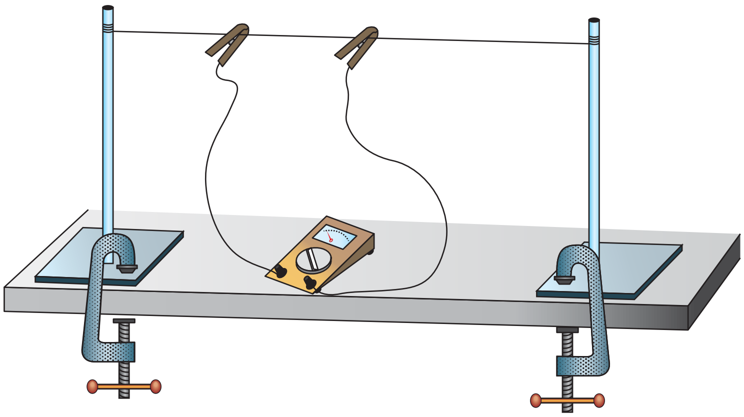
Aim

To find the resistivity of a wire

Materials

- 2 metres nichrome wire
- micrometer
- ruler
- digital multimeter to measure resistance
- two retort stands
- 2 G clamps

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
The wire could snap if it is too taut.	Do not tie the wire too tightly.



◀ Figure 5.30
Experimental set-up

Procedure

1 Set up the materials as shown in Figure 5.30.
2 Tie 1 m of wire between the two retort stands. Make sure that the wire is taut without kinks.
3 Use the micrometer to measure the thickness of the wire at three different points and calculate an average diameter. Use this measurement to calculate the cross-sectional area of the wire.
4 Measure the resistance for different lengths of the wire. Record your measurements in a table similar to the one below.
5 Collect at least 5 data points.

Results

Length	Resistance

Analysis of results

1 Plot a graph of resistance versus length with uncertainty bars.
2 Draw a line of best fit.
3 Use the gradient to find the resistivity of nichrome. Estimate the uncertainty in this value.

Discussion

1 Why did you measure data for at least 5 points?
2 Was the measured value for resistivity:
   a accurate?
   b precise? Explain your answer.
3 What could be some sources of error in your measurements?
4 Find a standard value of resistivity for nichrome.
   a What is the uncertainty in this standard value?
   b Considering the uncertainty you estimated for your result, does your value fit within the range of the standard value?
5 How could differences in temperature affect your results?

Ohm's law

Resistance affects current and voltage. If you have a 12 V battery in a circuit with a small resistance you will get a large current. If you have a 12 V battery in a circuit with a large resistance, you will get a smaller current.

Ohm's law describes the relationship between resistance, current and voltage. The resistance (R) of a circuit component is defined as the ratio of the potential difference (V) to the current (I):

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

Resistance is measured in ohms (Ω):

$$1\text{ ohm} = 1\,\Omega = \frac{1\text{ volt}}{1\text{ ampere}} = 1\,\text{V}\,\text{A}^{-1}$$

Ohm's law states that the current through a conductor between two points is directly proportional to the potential difference across the two points.

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

Ohmic devices

A resistor is an electrical component that restricts the flow of current. In an **ohmic device** the resistance is constant for a wide range of voltages and currents. For ohmic resistors:

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

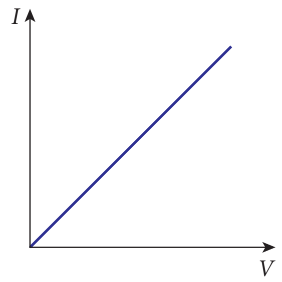
This can be transposed to become:

$$V = IR$$

Current through an ohmic device is directly proportional to the potential difference across it. This is known as Ohm's law, where R is the constant resistance.

A characteristic I–V graph for an ohmic resistor is shown in Figure 5.31. Ohmic devices have constant resistance, which can be calculated as the inverse of the gradient.

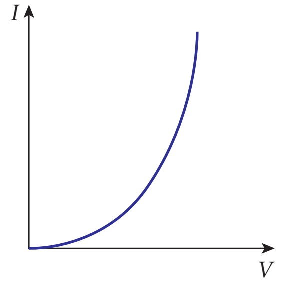
▼ Figure 5.31
In an ohmic device the resistance is constant as you change voltage and current. You can calculate the resistance as the inverse of the gradient.



Non-ohmic devices

The I–V graph of a **non-ohmic device** is shown in Figure 5.32. It is not a straight line, the resistance is not a constant for different amounts of voltage and current.

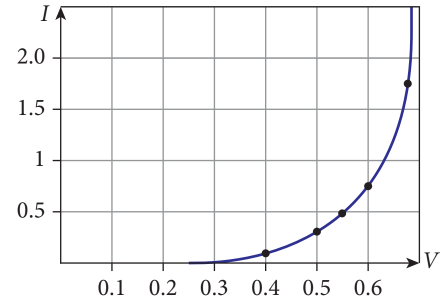
A number of different devices with non-constant resistance are used in electrical circuits. These include light globes, diodes, light-emitting diodes (LEDs), thermistors and light-dependent resistors (LDRs). To calculate the resistance for a certain voltage you need to use the equation $R = \frac{V}{I}$.



▲ Figure 5.32
In a non-ohmic device the resistance changes as you vary voltage and current.

WORKED EXAMPLE 5.5

The I–V graph for a diode is shown in Figure 5.33. Calculate the resistance when there is a potential difference of 0.6 V across the diode. (3 marks)



◀ Figure 5.33

Answer	Logic	
At 0.6 V the current is 0.75 A.	Read the correct value from the graph.	1 mark
$R = \frac{V}{I}$	Use the correct formula.	
$R = \frac{0.6\ V}{0.75\ A}$	Substitute the correct values.	1 mark
$R = 0.8\ \Omega$	Calculate the answer.	1 mark

Try these yourself

Travis measured the current through an LED for 5 different potential differences. You can see his measurements in the table.

Potential difference (V)	2.4	2.6	2.8	3	3.2
Current (mA)	0	1	4	12	25

a Plot the points on an I–V graph. (2 marks)

b Is this an ohmic or non-ohmic device? (1 mark)

c What is the resistance when there is a potential difference of 3 V across the LED? (2 marks)

QUESTION SET 5.5

Understanding

1 Why does resistance decrease with greater cross-sectional area?

2 The current in a simple circuit comprising a variable resistor and variable power supply increases. How could this occur?

3 What is the difference between an ohmic and a non-ohmic resistor?

Applying

4 When a potential difference of 16 V is applied across the ends of a wire, the current flowing in the wire is 2.4 A. Assume the wire is ohmic.

a What is the resistance of the wire?

b What potential difference is needed to make a current of 3.0 A flow through the wire?

5 What is the potential difference across a 1.8 kΩ resistor in which a current of 240 mA flows?

6 What is the resistance of a light bulb if it has a current of 0.8 A when there is a voltage of 240 V across its terminals? Note that this is the resistance at 240 V, light bulbs are not ohmic.

7 Two wires A and B of the same material have resistances of 6.0 Ω and 54.0 Ω, respectively. The length of A is double the length of B.

a What is the ratio of the diameter of A to the diameter of B?

b If the two wires are connected in parallel across a 6.0 V battery, what is the current in each wire?

Reflecting

8 Create an analogy for the factors affecting the resistivity of a wire where the charges are people.

Power

For a light globe to glow it needs to be provided with an energy source, such as a battery. The light globe transforms the electrical energy into light and heat energy. It is useful to know how quickly this energy is being transformed. Power is the rate of energy transformation.

Power is the rate of energy transfer per unit of time.

$$P = \frac{W}{t}$$

A 100 W globe uses 100 J of energy in 1 second. A 50 W globe takes 2 seconds to use the same amount of energy. The 100 W globe is therefore more powerful because it uses more energy per second than the 50 W globe. It is twice as powerful because it uses energy at twice the rate.

Power (P) is defined as the rate of energy transfer, either delivery or dissipation. It is the energy transfer per unit of time. Therefore:

$$P = \frac{W}{t}$$

When V is measured in volts (V), I is measured in amperes (A), t is measured in seconds (s) and energy (W) is measured in joules (J). The power is then measured in watts (W).

$$1 \text{ watt} = 1\,\text{W} = 1 \text{ joule per second} = 1\,\text{J s}^{-1}$$

Power delivery can be deduced from voltage and current values:

$$P = \frac{W}{t}$$

$$\Rightarrow P = \frac{Vq}{t} = V\left(\frac{q}{t}\right)$$

$$\Rightarrow P = VI$$

The amount of energy transferred can be deduced from:

$$P = \frac{W}{t}$$

$$\Rightarrow W = Pt$$

$$\Rightarrow W = VIt$$

W and *W*

W (not italic) is the unit watt or J s^{-1}, the unit of power. W (in italics) is the symbol we are using for energy, which is measured in J. Do not confuse the two!

Case study

Dr Jenny Riesz

Dr Jenny Riesz is a power system market analyst. She conducts research at the University of NSW into how to integrate renewable energy into power systems. Dr Riesz has also worked as an industry consultant at companies such as AECOM (a company that provides technical support in areas such as transportation, environment and water) and ROAM Consulting (a company that specialises in electricity market modelling).

Power systems are very complex. Generation of electricity needs to match demand at every point in the day. When the electricity market is designed, the aim is to create financial incentives for market participants to behave in ways that support the safe, secure and reliable operation of the power system.
Dr Riesz's particular focus is on designing the electricity markets that allow renewable technology to be integrated efficiently.


Courtesy of Dr Jenny Riesz

▲ Figure 5.34
Dr Jenny Riesz is a power system market analyst. She analyses power systems to look at ways to integrate renewable energy. This picture was taken at the Albany Wind Farm in Western Australia.

Solar photovoltaics and wind turbines are cheap compared with other types of renewable energy, so they are well suited to contributing to a cost-effective, 100%-renewable power system. However, these technologies are different from conventional fossil fuel technologies in a number of ways and we need to develop new strategies for integrating them. One challenge is that solar and wind generate different amounts of energy depending on the weather. Another challenge is that they are non-synchronous, which means that they don't stabilise the frequency of the system in the same way that fossil fuel generators do.

Dr Riesz and her team have discovered ways to make traditional power systems more flexible to incorporate electricity from renewable sources. The key is to create financial incentives for the companies to make it happen.

This work requires a solid understanding of power system engineering, as well as the fundamentals of economics and policy making. It is not enough to know how power systems work, you also need to know how they affect the financial market and how to change policies to change company behaviour.

Dr Jenny Riesz loves her job. She is passionate about reducing climate change and she gets to spend lots of time solving challenging problems to make it happen.

Questions

1. What are some of the reasons that power systems are complex?
2. Why do we need to develop new ways of managing power systems when we integrate lots of renewable energy generation?
3. Climate change is a complex process. Knowing how to produce electricity from renewable sources does not mean that the problem is solved. What else needs to happen to help solve the problem?

Energy and power

From $R = \frac{V}{I}$, we can deduce new relationships between power, voltage and current.

$$R = \frac{V}{I} \Rightarrow V = IR$$
$$P = VI = I \times (IR)$$
$$P = I^2R$$
$$R = \frac{V}{I} \Rightarrow I = \frac{V}{R}$$
$$P = VI = V \times \left(\frac{V}{R}\right)$$
$$P = \frac{V^2}{R}$$

For ohmic devices $R =$ constant and the calculations are a little simpler.

Converting energy values to kilowatt-hours

For domestic energy usage, the joule is too small. A larger unit, the kilowatt-hour (kWh), is used.

One kilowatt-hour (1 kWh) is the *energy* used by a 1 kW appliance in 1 hour:

$$E\,(\text{kWh}) = P\,(\text{kW}) \times t\,(\text{h}) = 10^3\,\text{W} \times (60\text{ min} \times 60\,\text{s})$$

Thus: $1\text{ kilowatt-hour} = 1\,\text{kWh} = 3.6 \times 10^6\,\text{J}$

Household appliances

Different appliances take different amounts of energy to run.

How much energy does it take to boil a kettle for a cup of tea?

If you put 500 mL of water into a 2000 W kettle, it takes approximately 90 seconds to boil.

$$\begin{aligned} W &= Pt \\ &= 2000 \times 90 \\ &= 180\,\text{kJ} \\ &= 0.05\,\text{kWh} \end{aligned}$$

If you boiled your 2000 W kettle 10 times each day, would it use more or less energy than running a 500 W fridge all day? You can find out by doing a household appliance energy-use audit.

ACTIVITY 5.2

HOUSEHOLD APPLIANCE ENERGY USE AUDIT

Aim

To compare the energy use of different appliances in your home

You will need

- copy of the table below
- pen
- calculator

What to do

Find out the power rating of all of the electrical devices in your house. Estimate the number of hours that you use each device each week. Complete the table to compare the amount of energy used by each device. An example is given in the first row of the table.

Room	Appliance	Power rating for electric appliances (kW)	Estimated weekly household use (h)	Amount of energy used by this appliance (kWh)	Cost (1 kWh = 28 cents)
Study	Computer	0.25	10	2.5	$0.70

What did you discover?

1 Were you surprised by any of your findings?

2 What is the most expensive appliance to run in your house?

Taking it further

Is it more energy efficient to heat up 500 mL of water in the kettle, microwave oven or on the stove? How could you find out?

WORKED EXAMPLE 5.6

A current of 2.0 A flowing in a heater for an hour converts 1.7 MJ of electrical energy into heat energy.

a How much charge is transferred through the heater? (2 marks)

b What is the potential difference across the heater? (2 marks)

c Find the power rating of the heater in kW. (2 marks)

d How much energy (in kWh) is used if the heater runs for 2 hours? (2 marks)

Answers	Logic	
a $I = \frac{q}{t}$, $q = It$	Use the correct formula.	
$\Rightarrow q = 2.0\ \text{A} \times (60\ \text{min h}^{-1} \times 60\ \text{s min}^{-1})$	Substitute the correct values.	1 mark
$\Rightarrow q = 7.2 \times 10^3\ \text{C}$	Calculate the answer.	1 mark

b $V = \frac{W}{q}$ — Use the correct formula.

$\Rightarrow V = \frac{1.7 \times 10^6 \text{ J}}{7.2 \times 10^3 \text{ C}}$ — Substitute the correct values. 1 mark

$\Rightarrow V = 240 \text{ V}$ — Calculate the answer. 1 mark

c $P = VI$ — Use the correct formula.

$\Rightarrow P = 240 \text{ V} \times 2.0 \text{ A} = 480 \text{W}$ — Substitute the correct values. 1 mark

$\Rightarrow P = 0.48 \text{kW}$ — Calculate the answer. 1 mark

d $E = Pt$ — Use the correct formula.

$\Rightarrow E = 0.48 \text{ kW} \times 2.0 \text{ h}$ — Substitute the correct values. 1 mark

$\Rightarrow E = 0.96 \text{kWh}$ — Calculate the answer. 1 mark

Try these yourself

A current of 3.0A flowing in a heater for 4 hours converts 5.2MJ of electrical energy into heat energy.

a How much charge is transferred through the heater? (2 marks)
b What is the potential difference across the heater? (2 marks)
c Find the power rating of the heater in kW. (2 marks)
d How much energy (in kWh) is used if the heater runs for 2 hours? (2 marks)

QUESTION SET 5.6

Remembering

1 What is power?

2 A circuit with *emf V*, a resistor, *R*, and a total current, *I*, transforms energy, *E*, in a time interval, *t*. Write all the equations that show how power, *P*, is related to these variables.

Understanding

3 How can you find the energy transformed in a circuit given *emf*, resistance and time?

Applying

4 A 100W light globe is left on for 1.0 hour.
How much energy is used by the globe? Give your answer in joule and kilowatt-hour.

5 A 1.5kW toaster cooks two pieces of toast in 80s. If all the energy is used in cooking the toast, how many kilowatt-hours of energy are used per piece of toast?

6 How much energy is transformed in a hair dryer if it has a voltage drop of 240V and 100C of charge passes through it?

Analysing

7 Two globes drawing 50 W are connected in series across a 10.0 V supply. Each globe has a resistance of 0.5 Ω. Find:

a the potential difference across one globe.
b the current in both globes.

CHAPTER SUMMARY

- Electricity is a convenient form of energy that can be easily converted into other forms of energy such as light, heat and sound.
- There are two types of charge: positive and negative.
- Like charges repel. Unlike charges attract.
- Charge is measured in coulomb, C. An electron has a charge of 1.602×10^{-19} C.
- Conductors are materials in which the electrons are relatively free to move. Insulators are materials that do not contain many free electrons.
- An electrical circuit is a complete loop through which charges can flow.
- Energy is required to separate opposite charges. Energy is required to move two like charges closer together.
- Charges in different parts of a circuit have different amounts of electrical potential energy. Potential difference is the difference in potential energy per unit charge between two points in the circuit:

  $$V = \frac{W}{q}$$

  where V = potential difference, W = potential energy and q = charge.
- Current is the flow of charge per unit time:

  $$I = \frac{\Delta q}{t}$$

  where I = current, Δq = the amount of charge that passes a point in the circuit and t = time interval.
- Conventional current is the flow of positive charge from the positive terminal to the negative terminal.
- In a series circuit, there is one pathway. The current is the same all the way around, but the potential differences are split and shared between each of the circuit elements.
- In a parallel circuit, the current is split and shared in the different pathways. The potential difference is the same across parallel circuit elements.
- Resistance is opposition to the flow of electrons.
- Resistance depends on the resistivity, the cross-sectional area and the length of the material:

  $$R = \rho\frac{\ell}{A}$$

  where R = resistance, ρ = resistivity, ℓ = length and A = cross-sectional area.
- Resistance affects current and voltage:

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

  where R = resistance, V = potential difference and I = current.
- Power is the rate of energy transfer:

  $$P = \frac{W}{t}$$

  $$P = VI$$

  $$P = I^2R$$

CHAPTER GLOSSARY

alternating current (AC) current that changes direction periodically

conduction band a range of electron energies in which the electrons are relatively free to move away from the atom

conductor a type of material that allows electrons to flow

current the rate of flow of charge

direct current (DC) current that is always in one direction

drift velocity the speed of electrons moving as a result of the electrical effect

electrical circuit a complete loop through which charges can flow

emf electromotive force; source of potential energy per charge

insulator a type of material that does not allow electrons to flow

metal lattice a regular arrangement of large numbers of metal atoms, ions or molecules

non-ohmic resistance is not constant: $R \neq \frac{V}{I}$

ohmic device a component with constant resistance for different values of V and I

parallel circuit circuit with multiple paths through which current can flow

potential difference (V) potential energy per charge: $V = \frac{W}{q}$, sometimes referred to as voltage

potential energy energy that can be considered to be 'stored' within a body due to its position, composition or molecular arrangement

resistance opposition to the flow of electrons

resistivity, ρ how much a material opposes the flow of charges

semiconductor a material that conducts electricity less than conductors but more than insulators

series circuit circuit with only one path through which the charge can flow

valence band a range of electron energies for which the electrons are still attached to the atom; energy levels involved in chemical reactions

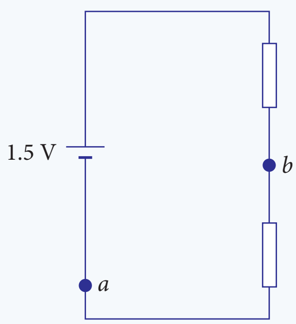
CHAPTER REVIEW QUESTIONS

Remembering

1 List some of the forms of energy into which electricity can be converted.

2 What is the difference between a conductor and an insulator?

3 Which positive charge, *a* or *b*, in Figure 5.35 is at a higher electric potential?



◀ Figure 5.35

4 Current in a metal is the flow of protons. True or false? Explain your answer.

5 What is the difference between a series circuit and a parallel circuit?

6 a What is resistance?

b What are some factors that affect resistance?

7 Power and current are both the measure of something over time. Is this true or false? Explain your answer.

8 Define these terms.

a Ohmic device

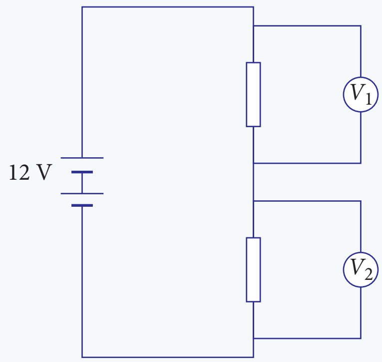
b Non-ohmic device

Understanding

9 A student claims that the charge on an ion in a solution is 2.4×10^{-19} C. Why must the student be incorrect?

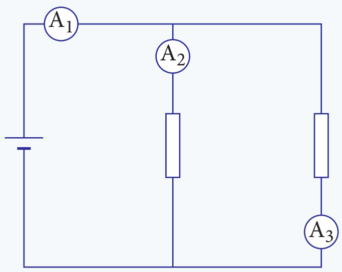
10 How does a potential difference create a flow of electrons?

11 The reading of V_1 in Figure 5.36 is 5 V. What is the reading of V_2?



◀ Figure 5.36

12 Figure 5.37 shows a parallel circuit. The current through A_1 is 100 mA and the current through A_2 is 25 mA. What is the current through A_3?



◀ Figure 5.37

13 Why does a longer wire have a larger resistance than a shorter wire?

14 A 100 W globe uses 100 J in one second. A 200 W globe uses 100 J in 0.5 seconds.

a Which globe used more energy?
b Which globe is more powerful?
c Which globe is brighter?
d Which globe is cheaper to run?

Applying

15 How much energy does a 12.0 V car battery supply to every coulomb of charge?

16 How much energy does a 12.0 V car battery supply to every electron?

17 A current of 0.50 A flows for 10 minutes in an electrical conductor. Calculate the number of coulombs of charge that pass a given point.

18 1.2×10^4 C of charge flows through a conductor in 1.0 hour. What is the electric current (assuming that this is a steady current)?

19 A steady current of 0.50 A flows in a wire. How many electrons flow past any point in the wire in 1.0 s?

20 Calculate the potential difference of a battery that supplies 6.0 J of energy to every 0.5 C of charge that passes through the battery.

21 What is the voltage drop across a 220 mΩ resistor if it carries a current of 4.0 A?

22 Calculate the potential difference of a battery that supplies 960 J of energy to every 80.0 C of charge that passes through the battery.

Analysing

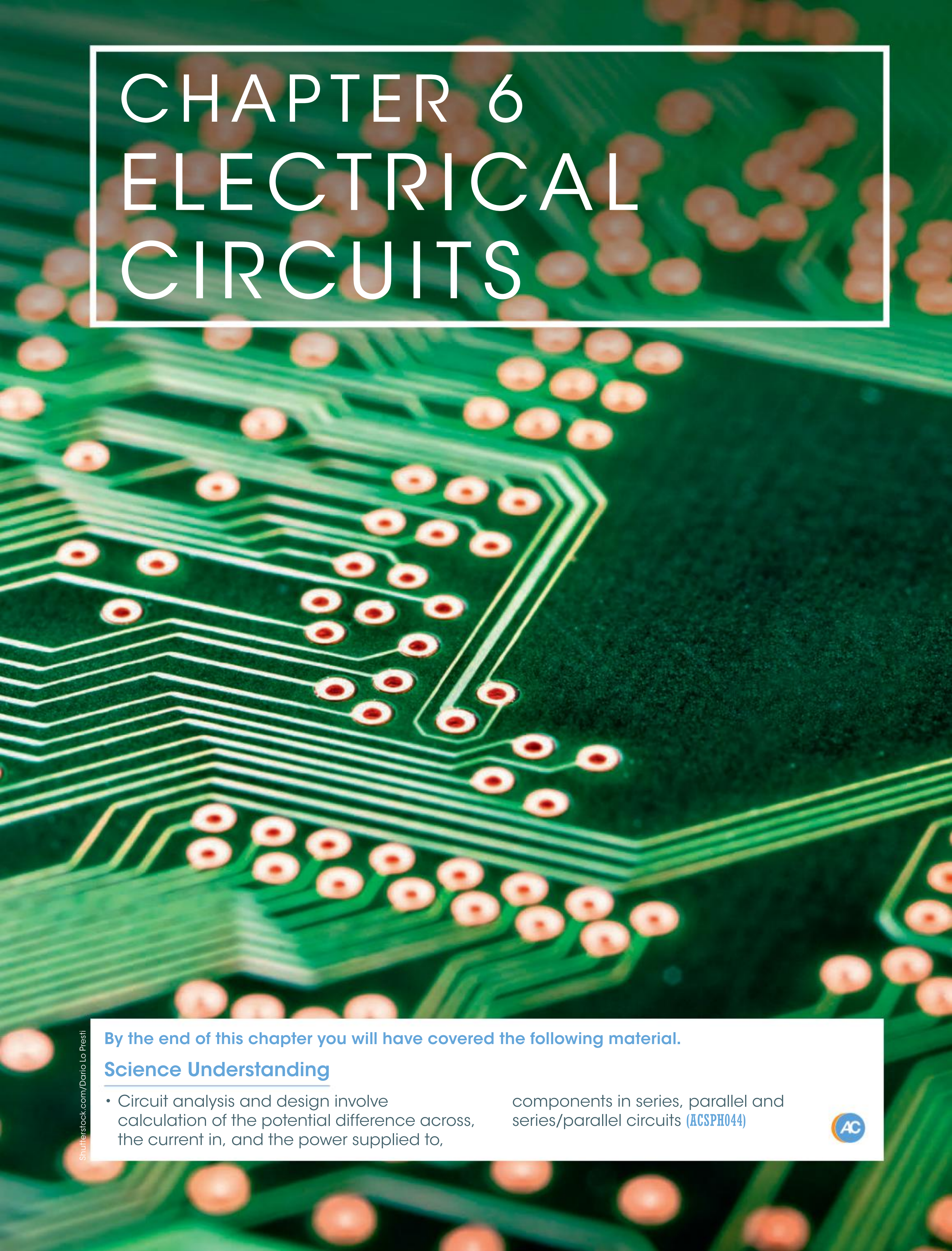
23 What is the difference between electrical potential energy and potential difference?

24 A current of 2.0 A flows in a battery when a light globe is connected across the terminals of the battery. The potential drop across the terminals is 6.0 V.

a What is the quantity of electrical charge that flows in the globe each second?
b How much energy is given to each coulomb of charge that passes through the battery?
c How long does it take the battery to supply 480 J of energy?
d How many coulombs of charge will pass through the battery in this time?

Reflecting

25 How do the words 'charge', 'energy', 'potential' and 'resistance' get used in everyday language? How does this differ from the scientific use of these words?



CHAPTER 6
ELECTRICAL CIRCUITS

Shutterstock.com/Dario Lo Presti

By the end of this chapter you will have covered the following material.

Science Understanding

- Circuit analysis and design involve calculation of the potential difference across, the current in, and the power supplied to, components in series, parallel and series/parallel circuits (ACSPH044)

Introduction

In Chapter 5 we looked at the theory behind electricity. We will now look at its application in electronic circuits and devices.

Electronics has revolutionised the way that we live. Computers and mobile phones look very different from those of 20 years ago. Technology is continually making devices faster, smaller and more user friendly.

Circuit analysis

To analyse a circuit you need to look at the energy source that is providing the potential difference and the components that are using the energy.

An energy source, for example a battery, is a source of electromotive force (***emf***), which provides a potential difference for current to flow. Sometimes batteries can be placed in series to provide a larger *emf*.

The flow of charge, the current, can be controlled by devices placed in the circuit, such as a switch or a variable resistor. The wires connecting the components and any measuring devices placed in the circuit can also have some effect, although most circuits are designed to minimise this effect.

Thévenin's theorem allows us to reduce complex circuits into a simpler form. Two types of arrangement can be combined to make complex circuits: elements in series combined with elements in parallel. Table 6.1 summarises the difference between series and parallel circuits.

Table 6.1 Summary of circuit elements in series and parallel circuits

Circuit elements in:	Current in each circuit element is:	Potential difference across each element is:
series	same	shared
parallel	shared	same

Thévenin's theorem

Every circuit, even if it has several energy sources and several loads, can always be reduced to the simplest of forms to make it easier to analyse. A circuit can be reduced to one source of *emf* that is the equivalent of all the sources, and one load that is the equivalent of all the loads. This is **Thévenin's theorem**.

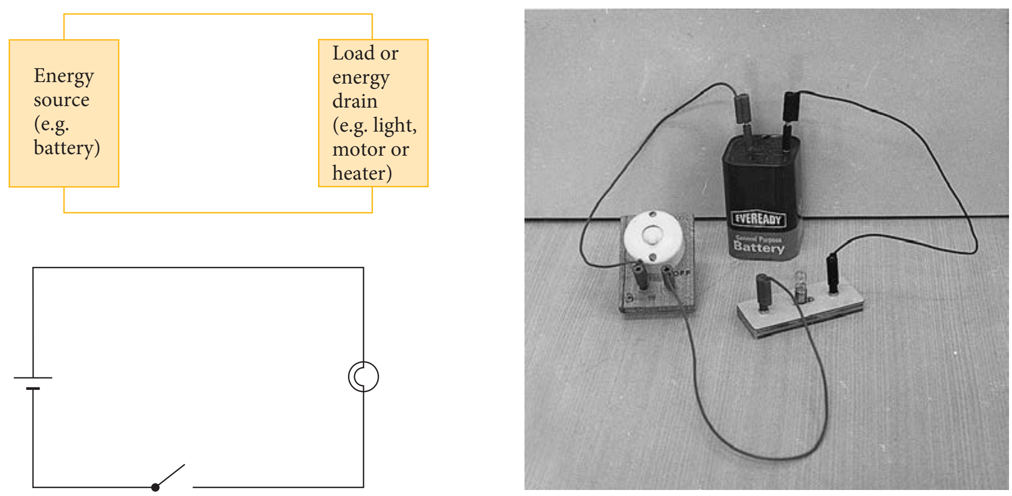


Figure 6.1 ▶ All circuits can be reduced to the simplest arrangement of one equivalent source and one equivalent load.

Thévenin's theorem

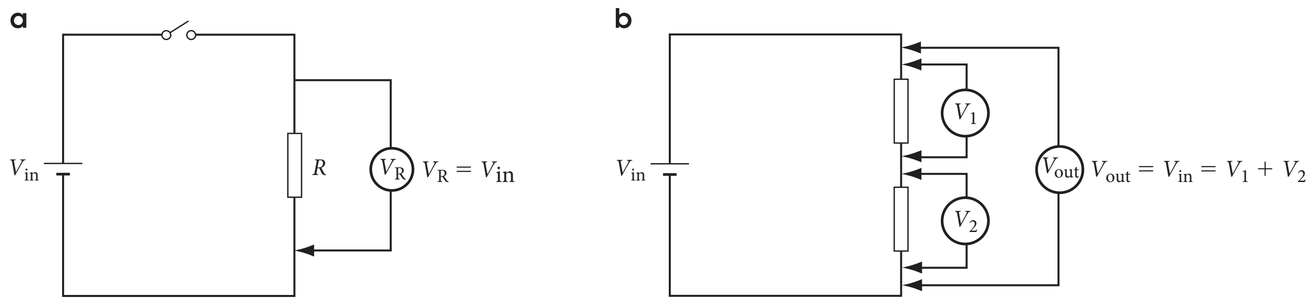
A circuit can be reduced to one source that is the equivalent of all the sources, and one load that is the equivalent of all the loads.

In Figure 6.1, which comprises a battery, a switch and a globe, the potential energy from the source (battery) is completely transformed in the load (light globe).

We will use Thévenin's theorem to analyse series circuits and parallel circuits. Thévenin's theorem is particularly useful when analysing combination circuits. Combination circuits contain both series and parallel sections, so it is important to be able to simplify them using equivalent loads.

Resistors in series circuits

The potential energy per unit charge, *emf*, provided by a battery is used by the circuit elements. If there is one resistor in a circuit, then all the energy per unit charge, the potential difference, is across the resistor, and is used in that resistor (Figure 6.2(a)). If two resistors are placed end to end in a series circuit (Figure 6.2(b)), the energy is shared between the two resistors. The current is the same in each resistor.



▲ Figure 6.2
a) A simple circuit comprising a source of emf and one resistor. The potential difference across the resistor is the same as the potential difference across the source. b) A series circuit comprising a potential difference source and two resistors in series. The potential difference across both resistors is the same as the potential difference across the source. The potential difference is shared between the resistors; the current is the same in both resistors.

Any series circuit can be reduced, using Thévenin's theorem, to a single source and a single equivalent resistor.

In a series circuit, the potential difference is shared:

$$V_T = V_1 + V_2$$

In a series circuit there are no junctions, so the current in each resistor is the same:

$$I_T = I_1 = I_2$$

Altogether, we can deduce the equivalent resistance:

$$\frac{V_T}{I_T} = \frac{V_1}{I_1} = \frac{V_2}{I_2}$$

$$\Rightarrow R_T = R_1 + R_2$$

The equivalent resistance in a series circuit is the sum of all the resistances:
$R_T = R_1 + R_2 + \ldots$

The Thévenin equivalent circuit comprises a potential difference, V, and the equivalent resistance, R_T.

Note that this result arises from the definition of resistance as the ratio of potential difference to current. It is applicable to all resistances in series.

Using Figure 6.2(b), we can show that the ratio of the potential differences across each resistor is equal to the ratio of the resistances. The current is the same in each resistor, so:

$$I = \frac{V_1}{R_1} = \frac{V_2}{R_2}$$

$$\Rightarrow \frac{V_2}{V_1} = \frac{R_2}{R_1}$$

This shows that the potential difference is divided in the ratio of the resistances.

WORKED EXAMPLE 6.1

Figure 6.3 shows a series circuit with two resistors.

a Use Thévenin's theorem to re-draw the circuit. (1 mark)

b What is the total resistance in the circuit? (1 mark)

c What is the current in the circuit? (1 mark)

d Calculate the potential difference across:

- **i** R_1. (1 mark)
- **ii** R_2. (1 mark)

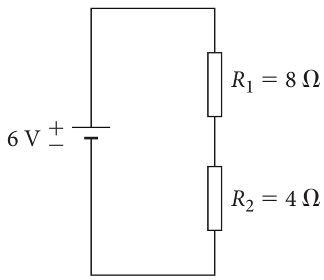
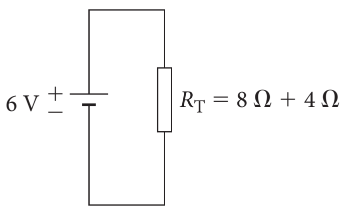


Figure 6.3 ◀

Answers	Logic	
a 	Draw correct equivalent circuit diagram.	1 mark
b $R_T = R_1 + R_2$ $R_T = 8\,\Omega + 4\,\Omega$ $R_T = 12\,\Omega$	Find the correct sum of resistors.	1 mark
c $R = \frac{V}{I}$ $I = \frac{V}{R}$ $I = \frac{6\text{ V}}{12\ \Omega}$ $I = 0.5\text{ A}$	Calculate current using total voltage and total resistance	1 mark
d **i** $I = 0.5\text{ A}$, $R = 8\,\Omega$, $V = ?$ $V_1 = I_1R_1$ $V_1 = 0.05\text{ A} \times 8\,\Omega$ $V_1 = 4\text{ V}$	Use the total current and particular resistance to calculate potential drop across V_1.	1 mark

▶

ii $V_T = V_1 + V_2$ | Use conservation of energy per charge to calculate potential drop across V_2. | 1 mark

$V_2 = V_T - V_1$

$V_2 = 6\,V - 4\,V$

$V_2 = 2\,V$

Try these yourself

Series circuits similar to that in Figure 6.3 were constructed with the following circuit elements:

A $V = 6\,V$, $R_1 = 2\,\Omega$, $R_2 = 6\,\Omega$

B $V = 12\,mV$, $R_1 = 6\,\Omega$, $R_2 = 6\,\Omega$

C $V = 1.5\,V$, $R_1 = 20\,\Omega$, $R_2 = 10\,\Omega$

a Draw each circuit diagram. (3 marks)

b For each circuit diagram draw the equivalent (Thévenin) circuit. (3 marks)

c Find the current in each resistor. (3 marks)

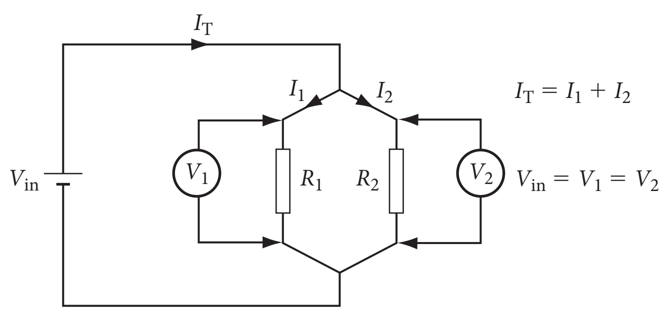
d For each circuit calculate the potential difference across:

i R_1. (3 marks)

ii R_2. (3 marks)

Resistors in parallel circuits

As you learnt in the previous chapter, the energy per charge provided by a battery, the *emf*, is used by the circuit elements. If two resistors are placed side by side, so that the current is shared, the elements are in parallel (Figure 6.4). Because the source of energy per charge is across both resistors, both resistors use the same amount of energy per charge. The energy per charge (potential difference) is the same across both resistors. The current is shared between the resistors.



◀ **Figure 6.4** In parallel circuits, current is shared between circuit elements such as resistors.

Any parallel circuit can be reduced, using Thévenin's theorem, to a single source and a single, equivalent resistor.

In a parallel circuit, the potential difference is the same across each resistor:

$$V_T = V_1 = V_2$$

In a parallel circuit, the total current in the circuit is shared between each resistor:

$$I_T = I_1 + I_2$$

Altogether, we can deduce the equivalent resistance:

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

Thus: $I_T = I_1 + I_2$

becomes: $$\frac{V_T}{R_T} = \frac{V_T}{R_1} + \frac{V_T}{R_2}$$

$$\Rightarrow \frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2}$$

Note that this result arises from the definition of resistance as the ratio of the potential difference to current. It is applicable to all resistors in parallel.

To calculate the equivalent resistance in a parallel circuit use the equation:

$$\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} \cdots + \frac{1}{R_n}$$

Using Figure 6.4, we can show that the ratio of the current in each resistor is equal to the inverse ratio of the resistances. The potential difference across each resistor is the same, so:

$$V = I_1R_1 = I_2R_2$$

$$\Rightarrow \frac{I_1}{I_2} = \frac{R_2}{R_1}$$

WORKED EXAMPLE 6.2

A circuit with two resistors in parallel is shown in Figure 6.5.

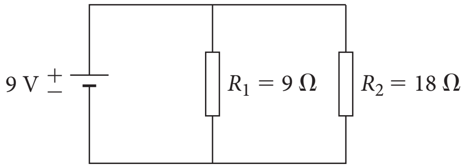


Figure 6.5

a Use Thévenin's theorem to re-draw the circuit. (1 mark)

b What is the total resistance in the circuit? (2 marks)

c What is the total current in the circuit? (2 marks)

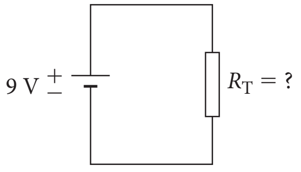
d What is the current in:

  i R_1? (1 mark)

  ii R_2? (1 mark)

Answers

a



Logic

Draw the correct Thévenin circuit. 1 mark

b $\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{1}{9\,\Omega} + \frac{1}{18\,\Omega}$ — Substitute the correct values. — 1 mark

$\Rightarrow \frac{1}{R_T} = \frac{2}{18\,\Omega} + \frac{1}{18\,\Omega}$

$\Rightarrow \frac{1}{R_T} = \frac{3}{18\,\Omega}$

$\Rightarrow R_T = \frac{18\,\Omega}{3} = 6\,\Omega$ — Calculate the answer. — 1 mark

c $R_T = \frac{V_T}{I_T}$ — Use Kirchhoff's energy law. — 1 mark

$\Rightarrow I_T = \frac{V_T}{R_T}$

$\Rightarrow I_T = \frac{9\text{ V}}{6\,\Omega} = 1.5\text{ A}$ — Calculate the answer. — 1 mark

d i $V_T = V_1 = V_2$ — Substitute the correct values. — 1 mark

$\Rightarrow I_1 = \frac{V_1}{R_1} = \frac{9\text{V}}{9\,\Omega} = 1\text{A}$ — Calculate the answer.

ii $I_T = I_1 + I_2$ — Substitute the correct values. — 1 mark

$I_2 = I_T - I_1 = 1.5 - 1.0 = 0.5\text{A}$ — Calculate the answer.

Try these yourself

For the following parallel circuits: (7 marks)

1. $V = 6\text{V}$, $R_1 = 2\,\Omega$, $R_2 = 6\,\Omega$
2. $V = 12\text{mV}$, $R_1 = 6\,\Omega$, $R_2 = 6\,\Omega$
3. $V = 1.5\text{V}$, $R_1 = 20\,\Omega$, $R_2 = 0.1\,\Omega$

a Draw the circuit diagram.

b What is the effective resistance of the circuit?

c What is the total current in the circuit?

d What is the current in:

i R_1?

ii R_2?

EXPERIMENT 6.1

INVESTIGATING CURRENT IN SERIES AND PARALLEL CIRCUITS

In a series circuit there is only one current path. In a parallel circuit there is more than one path. Changing the type of circuit affects the way the current and voltage are distributed.

Aim

To measure the change in current in series and parallel circuits

Materials

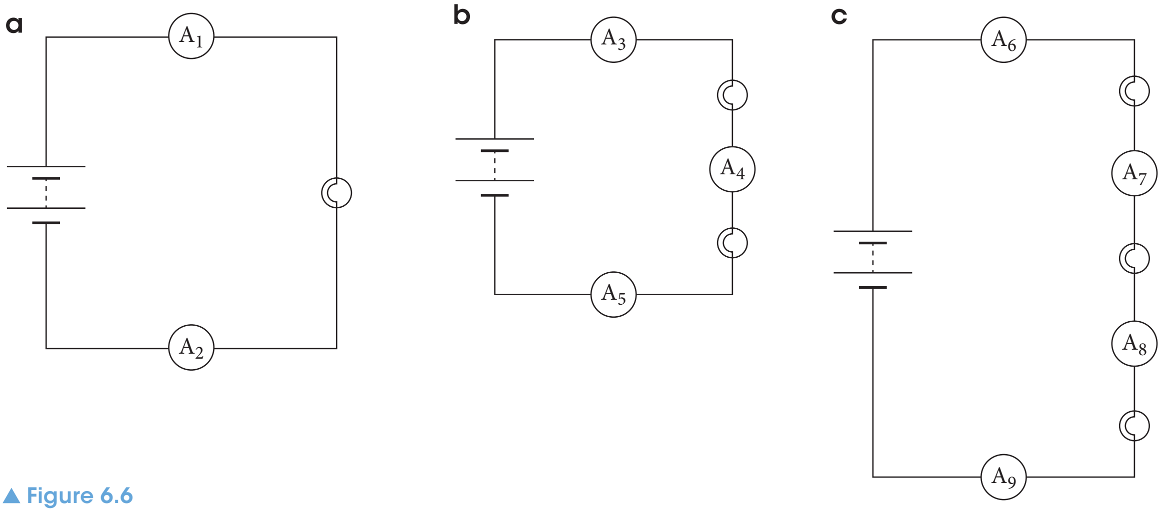
- multimeter
- 9V battery
- 3 globes rated to 9V
- wires

Procedure

1 Connect a series circuit with one globe as shown in Figure 6.6(a). Measure and record the current at each point marked in the circuit.

2 Connect a second globe in the series circuit as shown in Figure 6.6(b). Measure and record the current at each point marked in the circuit.

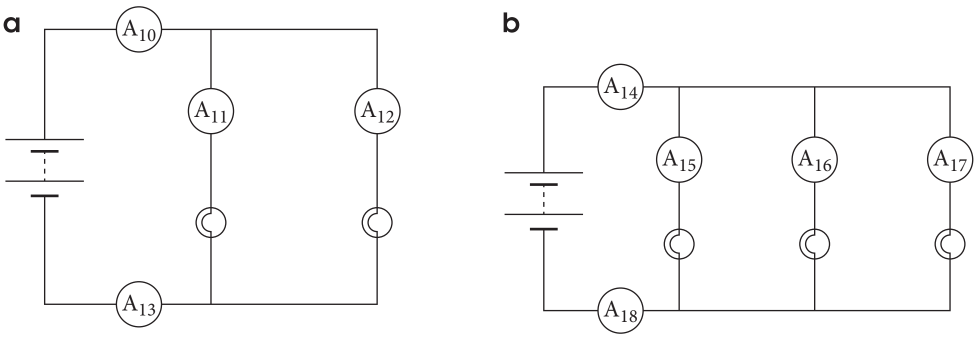
3 Connect a third globe in the series circuit as shown in Figure 6.6(c). Measure and record the current at each point marked in the circuit.



▲ Figure 6.6

4 Connect two lamps in parallel as shown in Figure 6.7(a). Measure and record the current at each point marked in the circuit.

5 Connect three lamps in parallel as shown in Figure 6.7(b). Measure and record the current at each point marked in the circuit.



▲ Figure 6.7

Results

Series circuit	Current
1 globe	A_1 =
	A_2 =
2 globes	A_3 =
	A_4 =
	A_5 =
3 globes	A_6 =
	A_7 =
	A_8 =
	A_9 =

Parallel circuit	Current
2 globes	A_{10} =
	A_{11} =
	A_{12} =
	A_{13} =
3 globes	A_{14} =
	A_{15} =
	A_{16} =
	A_{17} =
	A_{18} =

Analysis of results

Draw a graph comparing the number of lamps on the horizontal axis with the total current on the vertical axis.

Discussion

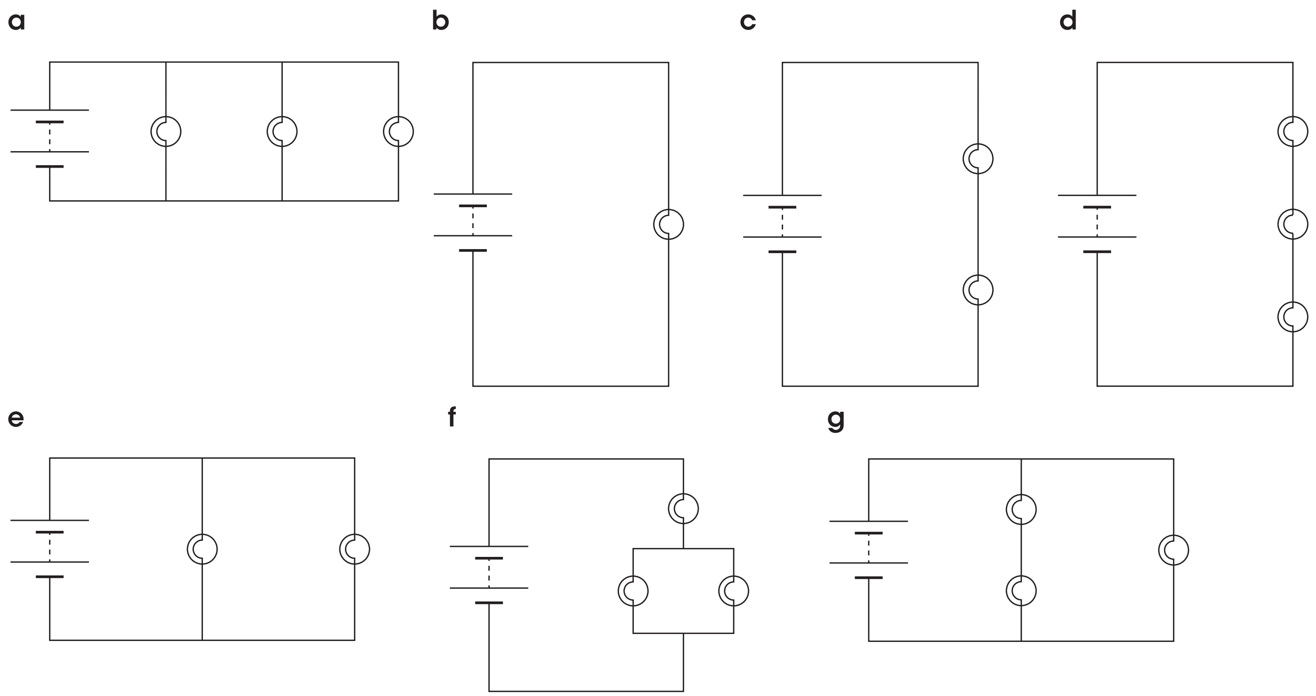
1 What happened to the current as the number of globes increased in the series circuit?
2 What happened to the relative brightness as the number of globes increased in the series circuit?
3 What did you notice about the values of current measured in the 3-globe series circuit?
4 What did you notice about the values of current measured in the 3-globe parallel circuit?

Conclusion

Write a conclusion that relates directly to the aim of this experiment and that uses the data appropriately.

Taking it further

Measure the potential difference across each globe for the circuits in Figure 6.8. What do you notice about the values?



▲ Figure 6.8

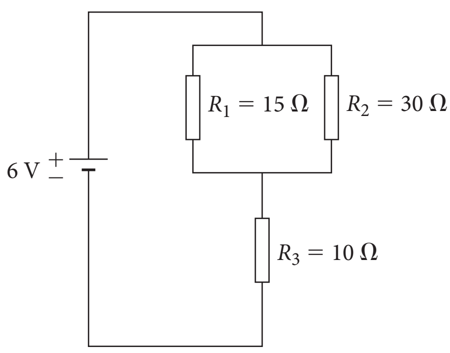
Combination circuits

Combination circuits have both series and parallel components, as shown in the circuit in Figure 6.9. It is particularly useful to use Thévenin's theorem to find the equivalent resistance for analysis.

Resistance is defined as the ratio of potential difference to current: $R = \frac{V}{I}$. Ohm's law, $V = IR$, is a useful transposition of the definition. Ohmic resistors have constant resistance over a relatively wide range of temperatures. Non-ohmic resistors have variable resistances that depend on the $V:I$ ratio.

Care needs to be taken when using Ohm's law, $V = IR$, to solve combination circuits. For a start, you must be sure you are working with ohmic resistors. Then, it is important to work out to which part of the circuit you are applying the equation. Are you applying it to the whole circuit, or are you applying it to one component of the circuit?

- To use $V = IR$ on the whole circuit, you need to know the total resistance.
- To use $V = IR$ on one component, you need to know the voltage drop and/or the current in that component.



▲ Figure 6.9
A combination circuit

When using Ohm's law:

- **use $V_T = I_T R_T$ on whole circuit.**
- **use $V_n = I_n R_n$ on individual components.**

A number of steps are involved in solving a combination circuit: that is, in specifying all resistances, potential differences and currents.

1. Calculate the equivalent resistance across each parallel section.
2. Redraw the circuit replacing the parallel resistors with the equivalent resistance.
3. Calculate current flowing through the resistors in series.
4. Calculate potential difference across the equivalent resistor in series.
5. Calculate potential difference across parallel sections.
6. Calculate the current in parallel sections.
7. Calculate the potential difference across the original series resistors.

WORKED EXAMPLE 6.3

The combination circuit shown in Figure 6.9 is redrawn here.

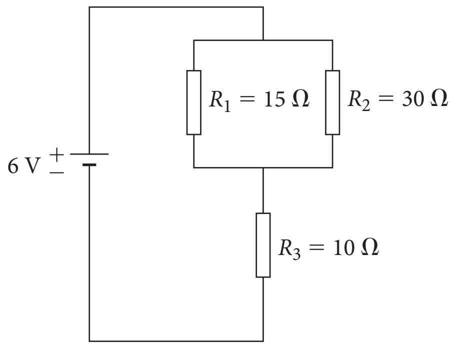
a Calculate the current in R_3. (3 marks)

b Draw one or more circuits to show the total resistance in the circuit as you proceed through the solution. (2 marks)

c Calculate the potential difference across R_3. (1 mark)

d Calculate the potential difference across R_1. (1 mark)

e Calculate the current in R_2. (1 mark)



Answers / Logic

a

Logic: Calculate the equivalent resistance of the parallel section.

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2}$$

$$\frac{1}{R_{eq}} = \frac{1}{15\ \Omega} + \frac{1}{30\ \Omega}$$

Logic: Substitute the correct values. — 1 mark

$$\frac{1}{R_{eq}} = \frac{2}{30\ \Omega} + \frac{1}{30\ \Omega}$$

$$\frac{1}{R_{eq}} = \frac{3}{30\ \Omega}$$

$$R_{eq} = 10\ \Omega$$

Logic: Calculate the answer. — 1 mark

In R_3:

$$V_T = I_T R_T$$

$$I_T = \frac{V_T}{R_T}$$

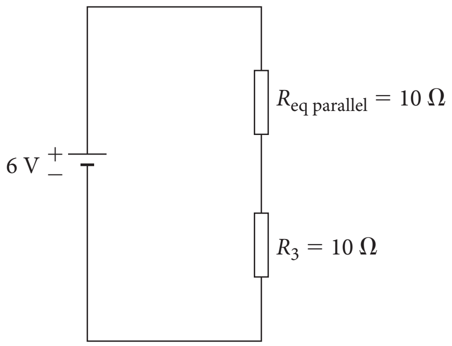
$$R_T = R_3 + R_{eq} = 10\ \Omega + 10\ \Omega$$

$$I_T = \frac{6\ \text{V}}{20\ \Omega}$$

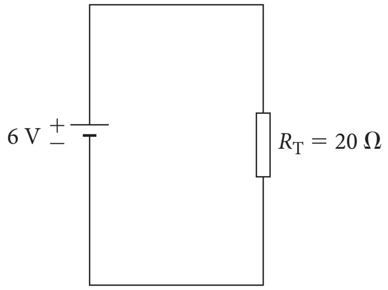
$$I_T = 0.3\ \text{A}$$

Logic: Substitute the correct values and calculate the answer. — 1 mark

b



Re-draw the correct series circuit for the parallel arrangement. 1 mark



Re-draw the correct equivalent circuit for the series arrangement above. 1 mark

c $V_3 = I_3R_3$

$V_3 = 0.3\ \text{A} \times 10\ \Omega$

$V_3 = 3\ \text{V}$

Substitute the correct values and calculate the answer. 1 mark

d $V_T = V_1 + V_3$

$V_1 = V_T - V_3$

$V_1 = 6\ \text{V} - 3\ \text{V}$

$V_1 = 3\ \text{V}$

Substitute the correct values and calculate the answer. 1 mark

e $V_2 = I_2R_2$

$$I_2 = \frac{V_2}{R_2}$$

$$I_2 = \frac{3\ \text{V}}{30\ \Omega}$$

$I_2 = 0.1\ \text{A}$

Substitute the correct values and calculate the answer. 1 mark

Try these yourself

Consider the combination circuit in Figure 6.10.

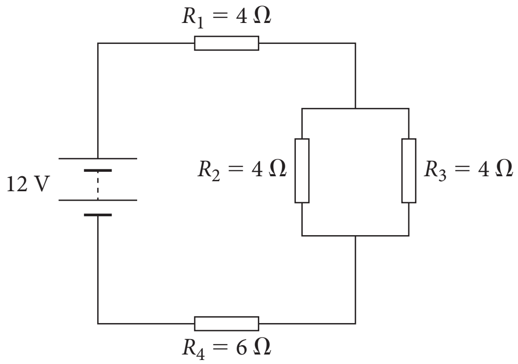
a Calculate the current in R_3. (2 marks)

b Draw one or more circuits to show the total resistance in the circuit as you proceed through the solution. (2 marks)

c Calculate the potential difference across R_3. (2 marks)

d Calculate the potential difference across R_1. (1 mark)

e Calculate the current in R_2. (1 mark)



▲ **Figure 6.10**
A combination circuit

QUESTION SET 6.1

Remembering

1 Define 'ohmic' and 'non-ohmic' resistors.

2 Write Thévenin's theorem in your own words. Explain how Thévenin's theorem is used in circuit analysis.

3 a Define 'solving a circuit'.

b Summarise the steps involved in solving a circuit.

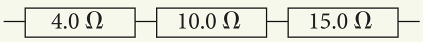
Understanding

4 Which has less equivalent resistance: two 5.0 Ω resistors in series or two 5.0 Ω resistors in parallel? Compare the total current when connected to a battery of *emf* V.

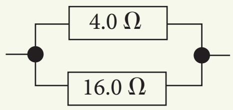
Applying

5 Four different combinations of resistors are below. What is the total or effective resistance of each combination?

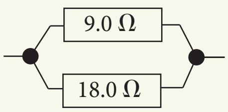
a



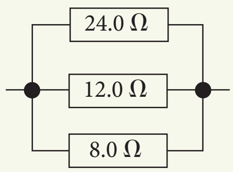
b



c



d



6 Two 12.0 Ω resistors and one 6.0 Ω resistor are connected in series across a 6.0 V battery.

a Draw a circuit diagram to show this arrangement.

b What is the total resistance of the circuit?

c What current flows through the 6.0 Ω resistor?

d What is the potential drop across each resistor?

7 Two conductors each of resistance 12 Ω are connected in parallel and the combination is joined in series with a 6 Ω resistor and placed across a 6 V battery as shown in Figure 6.11.

a Find the equivalent resistance of the circuit.

b What is the total current in the circuit?

c What is the current in each component?

d What is the potential drop across the parallel combination?

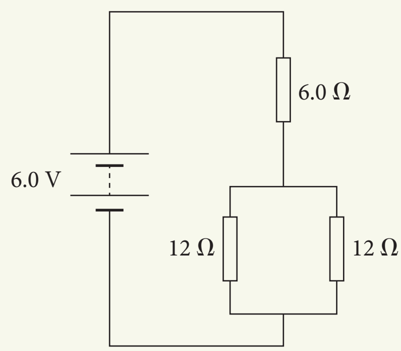


Figure 6.11 ▲

Analysing

8 Currents I_1 and I_2 respectively flow in resistances R_1 and R_2 placed in series across a potential difference *V*. Show that the potential difference is shared in the same ratio as the resistances:

$$\frac{V_1}{V_2} = \frac{R_1}{R_2}$$

9 Currents I_1 and I_2 respectively flow in resistances R_1 and R_2 placed in parallel across a potential difference *V*. Show that the current is shared in inverse ratio to the resistances:

$$\frac{I_1}{I_2} = \frac{R_2}{R_1}$$

10 A 100 Ω resistor is placed in series with a parallel combination comprising a 200 Ω and a 500 Ω resistance. A 20 V DC power supply is connected across this combination circuit. Draw the circuit diagram and solve the circuit.

Reflecting

11 Summarise the main ideas and procedures you have learnt in this section.

Voltage dividers

Sometimes the potential difference available from the power source is more than that required by the circuit. A **voltage divider**, or potential divider, takes advantage of the way a series circuit divides the potential difference between resistors. The simplest voltage divider uses two resistors. The potential drop across both resistors adds to the sum of the supply voltage, V_{in}. In a voltage divider the output voltage, V_{out}, is the potential difference across *one* of the resistors and so is less than V_{in}.

You can use Ohm's law and Thévenin's theorem to calculate V_{out}.

From Thévenin's theorem:

$$R_T = R_1 + R_{out}$$

From Ohm's law (transposed):

$$I = \frac{V_{in}}{R_T}$$

$$\Rightarrow I = \frac{V_{in}}{R_1 + R_{out}}$$

and

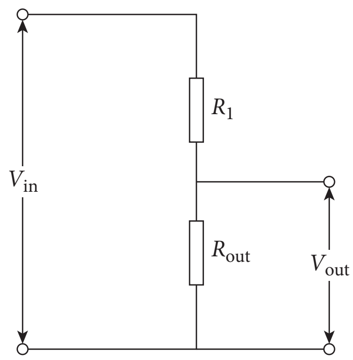
$$I = \frac{V_{out}}{R_{out}}$$

$$\Rightarrow \frac{V_{in}}{R_1 + R_{out}} = \frac{V_{out}}{R_{out}}$$

$$\Rightarrow \frac{V_{out}}{V_{in}} = \frac{R_{out}}{R_1 + R_{out}}$$

You can see that this is a ratio rule. The ratio of V_{out} to V_{in} is equal to the ratio of R_{out} to R_{total}.

We will return to voltage dividers later in the chapter to see how they can be used with electronic components to control circuits.



▲ **Figure 6.12**
A voltage divider is a simple series circuit. Part of the potential difference is used by each resistor.

WORKED EXAMPLE 6.4

In the circuit shown in Figure 6.12, calculate V_{out} if $V_{in} = 6.0\,V$, $R_1 = 1.0\,k\Omega$ and $R_{out} = 2.0\,k\Omega$. (3 marks)

Answer	Logic	
From: $\frac{V_{out}}{V_{in}} = \frac{R_{out}}{R_1 + R_{out}}$	Use the correct formula.	1 mark
$\Rightarrow \frac{V_{out}}{6.0\ V} = \frac{2000\ \Omega}{1000\ \Omega + 2000\ \Omega}$	Substitute the correct values.	1 mark
$\Rightarrow \frac{V_{out}}{6.0\ V} = \frac{2000\ \Omega}{3000\ \Omega}$		
$\Rightarrow \frac{V_{out}}{6.0\ V} = \frac{2}{3}$		
$\Rightarrow V_{out} = \frac{2}{3} \times 6.0\ V$		
$\Rightarrow V_{out} = 4.0\ V$	Calculate the answer.	1 mark

Try these yourself

In the circuit shown in Figure 6.12, calculate:

a V_{out} if $V_{in} = 8\,V$, $R_1 = 6\,k\Omega$ and $R_{out} = 2\,k\Omega$. (3 marks)

b the ratio $V_1 : V_{out}$. Show how you can use this ratio to find V_{out} when $V_{in} = 1.5\,V$, $R_1 = 6\,k\Omega$ and $R_{out} = 3\,k\Omega$. (3 marks)

Voltage and potential difference

Potential difference is measured in units of volt (V), which is the Système Internationale (SI) standard. Sometimes the word voltage is used to describe potential difference. Voltage, technically speaking, is the *measure* of potential difference between two points, which is not the same as potential difference. It is more correct to use potential difference than voltage when analysing circuits. However, as both terms are in common use we need to be careful to ensure we know what we mean.

QUESTION SET 6.2

Understanding

1 It is a common mistake to confuse a physical quantity with the unit used to measure that quantity. How does this relate to potential difference and voltage?

2 What is a voltage divider? Give an example of how a voltage divider could be used.

Applying

3 Draw each of the following circuits and calculate V_{out} across R_2 for the circuit shown in Figure 6.13.

a $V_{in} = 12\,V$, $R_1 = 3\,\Omega$, $R_2 = 9\,\Omega$

b $V_{in} = 24\,V$, $R_1 = 4\,\Omega$, $R_2 = 8\,\Omega$

c $V_{in} = 1.5\,V$, $R_1 = 10\,\Omega$, $R_2 = 5\,\Omega$

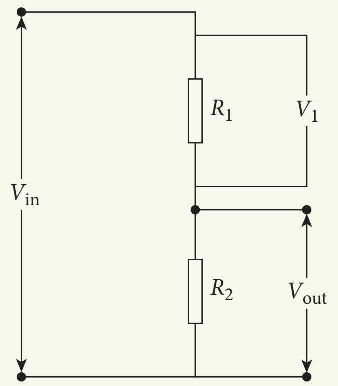


Figure 6.13 ◀

Analysing

4 a Copy and complete the following table, which relates to Figure 6.13.

V_{in}	R_1	R_2	I_1	I_2	V_1	V_2	$I_2 : I_1$	$R_2 : R_1$	$V_2 : V_1$
12	100	10							
12	100	100							
12	100	1000							
12	10 000	1000							
12	10 000	10 000							
12	10 000	100 000							

b Compare the last two columns. Are they the same or different?

c What can you conclude from this table?

5 In Figure 6.13, use Ohm's law and Thévenin's theorem to show that:

$$\frac{V_{out}}{V_1} = \frac{R_{out}}{R_1}$$

Hence, deduce the voltage divider formula:

$$\frac{V_{out}}{V_{in}} = \frac{R_{out}}{R_1 + R_{out}}$$

Reflecting

6 Summarise your understanding of voltage dividers. Make sure you identify circumstances under which you would be best to use one or other of the formulas:

$$\frac{V_{out}}{V_1} = \frac{R_{out}}{R_1} \quad \text{and} \quad \frac{V_{out}}{V_{in}} = \frac{R_{out}}{R_1 + R_{out}}$$

Electronic components

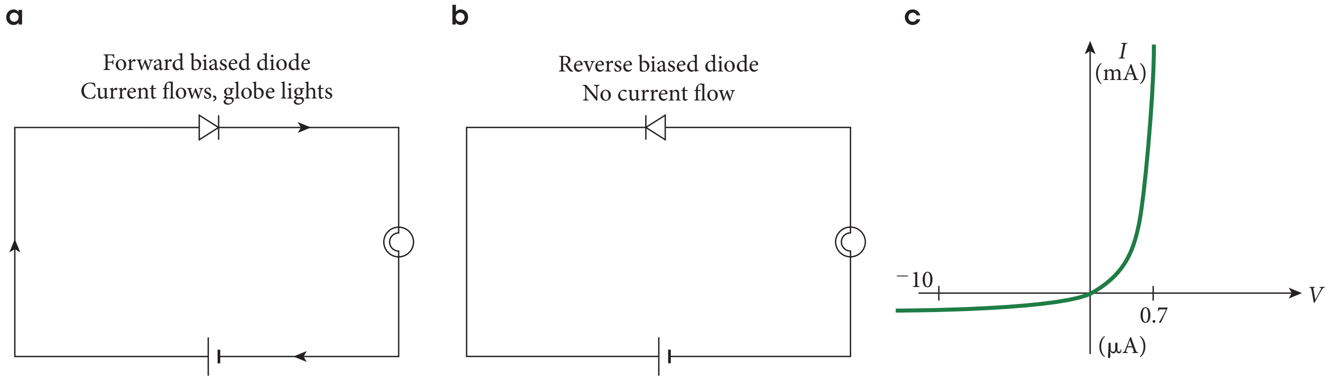
A number of different devices with non-constant resistance are used in electric circuits. These include diodes, photodiodes, thermistors, light-dependent resistors (LDRs) and light-emitting diodes (LEDs).

Diodes

A **diode** is an electronic device made from semiconducting materials. When used in a circuit, a diode allows current to flow in one direction only. The pointed end of the triangle in the circuit symbol shows the direction in which conventional current can flow.

Diodes can be connected to a circuit in two ways. When connected to allow current to flow, the diode is **forward biased** (Figure 6.14(a)). When **reverse biased**, as in Figure 6.14(b), the diode prevents current flow and only an extremely small leakage current flows in it (Figure 6.14(c)). The leakage current is used in photodiodes. If there is a large enough voltage across a diode, the diode is damaged and no current flows.

A forward biased silicon diode requires a potential difference of 0.7 V to allow current to flow. A germanium diode requires 0.3 V across it to allow current to flow.



▲ Figure 6.14
a) A forward biased diode allows current to flow. b) A reverse biased diode restricts the flow of current. c) When forward biased, the diode needs 0.7 V to allow current to flow. When reversed biased, there is only a very small leakage current until the voltage gets large enough to break the diode. Note: This graph has two scales marked on the x axis; there are two scales on the y axis too.

Input and output transducers

Electronic and electro-optical devices are used for many purposes. Voltage dividers are a significant component of these devices For example, all sorts of monitors, from automatic door openers to motion sensors and weather balloon data recorders and transmitters may include voltage dividers in their circuitry. All voltage dividers collect energy from the environment and use it to produce a change in an output.

A device that collects energy from a source and converts that energy to electrical form is an **input transducer**. A microphone collects energy in the form of sound from a person speaking or singing, and converts it to potential differences. A loudspeaker system uses the input potential differences to amplify the signal, which is then broadcast as sound. The loudspeaker is an **output transducer**. An output transducer in one circuit may become an energy source, an input transducer, for another circuit. For example, thermal energy from a fire might be sensed by a thermistor (input transducer). This changes the voltage division in the circuit. The change causes a warning LED (light-emitting diode) to turn on. Then, the warning light from the LED, now acting as an energy source for a new situation, might be used to affect a light-dependent resistor (LDR), such that a sprinkler is turned on. Table 6.2 shows a number of input and output transducers.

Table 6.2 Some transducers used in electronic and photonic circuits

Input transducers	Output transducers
Microphone	Speaker
Heat-dependent resistor, thermistor	Light globes
Light-dependent resistor, LDR	Light-emitting diode, LED
Photodiode	Liquid crystal display, LCD
Switch	Motor

LEDs

iStockphoto/arturoli

Figure 6.15 ▲ LEDs are now used in traffic lights. They last longer and are more efficient than incandescent globes.

An LED is an opto-electronic diode that can emit light. LEDs can produce light of many colours and are quite common. Small, red indicator lights, powerful torches and flashing bicycle lights are some uses. Modern traffic lights are now made of LEDs that appear as arrays of small dots (Figure 6.15). LED light bulbs are also an efficient way of lighting buildings.

The circuit symbol of an LED is shown in Figure 6.17.

Like simple diodes, LEDs require a minimum potential difference before allowing current to flow in them. A typical red diode needs 2.0 V while a blue diode needs 4.0 V. Like other diodes, LEDs have a small leakage current when they are reversed biased. LEDs only need very small voltages to operate. Large voltages across an LED result in large currents in the LED, which can damage it. They need to be connected in series with a resistor to protect them from large currents. Typically, a safe LED current is 20–30 mA. In high-brightness LEDs, this can be 10–20 times higher. When deciding on the size of the protecting resistor, the maximum current and the DC voltage supply need to be considered. For example, in Figure 6.18 a white LED is placed in a circuit with a 9 V supply.

The maximum safe current in the LED is 20 mA. The potential difference across the protecting resistor is (9V − 2V) = 7 V. The resistor and LED are in series, which means the current is the same in both. Application of Ohm's law gives a resistance:

$$R_P = \frac{V}{I} = \frac{7\ \text{V}}{(20 \times 10^{-3})\ \text{A}} = 350\ \Omega$$

Figure 6.16 ▶ Light-emitting diodes come in a range of colours.

Alamy/Schneider Sebastian

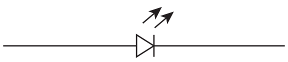
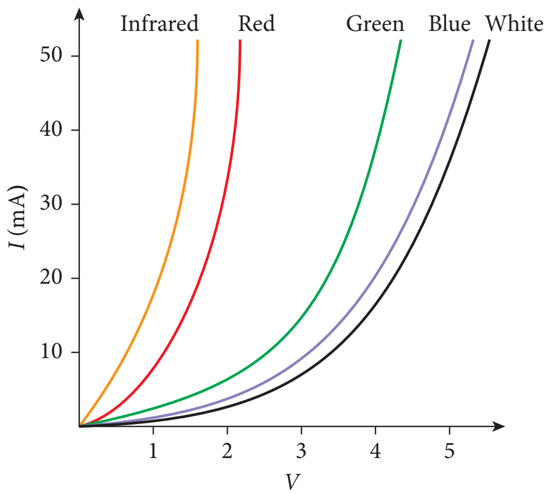


Figure 6.17 ▲ The circuit symbol of an LED

LEDs are better than normal light globes for a number of reasons.

- They have very low energy consumption.
- They have a longer lifetime.
- They respond to changes in potential difference very quickly.
- They can emit UV and infrared light, which is useful for remote controls.

▲ **Figure 6.18**
LED and protecting resistor



▲ **Figure 6.19**
LEDs of different colours require different amounts of potential difference to switch on.

LED PROTECTION

Calculate the resistance needed to protect an LED.

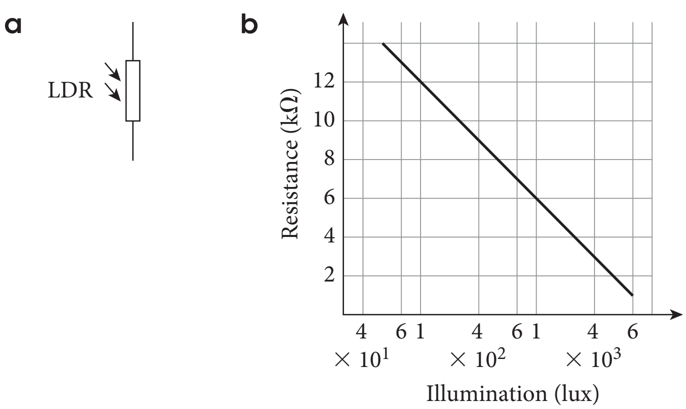
HOW MODERN LIGHT BULBS WORK

Watch this clip to learn about how different types of light bulbs work.

Light-dependent resistors (LDRs)

A **light-dependent resistor (LDR)** is made of a semiconductor material such as cadmium sulfide, whose resistance changes with the intensity of the light. As the intensity of the light falling on the surface of the LDR increases, the resistance of the semiconductor material decreases (Figure 6.20). LDRs can be used to turn on streetlights automatically or in a light meter to measure light levels for photography.

LDRs can take around a millisecond to respond to changes in light, which is relatively slow. Circuits often require very fast response to changes in light, so photodiodes are used.



◀ **Figure 6.20**
a) Circuit symbol and b) characteristic graph of an LDR. Note that the horizontal axis uses a logarithmic scale rather than a linear scale. This allows a very large range of values to fit in a small space.

Scientific literacy: New technology inspires a rethinking of light

… In the United States, lighting consumes more than 20 percent of electric power generated each year; the Energy Department says LEDs can cut consumption by up to 80 percent. LEDs – also called solid-state lighting – are already a $12.5 billion business worldwide, according to analysts at the research firm Strategies Unlimited in Mountain View, Calif. A 2012 McKinsey report estimates LEDs will be an $84 billion business by 2020.

But there is an obstacle or two facing the LED revolutionaries. One is existing modes of lighting: Edison's screw-based socket, the office's fluorescent ceiling tubes, and metal halide or sodium lights in parking lots are not going away anytime soon.

Another hurdle is public wariness after the environmental exhortations of the 2000s, which led to much-disputed federal legislation to phase out the old incandescents, often in favour of compact fluorescent bulbs. In pursuing their goals, advocates played down problems such as the harshness of fluorescent light, and difficulties with dimming the bulbs and dealing with the toxic mercury they contain. Now, some lighting scientists say, both consumers and investors are leery of buying into something they suspect might be substandard.

Another powerful force for continuity is the psychological legacy of light as we know it – from sun to candle to bulb. Isn't the cartoon shorthand for a new idea a glowing bulb over the thinker's head? So some companies are selling the new digital lighting in forms that will fit into the prerevolutionary world, with its sockets and streetlamps – including familiar bulb shapes.

Philips is producing a bulb called Hue that fits into the old sockets and not only dims and brightens, but also changes colors on command. Mr Crawford said that in his lamps division, 25 percent of sales income now comes from LEDs; he expects it to increase to 50 percent in two years. In 2008, that number was close to zero.

One reason adoption will speed up, Mr Crawford believes, is that in recent years, consumers have been asked to compromise on quality to get energy savings. With the latest generation of LEDs, he said, 'the consumer gets the energy savings without compromise.'

Shutterstock.com/Chones

Figure 6.21 ▲
An incandescent, a CFL and an LED globe

The cost barrier is getting lower. Until recently, it typically cost $30 to buy an LED that could replace a 60-watt glass incandescent bulb bought for less than a dollar. Now Cree, a semiconductor manufacturer, has 40-watt and 60-watt LED equivalents for $10 and $14.

James Highgate, an expert on the new technology who runs an annual LED industry conference, sees a transition period ahead 'for the next three to five years, until the eight billion sockets in the U.S. get filled' with LEDs. 'Some people will never change,' he added. 'They'll be in the alleys buying 100-watt incandescents.'

But a new poll done by the lighting company Osram Sylvania showed that fewer consumers were listing 'burned out or broken' as the main reason for switching bulb formats. According to a company news release, '68 percent of Americans say they have switched lighting for increased energy efficiency'.

Energy efficiency is only the beginning, according to experts on the lighting innovations. Take communication between lights. At the University of California, Davis, a bike path illuminated at night with a 'just in time' system has one light node alerting another and another down the line as a bicycle goes by, progressively lighting the rider's way, then dimming back into an energy-saving mode. …

Engineers such as Mr Maxik at Lighting Science are now imagining cities that light their streets as needed, without benefit of lampposts. He has created a fixture that could replace the reflective medians in highways south of the snow belt. Once installed along the road's centerline, they provide as much illumination as streetlamps. The metal and wiring that go into the streetlamp would be unnecessary. …

For the workplace, Osram Sylvania's researchers are looking to control light to improve office productivity. As Lori Brock, director of research and innovation at the company's technology lab in Massachusetts, said: 'It optimizes the illumination for the task you're doing. If you sat at your desk to use the computer, maybe the overhead light would dim, increasing the contrast so you could see better. Other lights could go to an energy-saving hue.' Ideally, productivity increases while energy costs decrease.

As for health applications, the Lighting Research Center of Rensselaer Polytechnic Institute has focused its research on the physiological and psychological impacts of light. This might lead to light fixtures in hotel rooms and elsewhere that enhance sleep or restore the circadian rhythms of jet-lagged travelers. ...

'This is where the promise is,' said Dr Siminovitch of the UC Davis Center. 'The promise is going to be on well-being, wellness, biology – lighting starts doing something for us that is inherently different.'

Barringer, F. (2013) 'New technology inspires a rethinking of light', *New York Times*, 24 April.

Questions

1. What are the two major obstacles facing LED revolutionaries?
2. How does the 'just in time' system save energy?
3. The article explains why it will take time for households to change to LEDs. Think about and list some reasons why it will also be hard for local councils to change to LEDs.
4. Compact fluorescent bulbs have affected people's interaction with light negatively. Compare compact fluorescent bulbs with LEDs to explain how LEDs are better.

WORKED EXAMPLE 6.5

An LDR is placed in a voltage divider circuit as shown in Figure 6.22. The characteristic curve for the LDR is shown in Figure 6.20(b).

What is V_{out} when 1×10^3 lux shines on the LDR? (3 marks)

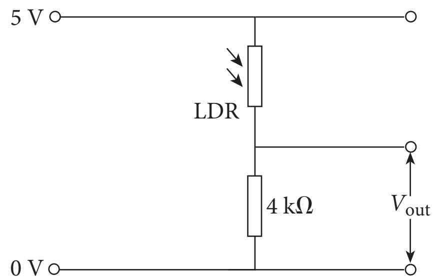


Figure 6.22 A voltage divider circuit with an LDR

Answer	Logic	
Reading from the graph, when there is 1×10^3 lux the resistance is 6 kΩ.	Read the graph correctly.	1 mark
$\frac{V_{out}}{V_{in}} = \frac{R_{out}}{R_1 + R_{out}}$	Use the correct formula.	
$V_{out} = \frac{R_{out}}{R_1 + R_{out}} \times V_{in}$		
$V_{out} = \frac{4 \times 10^3\ \Omega}{4 \times 10^3\ \Omega + 6 \times 10^3\ \Omega} \times 5\ \text{V}$	Substitute the correct values.	2 marks
$V_{out} = 2\ \text{V}$	Calculate the answer.	

Try these yourself

The resistance across the output in Figure 6.22 is changed to 8 kΩ. Find the output voltage when the incident light intensity is:

a 4×10^3 lux. (3 marks)

b 100 lux. (3 marks)

Thermistors

Thermistors, or temperature-dependent resistors, are used to detect changes in temperature. The resulting electrical changes are used to control devices such as heaters and air conditioners. They are non-ohmic resistors (Figure 6.23).

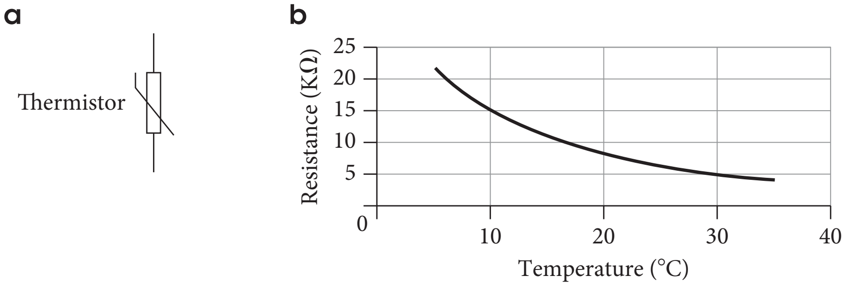


Figure 6.23 ▶ a) Circuit symbol and b) characteristic graph for a thermistor

EXPERIMENT 6.2

TESTING THE EFFECT OF TEMPERATURE ON THE RESISTANCE OF A THERMISTOR

Air conditioners and heaters switch off when the air reaches a set temperature. An electronic circuit with a thermistor can be used to switch the device on and off. In this experiment we will see how the resistance of a thermistor varies with a change in temperature.

Aim

To determine the relationship between the temperature measured by a thermistor and the resistance of the thermistor

Materials

- heat source
- 500 mL beaker
- 200 mL beaker
- water
- glycerol
- thermistor
- digital multimeter
- temperature probe

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
The heat source could cause a burn.	Take care to not touch the heat source while it is hot.

Procedure

1 Record the air temperature.
2 Use the multimeter to measure the resistance of the thermistor and record the result.
3 Put the glycerol in the small beaker with the thermistor and temperature probe.
4 Sit the small beaker in the large beaker and half fill the large beaker with water to make a water bath.
5 Place the large beaker on the heat source, as shown in Figure 6.24.
6 Get your teacher to check that you have set up the equipment correctly.
7 Turn on the heat source.

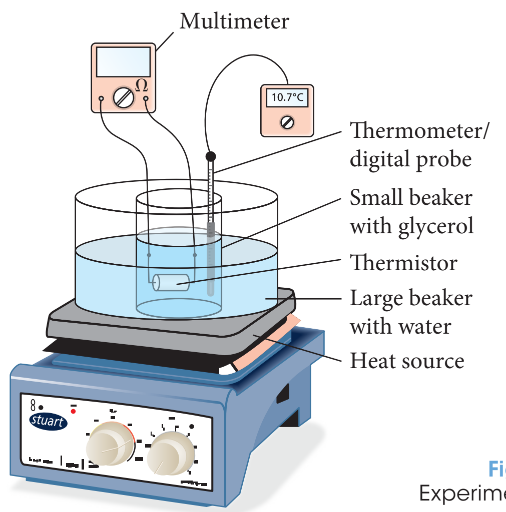


Figure 6.24 ◀ Experimental set-up

8 For every 5 degree increase in temperature record the resistance of the thermistor on a table such as the one below.

9 Estimate the uncertainties in the data points.

Results

Temperature														
Resistance														

Analysis of results

Plot a graph of temperature versus resistance, including uncertainty bars, for your results.

Discussion

1 Did the thermistor's resistance increase or decrease as the temperature increased?

2 What type of relationship is there between resistance and temperature? Give quantitative details, including any equation linking resistance and temperature. How confident are you in the quantitative relation?

3 How did your results compare with those of others in your class?

4 How could you improve the accuracy of this experiment?

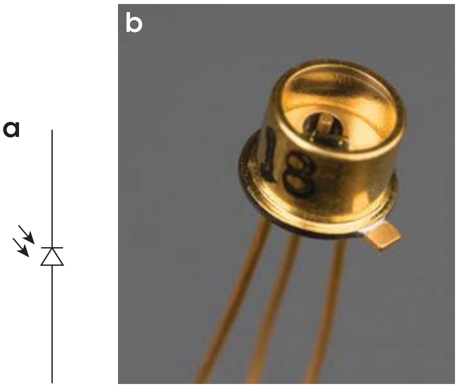
Conclusion

Write a conclusion that relates directly to the aim of this experiment and that uses the data appropriately.

Photodiodes

There are several types of optoelectronic devices that respond electrically to light intensity. They may respond by changing their conductivity, or by a change in the potential difference across or the current through them. The most common optoelectronic devices are photodiodes (see Figure 6.25). Photodiodes can operate in two modes: photovoltaic mode and photoconductive mode.

In **photovoltaic mode**, the photodiode absorbs light energy, for example from the Sun, and converts it into electrical energy. The result is a flow of electrons, or a current, through the photodiode and into an external circuit. This is how solar cells work. A solar cell is effectively a collection of very wide, flat, photodiodes that convert light energy into electrical energy (Figure 6.26).



Alamy/David J. Green - technology

▲ Figure 6.25
a) A photodiode circuit symbol; b) A photodiode

Shutterstock.com/Smileus

◀ Figure 6.26
In photovoltaic mode a photodiode generates a potential difference when it is exposed to light.

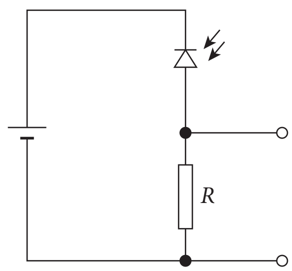


Figure 6.27 ▲ In photoconductive mode the photodiode operates in reverse bias.

In **photoconductive mode** the photodiode requires an external power supply, or source of potential difference. This potential difference is connected so that the photodiode is reverse biased, with the cathode positive relative to the anode as shown in Figure 6.27. Recall that any diode has high conductivity when connected in forward bias and low conductivity when connected in reverse bias. When light is incident on a reverse biased photodiode the conductivity increases, allowing current to flow. Hence a photodiode in photoconductive mode acts as a light-sensitive switch.

Photodiodes are used in photoconductive mode in CD and DVD players to read the signal on the disc. Light from a laser is reflected from the spinning disc to a reverse biased photodiode. The reflected light intensity varies rapidly as the pits in the disc move past the laser beam, and the current flow through the photodiode varies rapidly in response.

In photonic circuits, information is transmitted by light signals rather than, or as well as, electrical signals. The components need to react quickly to changes in light levels. Photodiodes react more rapidly than LDRs to light changes, typically a few nanoseconds for a typical photodiode, compared with 10 ms or more for an LDR.

The amount of current is directly proportional to the amount of light that the cell receives, as shown in Figure 6.28. The photocurrent versus reverse bias voltage for a photodiode is shown in Figure 6.29. For reverse bias voltages up to 10 V (−10 V on the *x* axis), the magnitude of the photocurrent is essentially constant.

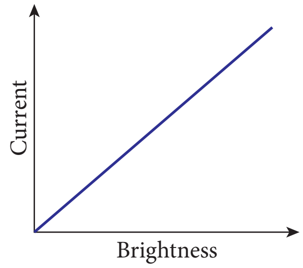


Figure 6.28 ▲ When a photodiode is reverse biased in a circuit, the amount of current is directly proportional to the amount of light that it receives.

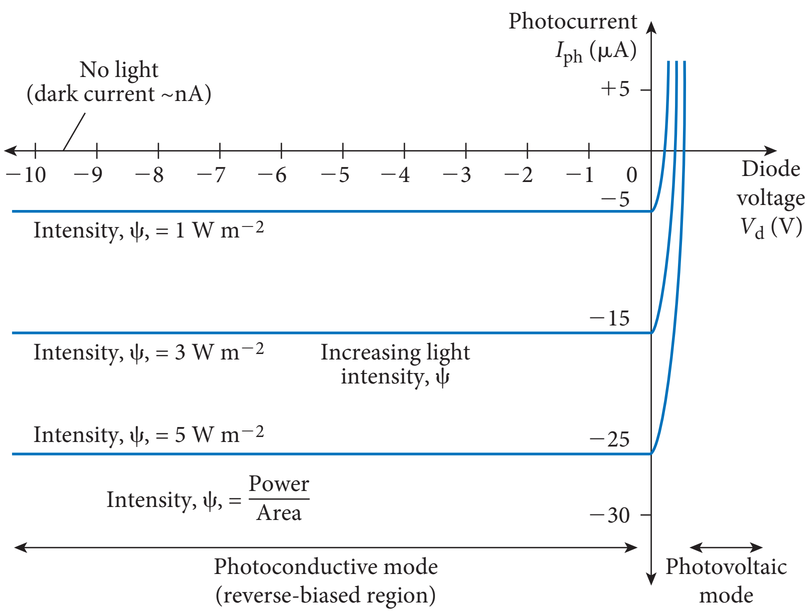


Figure 6.29 ▶ Reverse biased photodiode in photoconductive mode: the photocurrent is proportional to the light intensity.

WORKED EXAMPLE 6.6

A photodiode with characteristics similar to Figure 6.29 is illuminated by 4.0 W m^{-2} of light. It is working with a reverse bias potential difference of 6.0 V (Figure 6.30).

a Construct a data table and graph to find the quantitative relationship between intensity and photocurrent. Hence, find the photocurrent when the intensity is 4.0 W m^{-2}. (3 marks)

b Determine the value of the resistance, *R*. (2 marks)

c In what direction does current flow in this circuit? (1 mark)

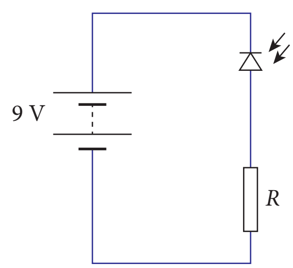


Figure 6.30 ◀ Resistor and photodiode in series across a 9.0 V battery

Answers

Logic

a From the graph in Figure 6.29, the data table is:

ψ (W m^{-2})	I (μA)
1	5
3	15
5	25

Put correct data in table. 1 mark

The graph of this data is shown below.

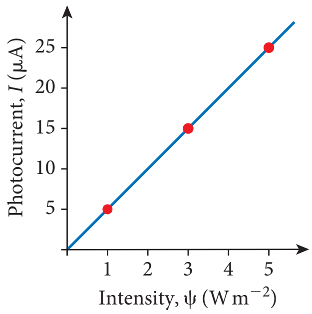


Figure 6.31 Photocurrent versus intensity for a photodiode

Draw correct graph. 1 mark

By inspection, the equation is $I = 5 \times 10^{-6}\ \psi$. Hence, the photocurrent when the intensity is 4.0 W m^{-2} is

$I = 5 \times 10^{-6} \times 4.0$ A

$\Rightarrow I = 20 \times 10^{-6}$ A

$\Rightarrow I = 20\ \mu$A

Calculate the answer. 1 mark

b The potential difference across the photodiode = 6.0 V.

Thus, the potential difference across R = 9.0 − 6.0 = 3.0 V.

The photodiode and resistor are in series, so the current in both is the same = 20 μA.

$$\Rightarrow R = \frac{V}{I} = \frac{3.0\text{ V}}{20 \times 10^{-6}\text{ A}}$$

Substitute the correct values. 1 mark

$\Rightarrow R = 1.5 \times 10^{5}\ \Omega$

$\Rightarrow R = 150\ \text{k}\Omega$

Calculate the answer. 1 mark

c The current flows clockwise in the circuit despite the reverse bias of the photodiode.

Give the correct direction. 1 mark

Try these yourself

A photodiode with characteristics similar to Figure 6.29 is placed across a 12.0 V battery. It is illuminated by 1.5 W m^{-2} of light. The reverse bias potential difference is 8.0 V.

a What is the photocurrent in the photodiode? (1 mark)

b Determine the value of the resistance, R. (2 marks)

c What is the potential difference across R, when the intensity is changed to 2.0 W m^{-2}? (1 mark)

Applications of electronic circuits

Figure 6.32 ▲
Climate controls in museum display boxes protect delicate objects from changes in atmospheric conditions

Museums

Museums and art galleries store and exhibit objects that are of scientific, historical, artistic or cultural interest. Often these objects are old and could deteriorate quickly if they are in the wrong climatic conditions. Too much light can cause fading, especially with paper and textiles. Changes in temperature cause swelling and contraction of the different materials in the objects. High humidity (more than 65%) increases the likelihood of mould and fungi, but low humidity (below 25%) means that objects can lose structurally important water. Typically museums try to maintain a constant temperature of (20 ± 2)°C and a humidity of approximately (45 ± 5)%.

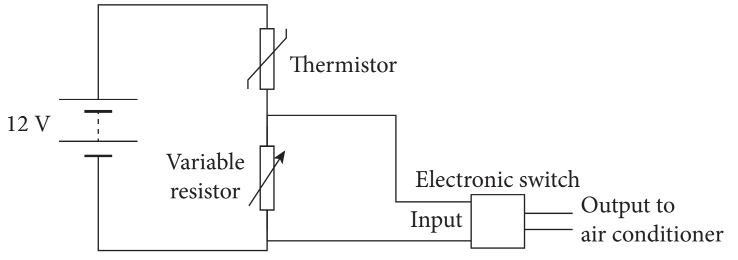


Figure 6.33 ▲
A temperature-control circuit

Delicate objects are kept in sealed glass cabinets because it is easier to maintain a constant climate in a smaller space. Sensors, such as thermistors, take temperature measurements. They are connected to circuits that control temperature and humidity.

Weather balloons

Weather balloons are released into the atmosphere to measure air pressure, temperature, humidity and wind speed. The helium-filled balloons are about 1.5 metres in diameter. Connected to the balloon is a small box called a radiosonde, which contains sensors to take the measurements. Some sensors that might be on board are a thermistor (temperature), a barometer (pressure), a hygrometer (humidity) and an anemometer (wind speed). The sensors are connected in voltage divider circuits so that, as the quantity changes, the output voltage changes. The information from the changed voltages in the sensors is used to transmit a radiofrequency signal back to base so that it can be analysed by meteorologists.

Figure 6.34 ▲
Weather balloons measure atmospheric conditions using electronic circuits.

Remote controls

When you turn up the volume on your TV using the remote control, you are using an opto-electronic control circuit. An LED in the remote control sends a signal to a photodiode in a voltage divider circuit in your TV. The LED emits infrared light to transmit the signal. When the photodiode receives the LED signal the output voltage in the voltage divider changes, which is then used to increase the volume.

Telecommunications

In modern communication systems, monochromatic (single frequency) light waves are used to transmit information. A fixed, high-frequency carrier wave, shown in Figure 6.35, is added to the low-frequency signal to produce a high-frequency signal. This is achieved by changing the amplitude of the light wave according to the signal. This process is called intensity **modulation**. Hence, the combined signal is called optical intensity-modulated light.

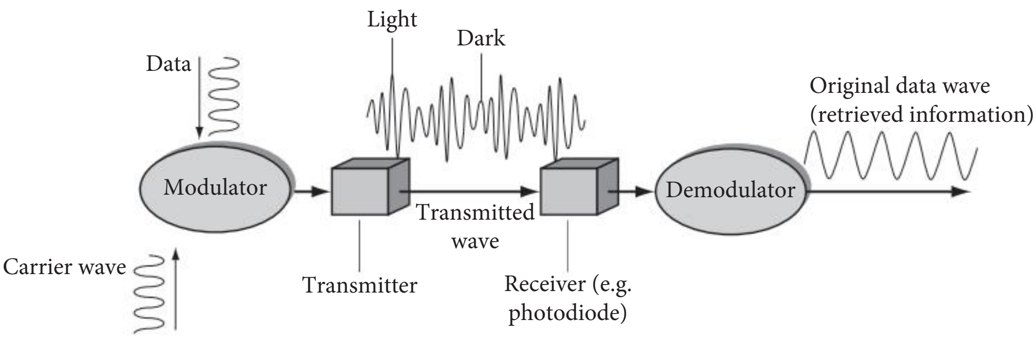


Figure 6.35
A fixed, high-frequency carrier wave

At the receiving end, a photodiode or phototransistor detects the intensity variation of light and identifies the signal. This process is called **demodulation**.

Over short distances, light signals can carry information in air. An example is a remote control for a television or stereo. These use infrared light, as this frequency is transmitted without much loss through the air. However, for efficient transmission over longer distances, optical fibres are used.

In normal metal wires, changes in pitch are carried by changes in the frequency of the electrical signal. This method can still be used with light and optical fibres but the laser or LED used would need to be able to change frequencies. The advantage of using intensity-modulated light is that the signal is carried by a single colour or frequency. This also means that another carrier frequency can carry other information at the same time, meaning that a single optical fibre can carry more than one signal at the same time.

QUESTION SET 6.3

Remembering

1 Define 'input transducer' and 'output transducer'. Give two examples of each.

2 What is the difference between a photodiode in photoconductive mode and one in photovoltaic mode?

Understanding

3 From the graph in Figure 6.36, estimate the resistance of the LDR when the illumination is 400 lux.

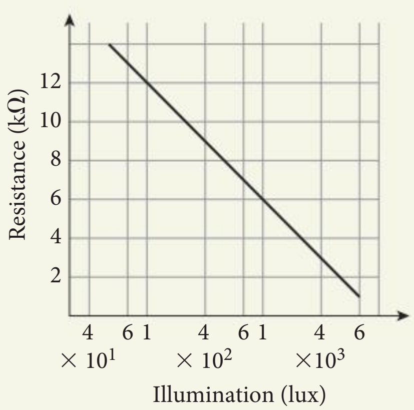
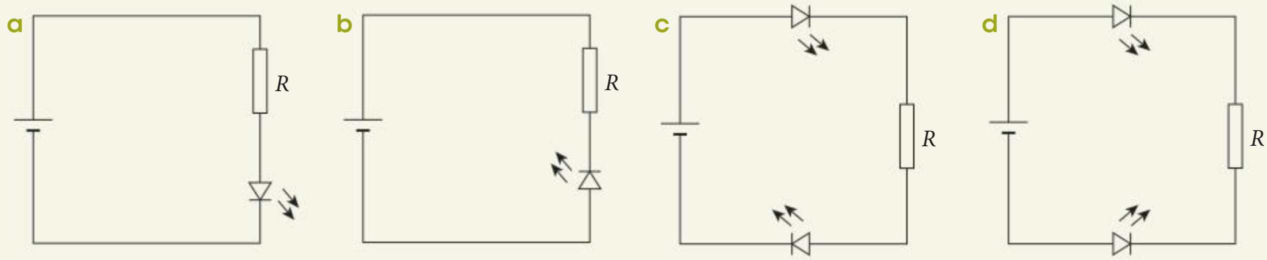


Figure 6.36

4 Both axes in Figure 6.14(c) have different positive scales and negative scales. Explain why.

5 In which circuit(s) below will the LED be lit? Why?



Applying

6 Figure 6.37(a) shows a voltage divider circuit. The characteristic curve of the thermistor is shown in Figure 6.37(b).

a What is the resistance of the thermistor when the temperature is 10°C?

b What is the output voltage when the temperature is 10°C?

c The temperature starts to fall. What happens to the value of the output voltage? Explain your answer.

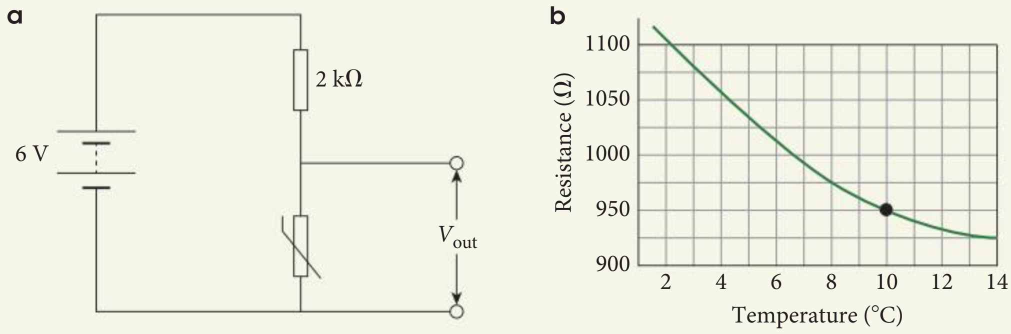


Figure 6.37

Analysing

7 The photodiode with characteristics shown in Figure 6.29 is placed in circuit Figure 6.38(a). It is then subjected to a fluctuating light source whose intensity varies with time as shown in Figure 6.38(b). Sketch the potential difference across the resistor versus time graph.

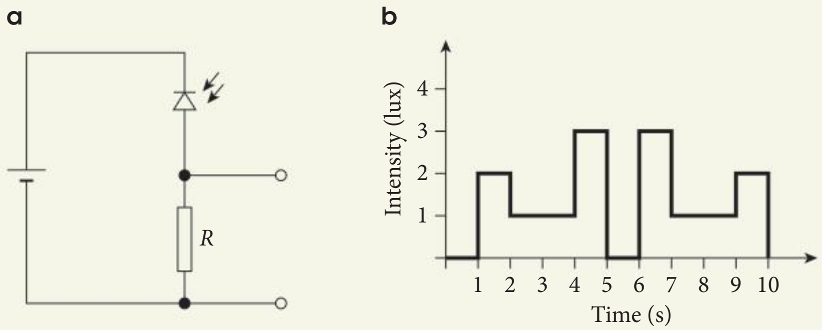


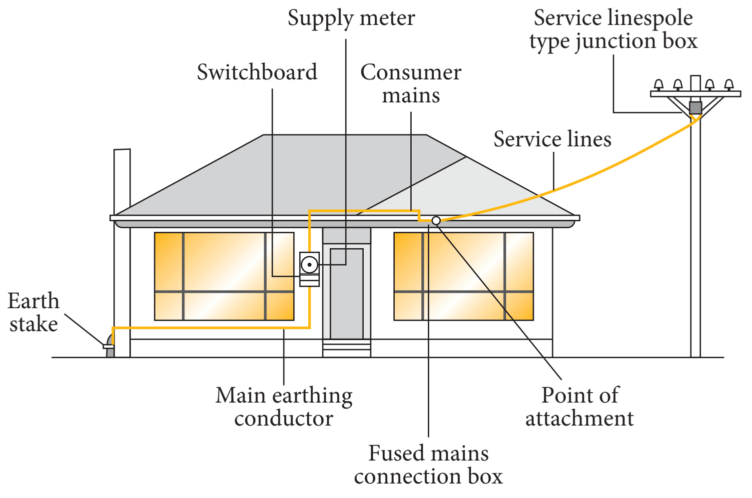
Figure 6.38
a) Circuit diagram;
b) Graph of intensity versus time for a fluctuating light source

Reflecting

8 List the electronic devices that you have used so far today. How do electronics affect our everyday lives?

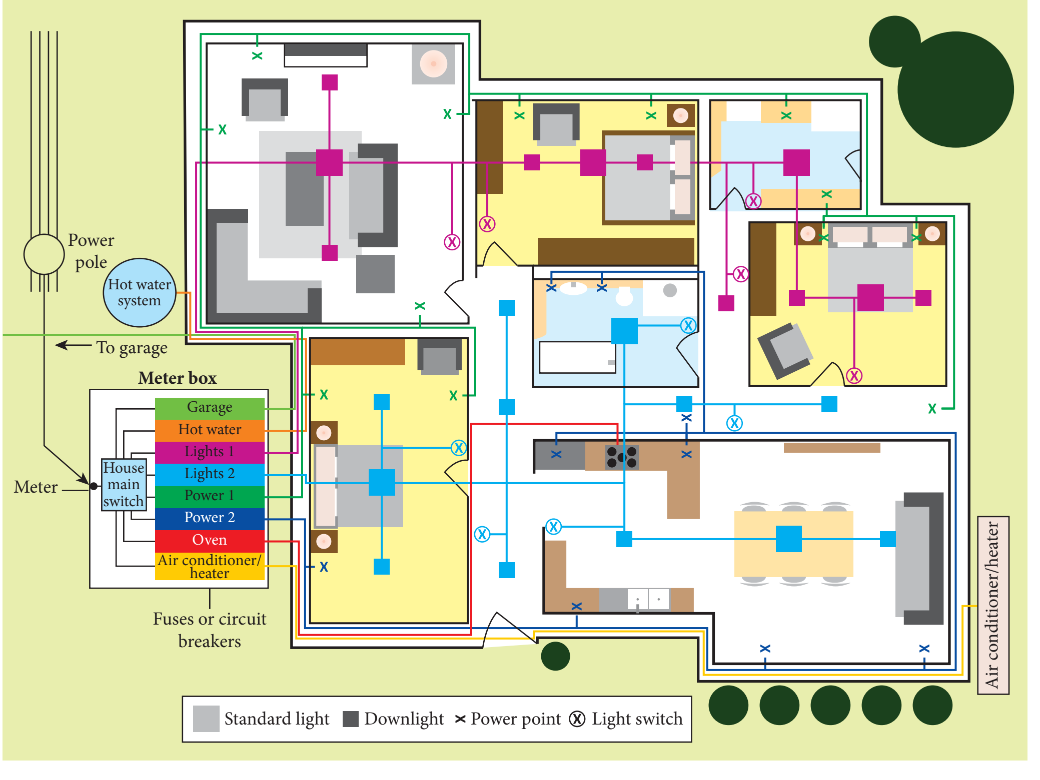
Household electricity

The electricity supplied to houses in Australia is 240 V, 50 Hz single phase. There are two wires in the cable to the house – one is the **active** or phase wire and the other is the **neutral** wire. These come to a mains connection box that contains a **fuse** in the active wire. From there, the cable goes to the fuse box or switchboard. The active lead is attached through an energy meter to the main switch and from there to a number of circuit breakers. The neutral wire is connected to a metal bar called the neutral bar, which is connected back to earth via a metal stake. The neutral wire is also earthed at the substation (Figure 6.39).



◀ **Figure 6.39** Typical household electrical supply system

A typical electrical circuit diagram for a house can be seen in Figure 6.40. Each colour represents a circuit breaker within the switchboard. There are actually a number of separate circuits for lights, power points and larger appliances. The power points and lights are connected in parallel. If they were not, then all the switches would have to be turned on to operate any power point or light!



◀ **Figure 6.40** A typical circuit diagram for a house

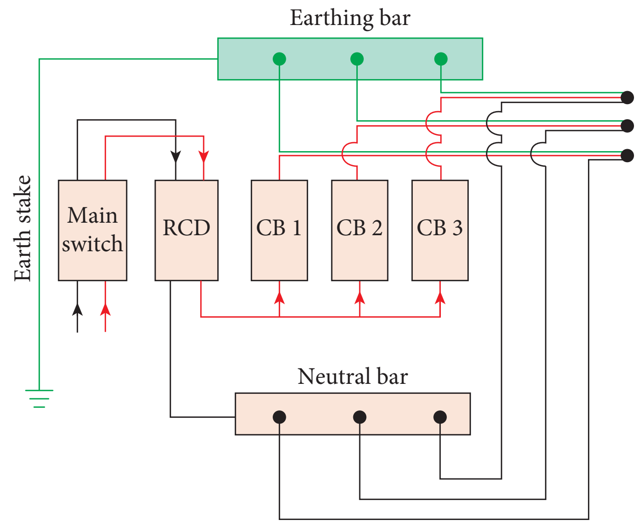


Figure 6.41 ▲
Switchboard diagram with three circuit breakers

Figure 6.41 shows a small switchboard with the main switch and three circuit breakers.

Power points

Appliances are connected to the power point through a three-pin plug, or a two-pin plug if the appliance is double insulated. The active wire and the neutral wire connect to the circuit, which means that power can be supplied to the appliance. The earth wire is connected to the metal frame of the appliance so that if a fault develops it will trip the circuit breaker.

Power points in homes accept three pins from a plug. The wiring is colour-coded to ensure the correct pin is connected to the active, neutral or earth wire (Figure 6.43).

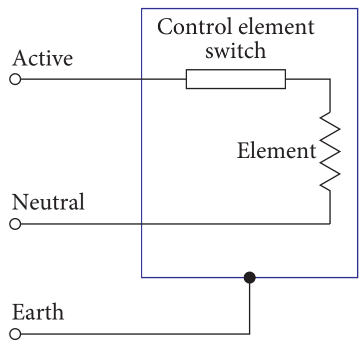


Figure 6.42 ▲
Circuit for a simple toaster or heater showing connections to the active, neutral and earth wires

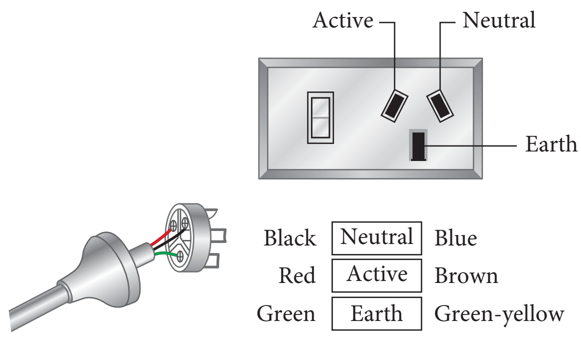


Figure 6.43 ▲
Colour code for a three-pin plug in a three-point power point

Short circuits

The most spectacular type of overloading occurs in a **short circuit**. A short circuit occurs when there is a current pathway between active and neutral, so that the current is able to bypass the appliance and use a far less resistive pathway. The effect is that, for much smaller resistance, the current becomes much greater than it is in normal use. This brings into play the fuse or other safety, tripping device.

A short circuit can occur, for example, when the insulation of two wires (active and neutral) in a cord wears through and the two bare wires touch each other. Alternatively, a bare active wire may touch a metal component on an appliance that is earthed. If you touch a 'live' appliance, you may provide the conduction path to earth. The result may be fatal.

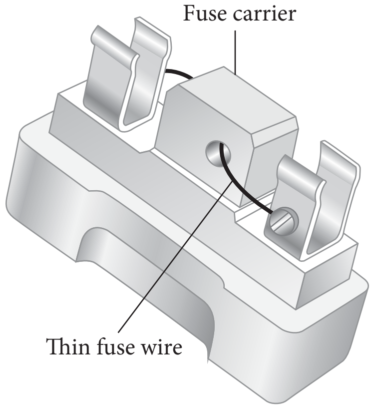
Accidents can be prevented by careful and regular attention to all electrical appliances and leads. It is essential that worn cords are replaced, that earth wires are properly connected and that care is taken at all times.

Safety devices

It is important to protect circuits from power surges that can damage the wiring and appliances in the home. Circuit breakers and fuses are designed to break the circuit before damage is done. They do this in different ways.

Fuse

A fuse in the active circuit wire is a short piece of wire selected so that it will melt or 'blow' when the current through it exceeds a certain value (Figure 6.44). The fuse protects the circuit from an oversupply of current. Fuses for a domestic power supply are typically rated at 30 A for appliance circuits and 15 A for lighting circuits.



▲ **Figure 6.44**
Ceramic fuse carrier and fuse wire

Circuit breaker

A **circuit breaker** is an electromechanical device that automatically opens a switch if overload occurs. It contains an electromagnet that becomes more powerful as the current increases. When the current reaches a certain value, the electromagnet is powerful enough to force apart a contact and so break the circuit. It can do this in a very short time – less than it takes for a piece of wire to burn through.

Residual current device (RCD)

A **residual current device (RCD),** or earth-leakage protection, provides an extra safety feature to prevent electrocution. If a person touches a live wire and electricity flows through their body, then there will be an imbalance in the amount of current flowing through the active and neutral wires. If the imbalance reaches 50 mA then the RCD will break the circuit within milliseconds. The RCD is typically located in the main switchboard, as seen in Figure 6.41.

Double insulation

Some equipment and appliances are not connected to the earth wire. Instead, they have two layers of insulation. The layers of **double insulation** are:

- the functional insulation, which is around the live parts such as insulation on wires.
- the protective insulation, which is the external plastic covering the device.

Failure of either part will not result in a current path to the outside of the appliance, and failure of both is almost impossible without complete mechanical fracture. These devices are not connected to the earth wire and only have a two-pin plug.

Safety

If a fuse blows or a circuit breaker operates, the cause of the fault should always be found and rectified. Never be tempted to use a fuse with a higher rating. If you use a higher rating fuse, the electrical wires in your house effectively become the fuse. If they overheat, they can set fire to your ceiling, where most of the wires can be found. The first you will know of a fire will be when the roof is alight and starts to fall in on top of you!

Only a licensed electrician should carry out wiring alterations, repairs and the installation of appliances and fittings that are permanently connected to the house wiring. Unqualified people should never make their own extension cords. Power cords, extension leads and fittings that show signs of wear should always be promptly repaired by a qualified person, or replaced.

Simple safety rules

1 Do not attempt electrical repairs unless qualified.
2 Never use electricity in a wet or damp situation, because water is a good conductor of electricity.
3 Replace all defective appliances and frayed cords.
4 Do not meddle with appliances that are switched on.
5 Keep clear of power lines.
6 Have sound knowledge of first-aid procedures.

QUESTION SET 6.4

Remember

1 What are the names and functions of the three main wires used in a house wiring circuit?

Understanding

2 How does a short circuit occur?

3 Why is it more convenient to have several separate circuits in your house rather than a single circuit?

Analysing

4 What are the similarities and differences between a circuit breaker and a fuse?

Reflecting

5 Most work places require all electrical cords to be tested and tagged once a year. This involves a visual inspection of the cord and measuring the earth and insulation resistance. Do you think this is a good use of time and money?

CHAPTER SUMMARY

- Every circuit can be reduced to the simplest form of one energy source and one energy sink.
- Resistors in series circuits: $R_T = R_1 + R_2 + \cdots + R_n$
- Resistors in parallel circuits: $\frac{1}{R_T} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_n}$
- In combination circuits use $V_T = I_T R_T$ on the whole circuit once you know the total resistance. Use $V_n = I_n R_n$ on individual components.
- A series circuit can be used as a voltage divider.

$$V_{out} = \frac{R_{out}}{R_1 + R_{out}} V_{in}$$

$$\frac{V_{out}}{V_1} = \frac{R_{out}}{R_1}$$

- Diodes only allow current to flow in one direction. They require a minimum potential difference to operate.
- Photodiodes can convert light energy into electrical energy.
  - Photovoltaic mode: a potential difference is produced from light energy.
  - Photoconductive mode: reverse biased current is proportional to incident light intensity.
- LEDs efficiently convert electrical energy into light.
- The resistance of an LDR varies with changes in light intensity.
- The resistance of a thermistor varies with changes in temperature.
- Electronic components can be used in voltage dividers to create useful circuits.

CHAPTER GLOSSARY

active connection to the 240 V household supply

circuit breaker electromechanical switch that trips when there is an overload; safety protection against overload

combination circuit elements connected in series and in parallel groups

demodulation process of separating the signal from the modulated wave

diode a semiconductor device that allows current to flow in one direction only

double insulation functional and protective layers of insulation that enhance safety devices

emf electromotive force; source of potential energy per charge

forward biased connected to allow current flow

fuse temperature-dependent wire that melts if an overload occurs; safety protection against overload

input transducer device that takes energy from the environment and converts it for use in an electric circuit

light-dependent resistor (LDR) device in which resistance depends on illumination

modulation process of combining the signal with the carrier wave

neutral zero potential in a household wiring system

output transducer device that uses energy from an electric circuit to convert to energy to be sent into the environment

photoconductive mode mode in which photodiodes exposed to light conduct current in reverse biased direction

photovoltaic mode mode in which photodiodes exposed to light generate potential differences

residual current device (RCD) earth leakage protection device; safety protection against overload

reverse biased connected to prevent current flow

short circuit connection between two points that allows current to flow with negligible resistance

thermistor temperature-dependent resistor; used to detect changes in temperature

Thévenin's theorem all circuits can be reduced to a single source and a single load

voltage divider device used to vary voltage at the output depending on a control resistor; also called a potential divider

CHAPTER REVIEW QUESTIONS

Remembering

1 a What effect does resistance have on current in a circuit?
  b Define 'ohmic resistance'.
  c Define 'non-ohmic resistance'.

2 Define electrical circuits that are in:
  a series.
  b parallel.

3 Write down the rules for calculating the equivalent resistance of any number, *n*, of resistors in:
  a series.
  b parallel.

4 Draw a combination series and parallel circuit. Write down the rules for solving such circuits.

Understanding

5 Is an LED an ohmic or non-ohmic resistor? Illustrate your answer with a sketch graph.

6 Why are photodiodes used in photonics circuits instead of LDRs?

7 What is the difference between an input transducer and an output transducer? Explain how input and output transducers can be used to operate a TV remote control.

8 Describe how a photodiode can be used in a voltage divider to respond to changes to the incident light.

Applying

9 Three $10\,\Omega$ and one $20\,\Omega$ resistors are joined in series across a 10.0 V supply. What is the potential difference across each?

10 Household globes, each 100 W, are connected in a parallel array, then connected in series to a 7.5 A fuse of negligible resistance. When 7.5 A or more flows in the fuse it melts and breaks the circuit. The power supply can be considered to be 240 V DC.

a What is the current in one 100 W globe?

b If three, 100 W globes are turned on, what is the current in the fuse?

c How many globes can be turned on in the array before the fuse melts?

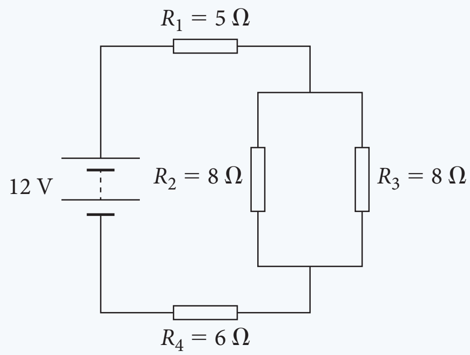


Figure 6.45 ▲

11 A combination circuit is shown in Figure 6.45.

a Calculate the effective resistance of the circuit.

b Draw the equivalent Thévenin circuit.

c Find the current in each resistor.

i R_1

ii R_2

iii R_3

iv R_4

d Calculate the potential difference across each resistor.

i R_1

ii R_2

iii R_3

iv R_4

12 Emily makes a voltage divider using two resistors and a 12 V battery as shown in Figure 6.46. V_{out} = 8 V and R = 10 kΩ.

a What is the resistance of R_{out}?

b What is the current in R?

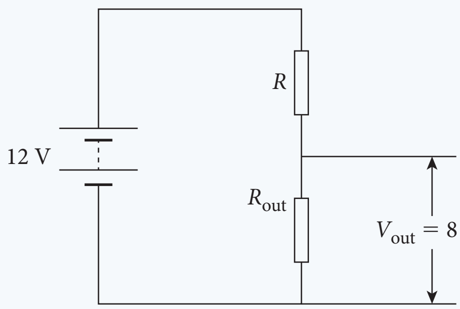


Figure 6.46 ▲

Analysing

13 Two 1 kΩ resistors are placed in series with a 9 V battery, an ammeter and an unknown resistor, R. The ammeter reads 1.5 mA.

a What is the resistance of the resistor, R? Show your working.

b What would the ammeter read if it were placed in parallel with one of the 1 kΩ resistors? Give a detailed explanation.

14 A slow cooker uses a thermistor to keep the cooking food at a constant low temperature. The variation of resistance with temperature for the thermistor is shown in Figure 6.47(a). It is connected in series with a $1.2 \times 10^4\ \Omega$ resistor and a 9.0 V power supply as shown in Figure 6.47(b).

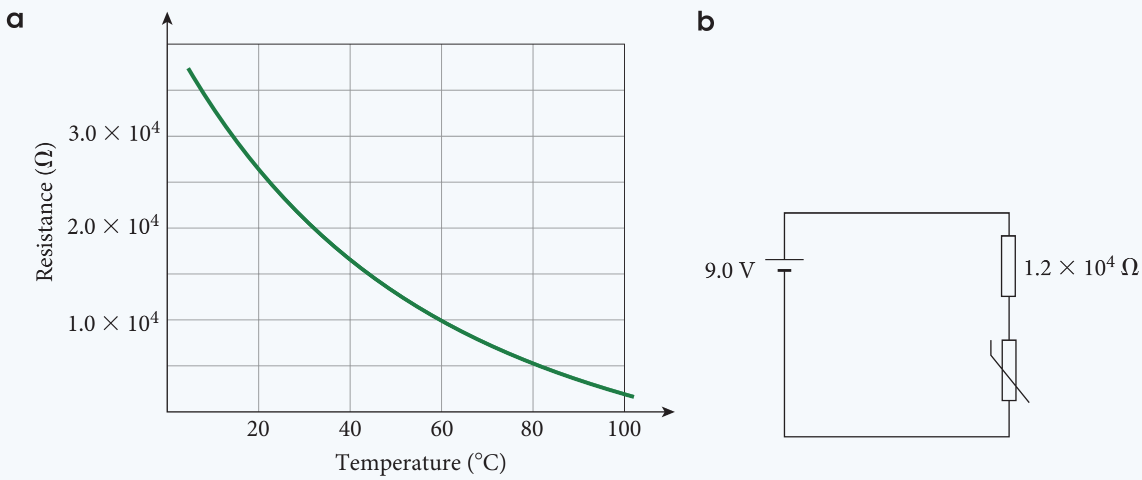


Figure 6.47 ◀

a What is the resistance of the thermistor when the temperature is 60°C?

b What is the potential difference across the thermistor when the temperature is 80°C?

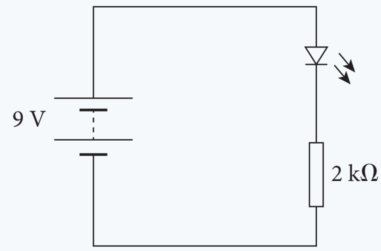
c What happens to the potential difference across the thermistor as the temperature drops? Give two examples of temperatures to show your reasoning.

15 The circuit in Figure 6.48 depicts a simple LED and resistor in a series circuit. The applied voltage is 9 V.

a What is the potential drop across the resistor if the LED requires a voltage drop of 1.8 V?

b What is the current flowing in the LED?

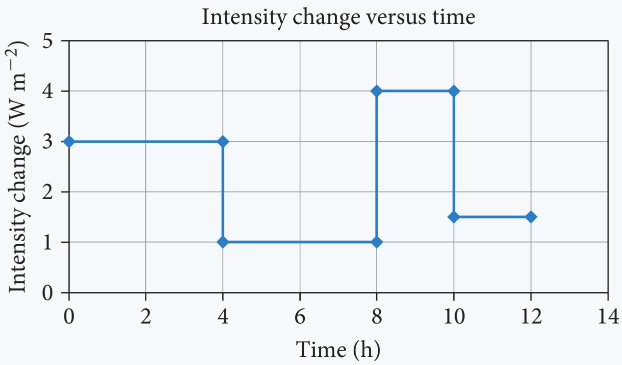
c What is the purpose of the 2 kΩ resistor?



▲ Figure 6.48

16 For a wide range of reverse biased potential differences, V, and incident light intensities, ψ, a photodiode current, I, is modelled by the equation $I = 3.5 \times 10^{-6}\psi$.

The photodiode is connected in series with a 15 V battery and a 100 kΩ resistor. The output voltage across the 100 kΩ resistor is to be used. The photodiode is placed in a room illuminated by a constant light intensity of 3.0 W m^{-2}. Over a 12-hour period, the intensity in the room is increased and decreased. The changes are shown in Figure 6.49.



◀ Figure 6.49

a Draw the circuit diagram for this arrangement.

b For the photocurrent versus reverse-bias voltage, draw characteristic lines for reverse bias voltages from 0 to 10 V when the intensity in W m^{-2} is 2.0, 4.0, 6.0 and 8.0.

c What is the current in, and output voltage across, the resistor when the light intensity is 9.0 W m^{-2}?

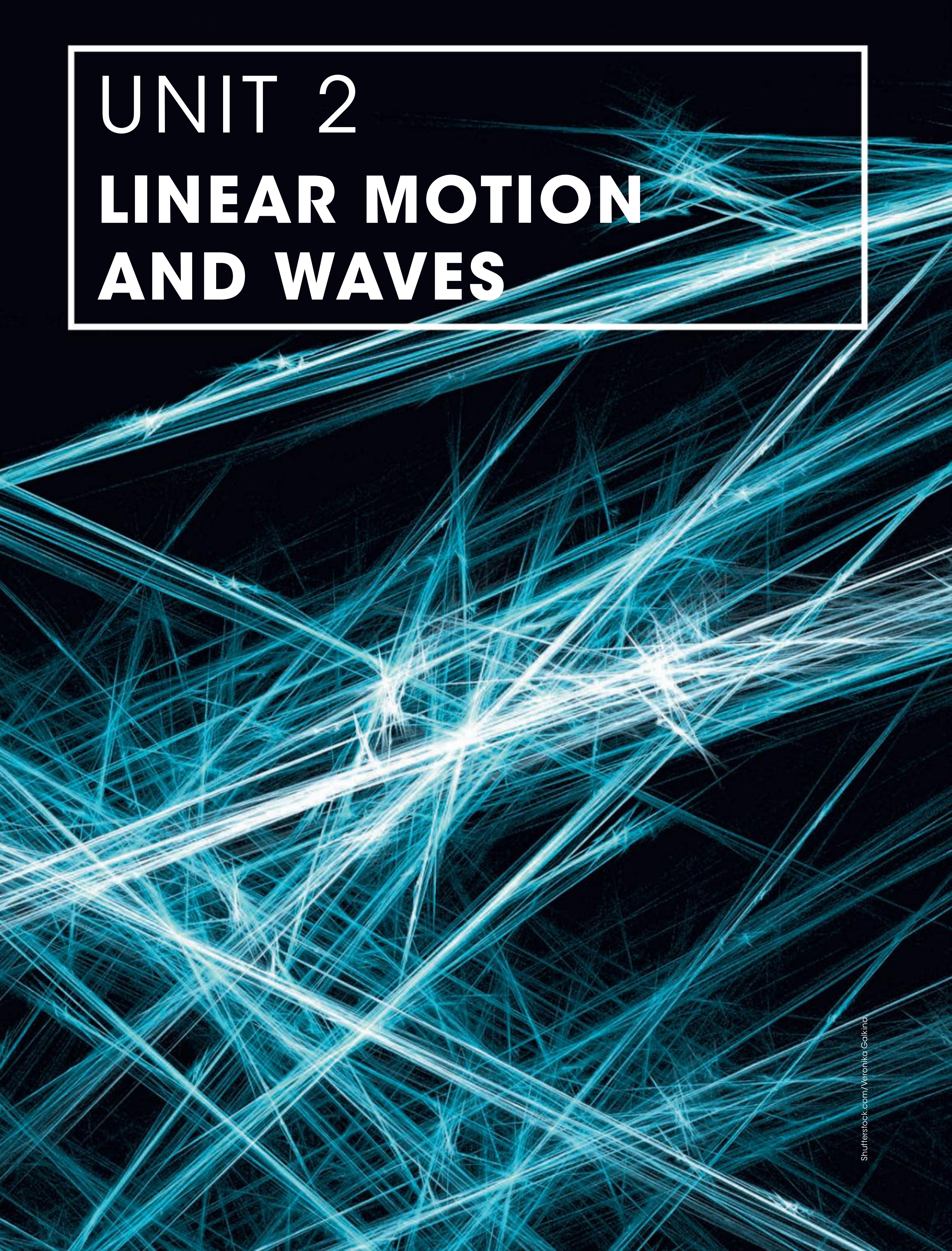
d Sketch the output voltage over the 12 hours for which the light intensity changed.

Reflecting

17 How has your understanding of the way electronic devices control systems changed?

18 Summarise your notes to include all relevant definitions, laws and equations. Consider how to integrate your learnings from Chapter 5 into this summary.

UNIT 2
LINEAR MOTION AND WAVES



Shutterstock.com/Denis Babenko

CHAPTER 7 MOTION

By the end of this chapter you will have covered the following material.

Science Understanding

- Uniformly accelerated motion is described in terms of relationships between measurable scalar and vector quantities, including displacement, speed, velocity and acceleration (ACSPH060)
- Representations, including graphs and vectors, and/or equations of motion, can be used qualitatively and quantitatively to describe and predict linear motion (ACSPH061)
- Vertical motion is analysed by assuming the acceleration due to gravity is constant near Earth's surface (ACSPH062)

Introduction

It was once thought that moving objects would slow down and stop when they got tired. Their motion was something that was natural to them. Aristotle (384 BCE–322 BCE) believed that heavier objects naturally fell faster than light objects. He said an efficient cause made an arrow travel in straight lines. It moved off a straight line by reason of violent force. Aristotle developed a sophisticated view about motion that was used extensively for nearly 2000 years. However, not every observation could be explained using these ideas. As a result, Aristotle's ideas came under scrutiny. Galileo and Newton changed the way motion is explained by using the results of many observations and investigations to develop new ideas about motion. These ideas are still used today to explain all but the most unusual forms of motion. Einstein's theories are required to explain the behaviour of objects and particles that are moving at speeds close to the speed of light.

These theories and laws are models for the behaviour of the motion of objects, which are themselves modelled as point masses. We use graphs and equations as equivalent and complementary representations of the motion of these model particles.

Movement along a straight line

Motion can occur in straight lines, on a surface (two dimensions) or in three dimensions. The simplest type of motion of an object is along a straight line, forwards or backwards, left to right or side to side. When we describe motion, a sensible **frame of reference** is chosen. For example, the stationary Earth is a sensible frame of reference for an aeroplane; for Venus, the Sun can be considered to be stationary. The chosen frame of reference can be used even if it is in motion, such as Earth moving around the Sun at about $30\,km\,s^{-1}$.

A car has many moving parts. However, in our studies we will model a car or other object as having its entire mass concentrated at one point, its **centre of mass**. This will simplify calculations about the motion of the car even though a real car is not a single point. Such models are often used in physics. Simplifications of real situations make it possible to analyse motion of an extended object, such as a car, even though it is obviously not a single point. More detailed analysis of the different parts of the car such as that undertaken by car designers might be required to understand car safety and how vehicles deform in crashes.

Distance and displacement

In order to describe the motion of an object, information about the position of the object and the time are needed. Position cannot change instantaneously. An object cannot move from one place to another without taking time to do so. It takes time for the position of anything to change.

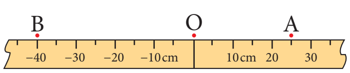


Figure 7.1 ▲ By defining a starting point or origin, opposite directions can be designated positive and negative values.

The position of an object is measured as the distance away from some reference point such as the starting place, or **origin**, O. In Figure 7.1, point A is a distance of 25 cm from O while point B is 40 cm from O. A position can either be to the right or to the left of the origin O in this example. A position to the right is given a positive value while a position to the left is negative.

In a method similar to a number line in mathematics, point A is said to be +25 cm from O while point B is −40 cm from O.

Distance

Distance, *d*, is the actual length between two points. It has no direction. For example, the distance between points A and B in Figure 7.1 is simply stated as 65 cm.

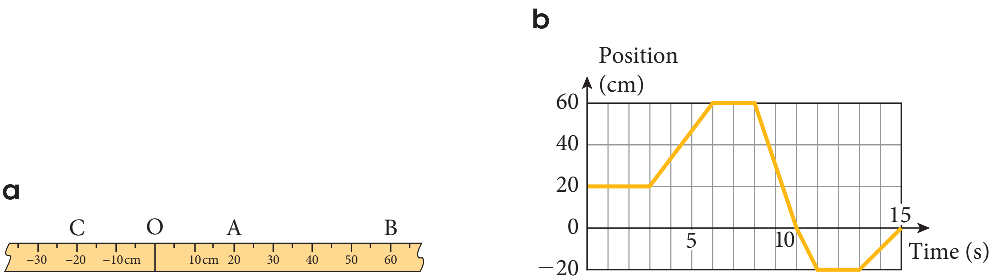
Displacement

Displacement is the position of an object relative to the origin, or starting point. In straight line motion, displacement must be given a positive or negative value to show which side of the origin the object is positioned.

In a marathon in which the finish line is the same as the starting line, a runner will have a final displacement of zero at the end of the race. The distance run will be 42.2 km, but the displacement measures how far you are from the start or origin. During the race, the runner's displacement may have been many kilometres, as this is the straight line distance from the start to the runner. We will use *d* as the symbol for distance and *s* for displacement. Both displacement and distance have the same SI unit, metre, *m*.

WORKED EXAMPLE 7.1

A snail starts at a position 20 cm from the origin and then moves to a new position 60 cm further away before going back past the origin to a position 20 cm on the other side of the origin. It finally ends up at the origin. The positions are shown in Figure 7.2(a).



◀ **Figure 7.2**
a) A snail moves along a straight line from +20 cm to +60 cm then to −20 cm and ends up at the origin, O.
b) Cartesian graph of the motion of a snail along a straight line

The position–time graph of the motion is shown in Figure 7.2(b).

1 What was the distance travelled by the snail's centre of mass on its trip? (2 marks)

2 What was the final displacement of the snail's centre of mass relative to the:
   a starting point? (1 mark)
   b origin? (1 mark)

Answers	Logic	
1 $d = 40\text{ cm} + [60\text{ cm} - (-20\text{ cm})] + (20\text{ cm} - 0\text{ cm})$	Read data from graph correctly.	1 mark
$= 40\text{ cm} + 80\text{ cm} + 20\text{ cm}$		
$= 140\text{ cm}$	Calculate the correct answer.	1 mark
2 a Displacement $d_f - d_i = 0\text{ cm} - 20\text{ cm}$	Calculate the correct answer.	1 mark
$s = -20\text{ cm}$ (i.e. 20 cm to the left of the starting point)		
b The snail is at the origin after the movement. This means its displacement is zero relative to the origin.	Interpret the answer correctly.	1 mark

Try these yourself

Tia lives on a long straight road 250 m from the shops and 400 m from her friend James, who lives in the opposite direction. Tia walks to the shops then goes to see James.

a Before she goes out, what is Tia's position relative to: (2 marks)
   i the shops?
   ii James?

b After the walk: (2 marks)
   i what distance did Tia walk?
   ii what was Tia's displacement relative to home?

When an object moves on a two-dimensional plane from one position to another, its change in position can be found geometrically. Let us consider a simple example. An orienteer is at a position that is 600 m, N30°W from the start or origin. In Figure 7.3(a) the displacement of the orienteer is shown by a yellow arrow that has been drawn to scale. The arrow represents both magnitude, 600 m, and direction, N30°W. A quantity that has both magnitude and direction is a **vector** quantity. Notice that the orienteerer has moved both north and west. The northwards **component**, d_N, and the westward component, d_W, can be found by scale drawing or trigonometry. By trigonometric ratios:

$$d_N = 600 \cos 30° \text{ m } (520 \text{ m})$$
$$d_W = 600 \sin 30° \text{ m } (300 \text{ m})$$

The orienteerer then moves to a new position that is 400 m N60°E from the start or origin. This is shown in Figure 7.3(b). The components in the northerly and easterly directions of this vector are:

$$d_N = 400 \cos 60° \text{ m } (200 \text{ m})$$
$$d_E = 400 \sin 60° \text{ m } (346 \text{ m})$$

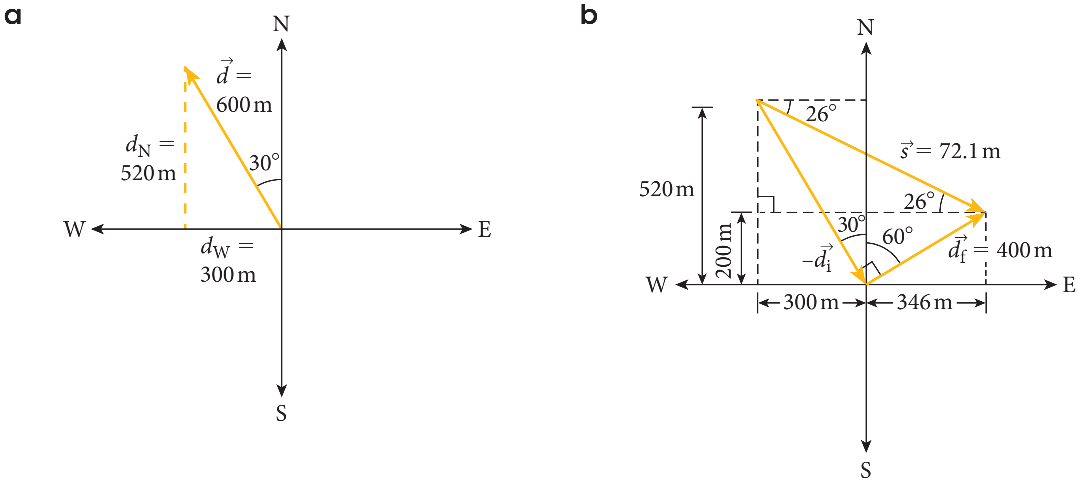


Figure 7.3 ▼ Positions of an orienteerer

When vectors are involved, the arrow goes over the whole symbol (e.g. $\vec{s}$).

The difference between two vector quantities has the same form as any other difference; that is, the final quantity minus the initial quantity:

$$\Delta\vec{d} = \vec{d}_f - \vec{d}_i$$

We give this the symbol $\vec{s}$:

$$\vec{s} = \vec{d}_f - \vec{d}_i$$

In order to arrive at the final position, you have, effectively, to take off the initial position and add the new position:

$$\vec{s} = \left(-\vec{d}_i\right) + \vec{d}_f$$

Figure 7.4 shows how this is done geometrically. The initial vector is reversed. Vectors are added head to tail, so now we have the vectors being added. You can easily see that the resultant vector, $\vec{s}$, has a southward component and an eastward component:

$$s_N = 200 \text{ m} - 520 \text{ m} = -320 \text{ m} = 320 \text{ m, S}$$
$$s_E = 346 \text{ m} + 300 \text{ m} = 646 \text{ m, E}$$

In this case, the magnitude of $\vec{s}$ can be found using Pythagoras' Theorem, because we note the right angle in the (geometric) vector sum triangle. The magnitude of the change in displacement is:

$$s = \sqrt{(600 \text{ m})^2 + (400 \text{ m})^2}$$
$$\Rightarrow s = 721.1 \text{ m}$$

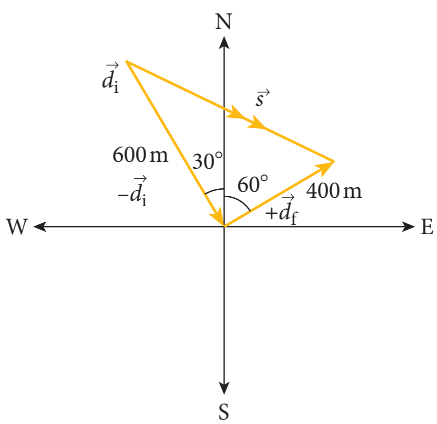


Figure 7.4 ▲ Geometric representation of orienteerer's position

In general, we cannot do this easily. What if the person was at 600 m, N51°W? Then the angles would not form a right triangle.

One way to do this is with a protractor and ruler. Draw a scale drawing, which is what has been done in Figure 7.3. Then, you just measure the required length and convert it with the scale. You can find the angle with a protractor. The measured angle is shown in Figure 7.3(b).

The final position of the orienteerer can also be found with a little more trigonometry:

$$\tan\theta = \frac{s_S}{s_E}$$

$$\Rightarrow \tan\theta = \frac{320\,\text{m}}{646\,\text{m}}$$

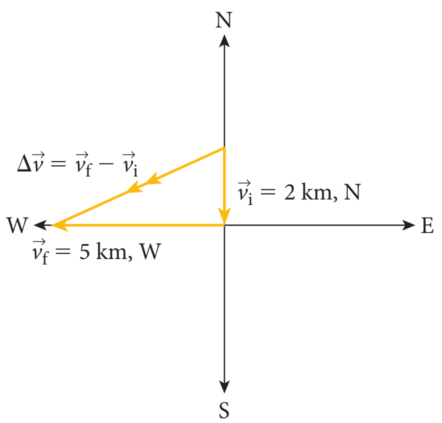
$$\Rightarrow \quad \theta = 26°$$

We are now in a position to specify the change of displacement completely:

$$\vec{s} = 721.1\,\text{m},\ \text{E}\,26°\text{S}$$

WORKED EXAMPLE 7.2

A rogainer runs from a position that is 2.0 km N of the start. Later, the rogainer is at a position 5.0 km W of the start. What is the change in the position of the rogainer? (5 marks)

Answer	Logic	Marks
 	Show the correct vector subtraction.	2 marks
Magnitude: $s = \sqrt{(2.0\,\text{km})^2 + (5.0\,\text{km})^2}$ $\Rightarrow s = 5.4\,\text{km}$	Calculate the magnitude.	1 mark
Angle: $\tan\theta = \frac{2.0\,\text{km}}{5.0\,\text{km}}$ $\Rightarrow \theta = 22°$	Calculate the correct angle.	1 mark
Specification: $s = 5.4\,\text{km},\ \text{W}22°\text{S}$	Give the correct specification.	1 mark

Try these yourself

What is the change of position when an object goes from:

a 30 m S to 40 m E? (5 marks)

b 45 km N30°W to 45 km S50°E? (5 marks)

Time

The position of an object is measured at a particular time t. The SI unit for time measures is the second, s.

Movement is the change in position as time changes. Any **time interval** can be shown as Δt, where:

$$\Delta t = t_2 - t_1 \text{ (Unit: s)}$$

The symbol Δt cannot be separated into Δ and t as it represents the amount of time that has passed from one measurement to the next. Δt is a single symbol. It is often replaced by t but retains the meaning Δt: $t = t_2 - t_1$

Speed and velocity

In everyday speech, the terms **speed** and **velocity** may be used to mean the same thing. In physics, speed relates to the distance covered in a time interval; velocity specifically relates to the *change* in displacement during a time interval.

The change in distance, called the **distance interval**, is given the symbol s.

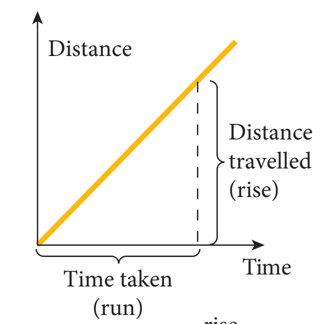
$$s = d_2 - d_1 \text{ (Unit: m)}$$

Speed

Speed, v, is a measure of how fast something is travelling. It is the rate at which distance changes as time changes. Speed can be measured by observing the distance travelled and dividing by the observed time taken. Speed may be reported as **instantaneous**, such as the reading on the speedometer of a car, or it may be **average**, where a whole trip is measured. Whether it is instantaneous or average, speed is always measured by measuring distance intervals and time intervals and performing the calculation:

$$v = \frac{s}{t}$$

The SI unit of speed and velocity is the unit of $\frac{\text{distance travelled (m)}}{\text{time taken (s)}}$, that is, m/s or m s^{-1}.



Gradient = slope = $\frac{\text{rise}}{\text{run}}$

$= \frac{\text{distance travelled}}{\text{time taken}}$

= speed

Figure 7.5 ▲ Speed is given by the gradient (rise/run) of the distance–time graph.

In Figure 7.5, the gradient of the position–time graph gives the speed. For constant speed, the gradient is the same at all points. The graph is a straight line. In this situation, the average speed would be the same as the instantaneous speed throughout the journey. If the speed were to vary, the gradient of the line would change too.

You cannot measure the speed of a car by taking a single photo. You can measure its speed over a very short time interval from a video of its motion, using the frame just before and just after the instant in time in question. Therefore, reference to instantaneous speed is really reference to an average speed that is the same value as all the speeds in the time interval measured. That is why the time interval needs to be very small.

> 'Instantaneous speed' is the rate at which distance is covered over a time interval that is so brief as to be negligible. All the speeds in the time interval do not vary significantly.

Velocity

The change in displacement, the **displacement interval**, is given the symbol $\vec{s}$. The arrow above the symbol is used to signify that this quantity has a direction associated with it. In straight-line motion, this is either positive or negative, so we do not need the arrow.

Velocity is the rate of change of displacement as time changes:

$$\Delta \vec{v} = \frac{\vec{s}}{t} \text{ (Unit: m/s, m s}^{-1}\text{)}$$

The arrow over the velocity symbol again signifies that this quantity has a direction associated with it. Velocity is therefore the speed with a direction attached.

Suppose that, in the previous case, where $\vec{s}$ = 721.1 m, E26°S, the change occurred over a 30 s time interval. The average velocity would then be:

$$\vec{v}_{\text{ave}} = \frac{\vec{s}}{\Delta t}$$

$$\Rightarrow \vec{v}_{\text{ave}} = \frac{721.1\,\text{m}}{30\,\text{s}}, \text{E } 26°\text{S}$$

$$\vec{s} = 24\,\text{m s}^{-1}, \text{E } 26°\text{S}$$

- Speed is the magnitude of the velocity.
- The direction of velocity is either positive or negative for straight line motion.
- In general, the direction of the velocity is the same as the direction of the *change* of displacement.
- The magnitude of the velocity is found geometrically by dividing the time interval over which the change takes place into the geometric difference between initial and final displacements.

Movement at constant speed

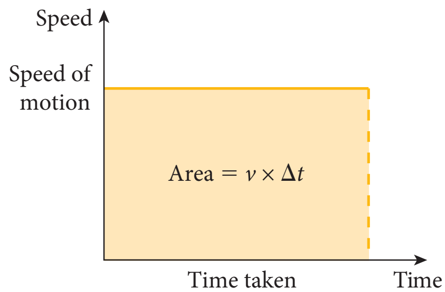
If you walk steadily taking two paces each second, you are walking at 2 paces per second (i.e. 2 paces s^{-1}). In 10 seconds you would have moved 20 paces. Using metres as the unit of length allows any motion to be described in the same way anywhere.

Calculations to find the distance moved can be made:

$$v = \frac{\Delta d}{\Delta t} = \frac{s}{t}$$

$$\Rightarrow s = vt$$

A graph of v versus t shows that the area under the line equals s, the distance travelled. This is shown in Figure 7.6.



Area shaded
= speed of motion × time taken
= distance travelled

▼ **Figure 7.6** A speed versus time graph of an object moving at constant speed

EXPERIMENT 7.1

THE SPEEDS OF COMMON OBJECTS

Aim

To measure the speeds of some human-propelled objects

Materials

- stopwatch
- measuring tape
- various bats, racquets and balls
- optional: video camera or motion data logger

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
A ball hit with a bat could hit a person or break a window.	Perform the experiment in an open space such as a school oval and keep bystanders well back.

Procedure

1. Measure out an appropriate length (e.g. 20 m) between two lines on the school oval or in a clear area.
2. By either throwing or hitting a ball with a bat or racquet, reproduce the actions of several different ball sports (e.g. cricket, tennis, hockey, golf) that propels a ball from one line past the other.

3 Measure the time it takes for the ball to travel the designated distance. For this, use a stopwatch or you may be able to video the motion and use the clock on the video. You might also have access to a motion data logger that is able to measure speed directly.

4 Repeat step 3 for the same sport several times.

5 Repeat steps 3 and 4 for a different sport.

Results

Use the results of your measurements to find the average speed of the ball for each sport. Include an estimate of the uncertainty in each value.

Analysis of results

1 Compare the average speed of the ball for each sport with the way in which that sport is played; for example, a tennis serve compared with a cricket bowler; a golf ball with a ball thrown by a fielder.

2 Convert the results from $m\,s^{-1}$ into $km\,h^{-1}$.

Discussion

Discuss the difficulties encountered during this experiment and suggest ways in which the data collection could be made to be more accurate.

FIRST MAN-HOLE COVER IN SPACE

Read more about this unusual event.

WOW

The first human-made object to be launched into space

Observers speculate that a metal plate being used as a man-hole cover for an underground nuclear test was the first artificial object to make it into space. Nuclear testing in the Nevada desert in 1957 was being filmed with a high-speed movie camera. The man-hole cover was caught in only a single frame, leading to speculation that it was moving with more than the required escape velocity of $11\,km\,s^{-1}$. However, it is more likely that the first human-made object to reach space (about 100 km according to NASA) was a V2 rocket launched in the 1940s.

Movement when speed changes instantaneously

When an object changes speed in a very short amount of time, the change in speed is often regarded as being instantaneous. Truly instantaneous speed change is impossible for the same reasons that truly instantaneous distance change is impossible. In Worked example 7.3, we will consider that the runner changes her speed almost instantaneously (at 15 minutes in Figure 7.7). In fact, some time interval must pass for the speed to change.

WORKED EXAMPLE 7.3

An athlete runs at $20\,km\,h^{-1}$ for 15 minutes. She then gets a stitch and slows to $15\,km\,h^{-1}$ for the next hour. The speed–time graph is shown in Figure 7.7.

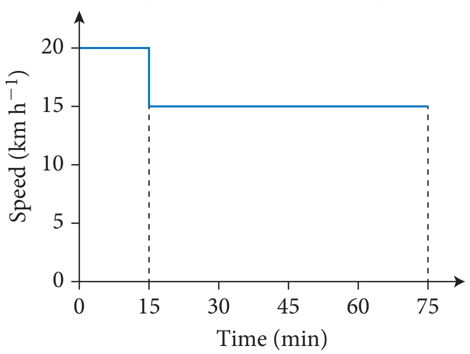


Figure 7.7

1 How far does the runner travel in the first 15 minutes? (2 marks)

2 How far does she run in total? (2 marks)

Answers	Logic	
1 Area $= s_1 = v_1\Delta t_1 = 20\,\text{km h}^{-1} \times 0.25\,\text{h}$	Identify area under the graph.	1 mark
$= 5\,\text{km}$	Calculate the correct answer.	1 mark
2 Area $= s_1 + s_2 = v_1\Delta t_1 + v_2\Delta t_2 = 5\,\text{km} + 15\,\text{km h}^{-1} \times 1\,\text{h}$	Use area under the graph.	1 mark
$= 20\,\text{km}$	Calculate the correct answer.	1 mark

Try this yourself

If the runner continues at $15\,\text{km h}^{-1}$ for another hour, how far will she have run in total? (2 marks)

Average speed, v_{ave}

When you are travelling through the city in a car your speed changes all the time. If you have travelled 20 km in half an hour you would say that your average speed, v_{av}, was $40\,\text{km h}^{-1}$ for that trip. It does not mean that you were always moving at $40\,\text{km h}^{-1}$; however, if you had been travelling at a constant $40\,\text{km h}^{-1}$, the same trip would have taken the same time. Average speed is the one single speed that would enable the car to cover the same distance in the same time interval:

$$v_{ave} = \frac{\Delta d}{\Delta t} = \frac{s}{t} \quad \text{(Unit: m/s or m s}^{-1}\text{)}$$

WORKED EXAMPLE 7.4

What was the average speed of the athlete in Worked example 7.3? (2 marks)

Answer	Logic	
The average speed is found by dividing the total distance travelled by the total time interval taken for the entire event:	Use the correct formula.	1 mark
	Substitute the correct values to find the correct answer and units.	1 mark

$s = 20\,\text{km}$, $\Delta t = 1.25$ hours, so:

$$v_{ave} = \frac{20\,\text{km}}{1.25\,\text{h}} = 16\,\text{km h}^{-1}$$

This answer can be checked by calculating the distance travelled if moving at a constant $16\,\text{km h}^{-1}$ for 1.25 hours: $s = 16 \times 1.25\,\text{km} = 20\,\text{km}$.

N.B. It is WRONG to add the two different speeds and divide by two to find the average.

In this example, this would give $n_{ave} = \frac{20\,\text{km} + 15\,\text{km}}{2\,\text{h}} = 17.5\,\text{km h}^{-1}$.

At this average speed the athlete would cover $17.5\,\text{km h}^{-1} \times 1.25\,\text{h} = 21.9\,\text{km}$ – too far!

Try this yourself

A car trip involves travelling at $60\,\text{km h}^{-1}$ for 1 hour, and then at $100\,\text{km h}^{-1}$ for the next 30 minutes. Find the average speed of the car for the entire trip. 2 marks

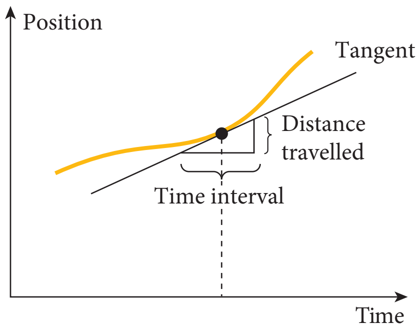


Figure 7.8 ▲ Instantaneous speed, v_{inst}, is the value of the gradient of the tangent to the distance time graph.

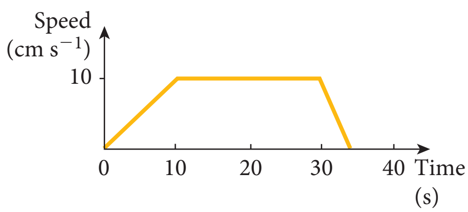
Instantaneous speed, v_{inst}

Glancing down at the speedometer of a car will give information about the vehicle's speed at that moment. This is the car's instantaneous speed, v_{inst}. When observing an object in motion it is often very difficult to measure its instantaneous speed. To find a speed, we need to measure the distance and the time intervals. This means that every measurement of speed is really a measurement of some average speed. In reality, all measurements are averages. However, if the time interval is very small, the average speed becomes very close to the instantaneous speed: $v_{inst} \approx v_{ave}$. Instantaneous speed is really a model for the idea of an object's speed at one instant in time.

If the time interval is made smaller and smaller so that it approaches zero, the value of the v_{ave} being measured approaches the gradient of the d versus t graph.

WORKED EXAMPLE 7.5

1 Refer to the speed versus time graph for an object shown in Figure 7.9.
- a Find the distance travelled:
  - i in the first 10.0 s. (1 mark)
  - ii between 10.0 s and 30.0 s. (1 mark)
  - iii between 30.0 s and 35.0 s. (1 mark)
- b Find the average speed of the object for the trip. (1 mark)
- c Find the distance travelled in 15 s. (1 mark)
- d How long does the object take to travel 70 cm? (2 marks)



▲ **Figure 7.9**

Answers

Distance = area under the graph

a i Distance $= \frac{1}{2} \times 10\,cm\,s^{-1} \times 10\,s = 50\,cm$

ii Distance $= 10\,cm\,s^{-1} \times 20\,s = 200\,cm$

iii Distance $= \frac{1}{2} \times 10\,cm\,s^{-1} \times 5\,cm\,s^{-1} = 25\,cm$

b Average speed $= \frac{\text{total distance}}{\text{time interval}}$

$= \frac{275\ cm}{35\ s}$

$= 7.9\ cm\ s^{-1}$

c Distance $= \frac{1}{2}(10\,cm\,s^{-1} \times 10\,s) + (10\,cm\,s^{-1} \times 5\,s) = 100\,cm$

d Somewhere between 10 s and 15 s the object has travelled 70 cm.

Let t be the time at which the object has travelled 70 cm.

$s = (t - 10\,s) \times v$ (i.e. the area under the graph after $t = 10.0$ s, i.e. 20 cm)

$t - 10\,s = \frac{s}{v}$

$t = \frac{s}{v} + 10\,s$

$t = \frac{20\ cm}{10\ cm\ s^{-1}} + 10\,s$

$t = 12\,s$

Logic

Identify distance as area under the graph.	1 mark
Calculate the area.	1 mark
Calculate the area.	1 mark
Calculate the gradient.	1 mark
Calculate the area.	1 mark
Calculate the area.	1 mark
Calculate the answer.	1 mark

Try these yourself

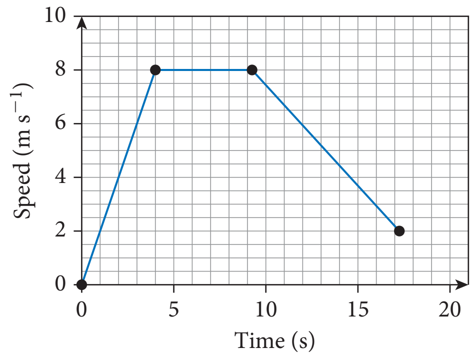
At the school sports, Elle runs the 100 m in 17.25 s (see Figure 7.10).

a Calculate the distance covered: (2 marks)

i in the first 6.0 s.

ii from 6.0 s to 8.0 s.

b Calculate the average speed in the first 8.0 s. (1 mark)



▲ Figure 7.10

QUESTION SET 7.1

Remembering

1 What is the difference between:

  a distance and displacement?

  b speed and velocity?

  c average and instantaneous speed?

  d average and instantaneous velocity?

2 What does the value of the gradient of a distance versus time graph for an object represent?

Understanding

3 Explain why instantaneous speed at a time during a journey can be quite different from the average speed for the whole journey.

4 For a speed versus time graph, show how to find the units of:

  a gradient.

  b area.

Applying

5 Jane averages 80 km h^{-1} for a 120 km journey. For the first 60 km, she averages 60 km h^{-1}. What must her average speed have been for the remainder of the journey?

6 A battery-operated car travels 6.0 m north in 2.4 s and then 6.0 m south in 1.8 s.

  a What is the displacement of the car?

  b What is the average velocity of the car?

  c What is the distance travelled by the car?

  d What is the average speed of the car?

7 Find the change in displacement and velocity for the following movements.

  a 50 km, S to 120 km, W in 1.5 h (m s^{-1})

  b 250 m, SW to 550 m, NW in 88 s (m s^{-1})

  c 700 m, N35°E to 950, S20°W in 35 min (m s^{-1})

Analysing

8 A cyclist travels at 30 km h^{-1} for 0.5 hours and then at 50 km h^{-1} for the next hour until the destination is reached.

  a How fast would a second cyclist, travelling at a constant speed, need to ride to arrive at the destination in the same time?

  b Explain why the answer to part a above is not simply 40 km h^{-1}.

9 The world record for the 100 m sprint is approximately 10 s. At this average speed, in what time would the record holder run 1500 m?

10 The Blue Orchid and the Yellow Devil taxi services pick up passengers at an airport at midday to drive to the same destination. The driver of the Yellow Devil taxi averages 100 km h^{-1} for 4 h, while the driver of the Blue Orchid taxi travels more sedately for 3 h, at an average speed of 80 km h^{-1}.

At what speed must the Blue Orchid driver travel during the next hour so that the two taxis arrive at the same place at 4 pm?

Constant acceleration along a straight line

The speed of sprinters in a 100 m race increases very quickly in the first few seconds of the race. Once they cross the finish line their speeds gradually decrease until they stop. In everyday language, this would be described as **acceleration** and then **deceleration**. Both involve a change in speed over a time interval, so they are both examples of acceleration – positive acceleration and negative acceleration. When speed in the positive direction is increasing, it is positive acceleration. When the speed in the positive direction is decreasing it is an example of negative acceleration.

Acceleration can therefore be positive or negative. A car becoming faster in the negative (reverse) direction would be undergoing negative acceleration, or acceleration in the negative direction. Acceleration occurs over a time interval, and is defined as the change in velocity divided by the time interval:

$$a_{ave} = \frac{\Delta v \text{ (m s}^{-1})}{\Delta t \text{ (s)}} = \frac{v}{t} \text{ (Unit: m/s}^2 \text{ or m s}^{-2})$$

Instantaneous acceleration can be found using the same method as instantaneous speed. As Δt approaches zero, the value of the acceleration is the value of the gradient of the velocity versus time graph at that time.

As velocity is speed with direction, acceleration will be occurring when there is a change in the direction of motion, even if the speed is constant. This affects our analysis of projectile motion (page 226) and circular motion. However, for the time being, we will concentrate only on motion in a straight line.

Circular motion is studied in Unit 3.

The unit of acceleration is m s^{-2}, spoken as metres per second per second. The reason for this sometimes confusing unit is explained in Worked example 7.6.

WORKED EXAMPLE 7.6

The speed of a car increases at a steady rate from 5.0 m s^{-1} to 15.0 m s^{-1} in 4.0 seconds. What was the car's average acceleration? (2 marks)

Answer	Logic	
$a_{ave} = \frac{\Delta v}{\Delta t}$		
$a_{ave} = \frac{(15.0 - 5.0) \text{ m s}^{-1}}{4.0 \text{ s}}$	Substitute the correct values.	1 mark
$a_{ave} = \frac{2.5 \text{ m s}^{-1}}{\text{s}}$		
$a_{ave} = 2.5 \text{ m s}^{-2}$	Calculate the correct answer and units.	1 mark

Expressed in words, the speed of the car is increasing by 2.5 m s^{-1} every second.

Try these yourself

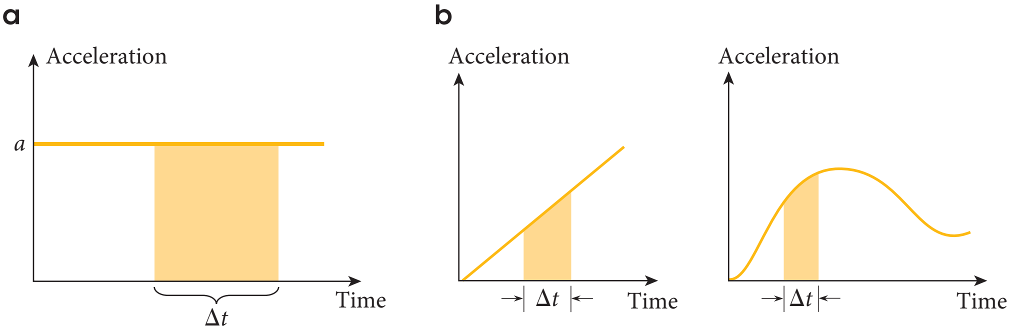
Find the average acceleration, in $m\,s^{-2}$, for a car when its speed:

a increases from $16\,m\,s^{-1}$ to $40\,m\,s^{-1}$ in 3.0 s. (1 mark)

b decreases from $75\,km\,h^{-1}$ to $40\,km\,h^{-1}$ in 5.0 s. Give your answer in $m\,s^{-1}$. (1 mark)

The area under an acceleration versus time graph has the units of speed, $m\,s^{-1}$. Finding the change in the speed between two values of time will involve finding the rectangular area under the acceleration versus time graph. This applies to objects with constant acceleration, as shown in Figure 7.11(a).

For non-constant acceleration, the change in speed is still found using the area under the acceleration versus time graph, as shown in Figure 7.11(b).



◀ **Figure 7.11**
a) The area under the acceleration–time graph gives the change in speed. b) The area under the acceleration–time graph is the change of speed, even for non-constant accelerations.

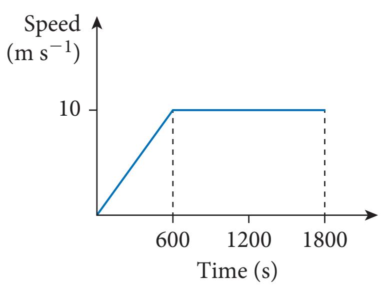
WORKED EXAMPLE 7.7

A cruise ship accelerates at a constant rate for 10.0 minutes until it reaches a speed of $10\,m\,s^{-1}$. It then continues to travel in a straight line for 20.0 minutes at $10\,m\,s^{-1}$.

a Sketch a speed ($m\,s^{-1}$) versus time (s) graph for the ship for the 30 minutes. (3 marks)
Note: convert minutes to seconds, so that the time axis goes from 0 to 1800 s.

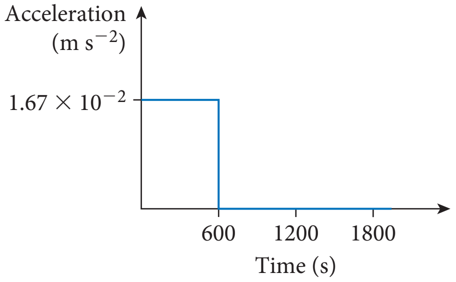
b What was the ship's acceleration for the first 10.0 minutes, in m^{-2}? (2 marks)

c Sketch an acceleration versus time graph for the ship for the 30 minutes. (3 marks)

Answers	Logic	
a 	Mark the correct axes and labels. Plot the line.	1 mark 2 marks
b Acceleration ($m\,s^{-2}$) $a = \frac{\Delta v}{\Delta t}$ $= \frac{10\,m\,s^{-1} - 0\,m\,s^{-1}}{10\,min \times 60\,s\,min^{-1}}\,m\,s^{-1}$	Use the correct equation and substitute the known values.	1 mark
$= 1.67 \times 10^{-2}\,m\,s^{-2}$	Calculate the correct answer.	1 mark

▶

c

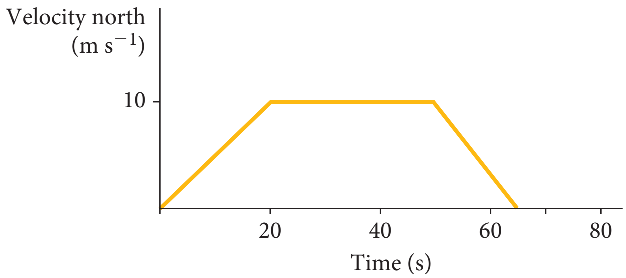


Mark the correct axes and labels. 1 mark

Plot the line correctly. 2 marks

Try these yourself

The velocity versus time graph for a bus travelling along a straight road is shown in Figure 7.12.



◀ Figure 7.12

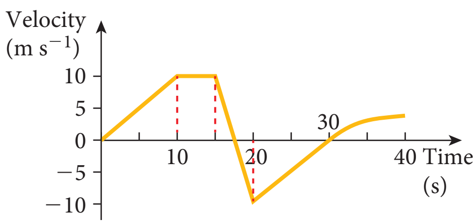
a What distance is travelled by the bus while it is accelerating in a positive direction? (1 mark)

b What is the distance between the two bus stops? (1 mark)

c What is the acceleration of the bus in the first 10 s of motion? (1 mark)

WORKED EXAMPLE 7.8

Figure 7.13 shows the velocity–time graph for an object as it moves along a straight line.



◀ Figure 7.13

a Find the displacement after 20 s. (1 mark)

b Find the acceleration at 17.0 s. (1 mark)

c What is the acceleration at t = 35.0 s? (1 mark)

Answers	Logic	Marks
a Displacement = area under graph	Use the correct area data.	1 mark
$s = \frac{1}{2}(10\,\text{m s}^{-1} \times 10\,\text{s}) + (10\,\text{m s}^{-1} \times 5\,\text{s}) + \frac{1}{2}(10\,\text{m s}^{-1} \times 2.5\,\text{s}) + \frac{1}{2}(-10\,\text{m s}^{-1} \times 2.5\,\text{s})$ $= 50\,\text{m} + 50\,\text{m} + 12.5\,\text{m} - 12.5\,\text{m}$ $= 100\,\text{m}$	Calculate the correct answer.	1 mark

b Acceleration is the gradient of the velocity versus time graph:

$$a = \frac{(-10\,\text{m s}^{-1}) - (+10\,\text{m s}^{-1})}{20\,\text{s} - 15\,\text{s}}$$

$$= -4.0\,\text{m s}^{-1}$$

Use the correct area data to calculate answer. 1 mark

c Finding the gradient of the tangent to the graph at $t = 35.0\,\text{s}$ gives

$$a = \frac{5.0\,\text{m s}^{-1} - 0\,\text{m s}^{-1}}{40\,\text{s} - 25\,\text{s}}$$

$$\Rightarrow a = \frac{5\,\text{m s}^{-1}}{15\,\text{s}}$$

$$\Rightarrow a = 0.3\,\text{m s}^{-2}$$

Use the correct area data to calculate answer. 1 mark

Try this yourself

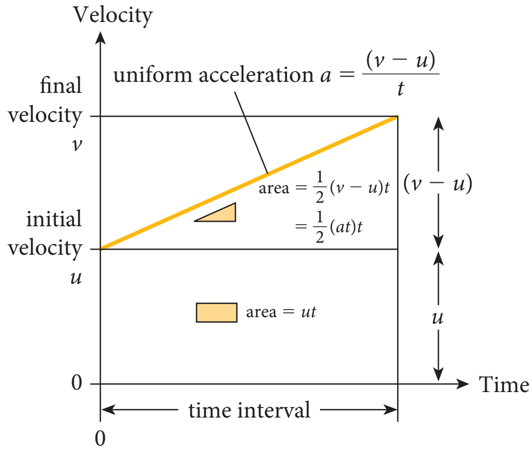
A motorbike, initially, at rest accelerates constantly at $5\,\text{m s}^{-2}$ until it reaches a speed of $20\,\text{m s}^{-1}$.Sketch a speed versus time graph until it reaches $20\,\text{m s}^{-1}$. Clearly show the time when this final speed is reached. (4 marks)

Analysing constant acceleration along a straight line

Graphs and algebraic formulas can both be used to represent motion. They are, in fact, equivalent representations or models of motion. There is no reason why one representation of motion should be taken to be more important than another. For motion at constant acceleration, the speed versus time graph is a straight line. The gradient represents the acceleration and the area under the graph represents the distance covered. From the graph, a number of simple equations can be derived. These equations represent the same motion as the graph and are, therefore, exactly equivalent to the graphs.

Figure 7.14 shows a velocity versus time graph for a car moving in a straight line at constant acceleration.

The car has an initial velocity of u. It is accelerating at a constant rate, shown by the straight line on the graph. After a time interval $\Delta t = t$, the car's final velocity is v. During this time it covers a distance interval of $\Delta d = s$.



▲ **Figure 7.14** Motion of a car along a straight line

> In the equations for constantly accelerated motion in a straight line:
> - the symbol t represents a time interval not an instantaneous time.
> - the symbol s represents a distance interval, not a particular point on a line.

In summary:

Initial speed $= u$

Final speed $= v$

Acceleration $= a$

Time interval, $\Delta t = t$

Distance interval, $\Delta d = s$

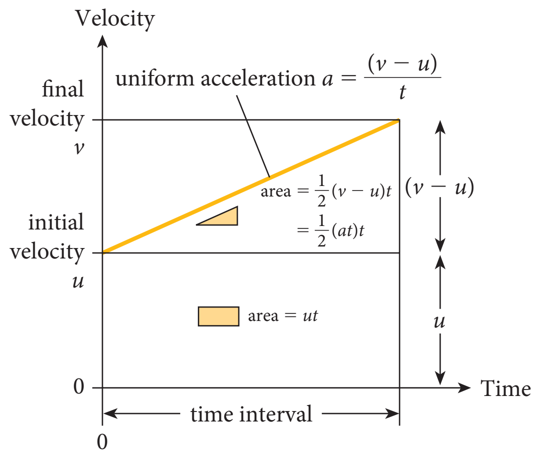


Figure 7.14 is repeated at the left for reference.

From the definition of acceleration, we get:

$$a = \frac{\Delta v}{\Delta t} = \text{gradient of the line}$$

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

$$v - u = at$$

$$v = u + at$$

Finding the area under the graph as the sum of the small rectangle and the triangle gives:

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

$$s = ut + \frac{1}{2}(at)t$$

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

Using the same graph, finding the area under the graph as a trapezium gives:

$$s = \text{average height} \times t$$

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

Another useful equation that can be derived from these two equations can be found using some algebra on the first two equations:

$$v = u + at$$

Squaring both sides:

$$\Rightarrow v^2 = (u+at)^2$$

$$v^2 = u^2 + 2uat + a^2t^2$$

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

$$v^2 = u^2 + 2a\left(ut + \frac{1}{2}at^2\right)$$

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

Summary of motion in a straight line at constant acceleration

The two methods used for analysing motion in a straight line are graphical and algebraic. Graphical analysis is often a simpler, more visual way that will give the same answer to a problem as algebraic techniques. Graphical analysis should be attempted wherever possible.

Graphically

$$v = \frac{s}{t} = \text{gradient of the } s \text{ versus } t \text{ graph}$$

$$s = \text{area under the } v \text{ versus } t \text{ graph}$$

$$a = \frac{\Delta v}{\Delta t} = \text{gradient of the } v \text{ versus } t \text{ graph}$$

$$\Delta v = \text{area under the } a \text{ versus } t \text{ graph}$$

Algebraically

The equations for straight-line motion with constant acceleration are:

$$v = u + at$$

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

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

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

Each of these equations involves four variables. When solving problems algebraically, you will need to know the values of three of the five variables, v, u, a, t or s. The fourth can be found by simple substitution in the appropriate equation. It is then possible to use another equation to find the fifth variable.

Graphical analysis is often simpler and more obvious than algebraic analysis. Acceleration (gradient) and distance (area) can often be computed easily once a v–t graph is sketched and relevant data points identified.

Both methods yield the same answers because they are both models of the same motion.

WORKED EXAMPLE 7.9

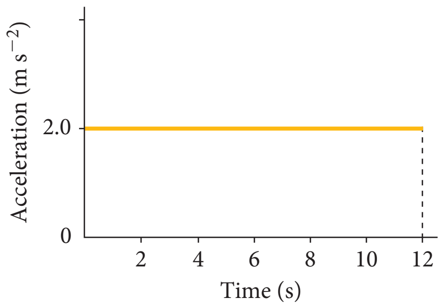
A car initially travelling at a speed of $4.0\,\text{m s}^{-1}$ accelerates at $2.0\,\text{m s}^{-2}$ for $12\,\text{s}$.

a Sketch the acceleration versus time graph for the car. (1 mark)

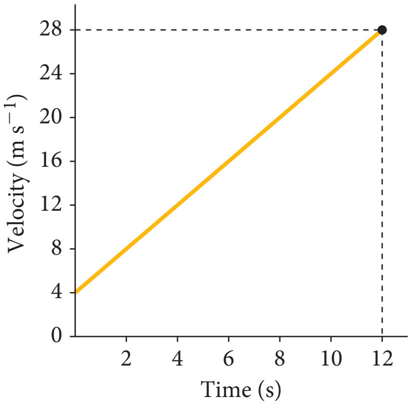
b Find the velocity v of the car after 8.0 s. (1 mark)

c Sketch the v versus t graph. (1 mark)

d Find the distance moved by the car in the 8.0 s. (1 mark)

Answers	Logic	
a 	Correctly sketch graph and label points.	1 mark
b Area under the a versus t graph: $\Delta v = 2.0\,\text{m s}^{-2} \times 8.0\,\text{s} = 16\,\text{m s}^{-1}$ Add the car's initial velocity: $v = 4.0\,\text{m s}^{-1} + 16\,\text{m s}^{-1} = 20\,\text{m s}^{-1}$.	Identify velocity as the area under the curve and calculate it.	1 mark

c



Correctly sketch the graph. 1 mark

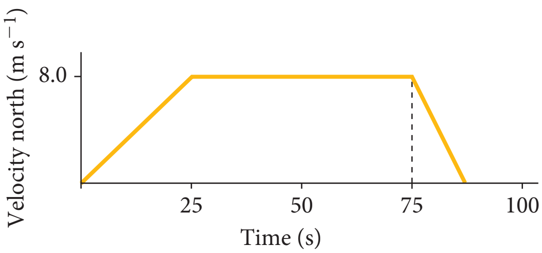
d Area under the *v* versus *t* graph using the trapezium:

$s = \frac{1}{2}(4.0\,\text{m}\,\text{s}^{-1} + 20\,\text{m}\,\text{s}^{-1}) \times 8.0\,\text{s} = 96\,\text{m}$

Identify distance as the area under the curve and correctly calculate it. 1 mark

Try these yourself

A train accelerates uniformly from rest to 8.0 m s^{-1} in 25 s. It then travels for 50 s at 8.0 m s^{-1} before coming uniformly to a stop in 12 s (see Figure 7.15).



◀ **Figure 7.15**

a What is the distance travelled by the train in the first 20 s? (1 mark)

b What is the distance between the two stops? (1 mark)

c What is the acceleration of the train in: (1 mark)

  i the first 10 s of the train's motion?

  ii the last 10 s of the train's motion? (1 mark)

d Sketch the acceleration versus time graph for the motion of the train. (2 marks)

e After what time interval had the train travelled 300 m? (1 mark)

REACTION TIME

Test your own reaction time at the traffic lights.

Non-uniform motion

Non-uniformly accelerated motion refers to motion of objects when the acceleration is changing. Where this occurs, the equations of motion cannot be used easily. It is better to refer to the graphs of the motion, especially the speed versus time graphs. Gradients and areas under the graph can be used to find approximate answers within experimental limits.

Emergency braking

Stopping a car in an emergency situation can be a frightening experience. Inexperienced drivers often say that it took a longer distance to stop the car than they thought it would.

Driver reaction

It takes a measureable time for a driver to recognise and react to a hazard. The **reaction time** is the time it takes for the driver to apply the brakes after the hazard is first identified. During this time, the car continues to move with its original velocity. At 90 km h^{-1}, this is 25 m every second.

WOW

Varieties of distractors

Neuroscience has identified several different ways in which people can be distracted. Three of these are sensory, emotional and cognitive. If there are a lot of visual stimuli, such as scenery or a pageant, drivers can be distracted and not notice a hazard. Having an argument or feeling upset can be distracting. Talking about a difficult problem while driving may also increase reaction time.

MOST DISTRACTING

Take the test to see what distracts you most.

Reaction times vary considerably between drivers and situations. Factors that affect driver reaction time include driver skill, alertness, drug and alcohol effects and inattention. Inattention due, for example, to texting while driving, adds an unacceptable and dangerous amount to reaction time, and has been fatal to drivers, pedestrians, cyclists and other road users. A prompt reaction time is about 0.2 s or less. Inattention can cause reaction times well over 1.0 s.

Stopping time is the total time elapsed from the recognition of the hazard to coming to a stop. It is the sum of the reaction time and the braking time.

The **reaction distance** is the distance the car travels before the driver takes appropriate action, such as applying the brakes. At 100 km h^{-1}, a distracted driver who reacts after 1.5 s will travel a reaction distance of approximately 42 m. This is one reason why there are special 40 km h^{-1} school zones near schools in the mornings and afternoons. These schools zones alert drivers to potential hazards, such as children on the road. Early warning and slower speeds reduce the reaction time – drivers are more alert – and slower speeds mean shorter reaction distances. The overall effect is to reduce the distance needed to come to a stop.

Braking distance is the distance covered while the brakes are applied.

Stopping distance is the total distance covered from the recognition of the hazard to coming to a stop. It is the sum of the reaction distance and the braking distance.

ACTIVITY 7.1

MEASURING REACTION TIME

Aim

To measure a person's reaction time

You will need

- 30 cm ruler

What to do

1. One person holds the ruler vertically at the 30 cm mark.
2. The second person places the ruler between thumb and forefinger at the zero end of the ruler, without holding it.
3. Without warning, the ruler is released and the second person must catch the ruler between thumb and forefinger.
4. Record the distance the ruler fell in cm. Convert this to metres.

Analysis of results

Calculate the reaction time, the time interval between when the ruler was released and caught, using the formula:

$$t = \sqrt{\frac{2 \times \text{distance fallen in metres}}{9.8\,\text{m s}^{-2}}}$$

What did you discover?

1 Suggest reasons why repeating this procedure was not part of the experimental procedure.

2 There are other ways to test reaction time, including apps found on iTunes and Google Play. Compare the results from this experimental procedure with those from other sources.

Taking it further

Compare your results with those of others in the class. Analyse the results by asking participants if they felt alert or tired when the experiment was performed. Suggest how this might be linked to road safety.

Bringing the car to a stop

Once the brakes of the car are applied, the car will come to a stop in a distance that is affected by many conditions. The braking distance, the distance the car takes to come to a stop once the brakes are applied, can vary enormously from car to car and from day to day. The conditions of the brakes, the road surface, the vehicle's tyres, whether it is wet or dry as well as modern technologies such as ABS (anti-skid braking system) all affect the braking distance. When braking, the car is undergoing deceleration, or negative acceleration. A deceleration of $9.8\,\text{m s}^{-2}$, achievable only by some cars in ideal circumstances, is said to be '1*g*'. It is the same as the acceleration due to gravity. A deceleration of $4.9\,\text{m s}^{-2}$ would be 0.5*g*, or half the acceleration due to gravity. Some road authorities use $3.4\,\text{m s}^{-2}$ as a comfortable rate of stopping. An emergency stop could be at twice this rate. A crash stop may involve rates too great for occupants to survive.

Stopping time = reaction time + braking time
Stopping distance = reaction distance + braking distance

Stopping times and distances can be represented graphically (see Figure 7.16).

The biggest factor by far that influences braking distance is the car's speed when the brakes are applied. The braking distance of a car is proportional to the square of the car's speed. Thus, if the braking distance for a car travelling at $40\,\text{km h}^{-1}$ is 15 m, the braking distance at $80\,\text{km h}^{-1}$ increases by 4 times to 60 m.

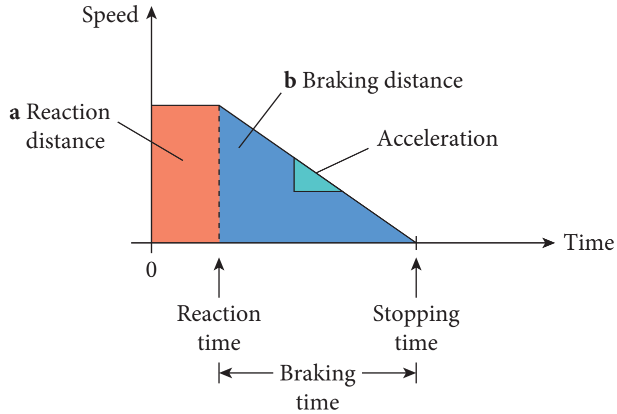


Figure 7.16 ▶ Graph of a vehicle as it comes to a stop. The time axis shows the reaction time, stopping time and, by difference, the braking time. The areas show: a) the reaction distance, b) the braking distance and, by sum, the stopping distance. The gradient of the braking section is the acceleration of the vehicle.

QUESTION SET 7.2

Remembering

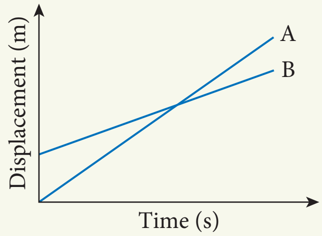
1 Define the relationship between speed, distance and time.

2 Describe the difference between stopping distance and braking distance.

Understanding

3 Explain why we can say, truthfully, that algebraic formulas and graphs of motion are equivalent representations.

4 For the motion of the two cars, A and B, in Figure 7.17, explain how it can be deduced that they never travel at the same speed.



▲ Figure 7.17

Applying

5 The velocity versus time graph for the motion of a sprint runner is shown in Figure 7.18.

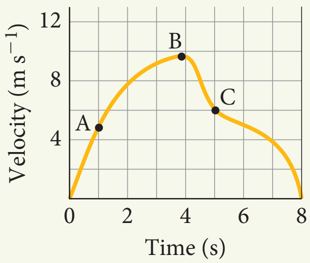
a From the graph, find the acceleration of the athlete at points:

i A ii B iii C

b Estimate the distance travelled by the athlete in the first 4.0 s.

c At what time is the acceleration of the athlete the greatest?

d What is the maximum velocity reached by the athlete?



▲ Figure 7.18

6 A car that is accelerating uniformly from rest travels 22.5 m in the eighth second of its motion.

a Calculate the acceleration of the car.

b Calculate how far the car travels in the first 6.0 s.

Analysing

7 Explain why the velocity versus time graph of a car can never have a line parallel to the velocity axis.

8 A car accelerates uniformly from rest to 25 m s^{-1} in 10 s.

a Sketch the graphs for:

i displacement versus time.

ii velocity versus time.

iii acceleration versus time.

b From the velocity versus time graph, state the car's:

i acceleration.

ii speed at 5.0 s.

iii displacement at 10 s.

9 A driver travelling at 75 km h^{-1} has a reaction time of 0.31 s. When a hazard occurs 50 m ahead, the driver slams on the brakes and comes to a stop just a few millimetres from the hazard.

Calculate and show the following on a carefully drawn sketch graph.

a Reaction distance (m)

b Braking distance (m)

c Braking time

d Stopping time

e Stopping distance

f Acceleration when braking

Reflecting

10 Identify and describe two or more advantages of being able to graph a car's motion rather than dealing with its motion using only equations.

Projectile motion

Once an object is released or thrown, the only force acting upon it is the **gravitational force**. Gravitational force is the force applied by one mass on another mass. **Gravity** is the common name given to the force applied by Earth's rather large mass, 6.0×10^{24} kg, on ordinarily rather small masses on and near Earth. Near Earth's surface, gravity will cause a force on an object that results in that object accelerating vertically downwards with an acceleration of $9.8\,\text{m s}^{-2}$. This is known as **gravitational acceleration**, and is given the symbol g. Although the value of g varies slightly around the world, using a value of $9.8\,\text{m s}^{-2}$ gives sufficiently accurate answers for our purposes.

Why then, does a hammer fall faster than a feather when both are released together? Mostly this is because air friction has a greater effect on lighter, fluffier objects than it does on denser, smoother objects. There is also a buoyancy force that is more significant for lighter objects. An experiment to show this was done by Apollo 15 astronaut Commander David Scott. He dropped a hammer and a feather on the Moon, which has no atmosphere. With no air friction, both objects fell at the same rate.

HAMMER VERSUS FEATHER

Apollo 15 Commander David Scott dropped a feather and a hammer on the Moon in 1971.

Because of the complexity of calculations involving air friction, we will, for this course, ignore its effects. Therefore, the only acceleration for any projectile motion will be $9.8\,\text{m s}^{-2}$ vertically downwards. Graphs of the motion of an object that is dropped are shown in Figure 7.19 for displacement, velocity and acceleration versus time. Notice that the y axis is for downwards values.

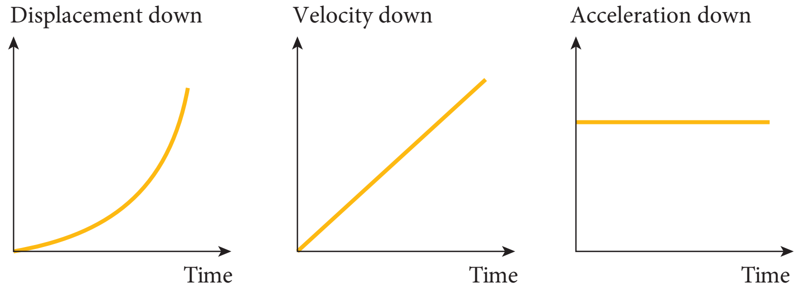


Figure 7.19 ▶ Graphs showing the motion of a falling object

The consequence of *g*

The value of gravitational acceleration $9.8\,\text{m s}^{-2}$ means that anything dropped from a height will reach a speed of $9.8\,\text{m s}^{-1}$ after the first second. This is about $35\,\text{km h}^{-1}$. After the next second, it will be falling with a speed of $9.8 + 9.8 = 19.6\,\text{m s}^{-1}$, or about $70\,\text{km h}^{-1}$. For every second it falls, the speed of the object will increase by $9.8\,\text{m s}^{-1}$, or $35\,\text{km h}^{-1}$. With no air friction, a brick would be falling at $98\,\text{m s}^{-1}$ or about $350\,\text{km h}^{-1}$ 10 seconds after being dropped. This is why we say $9.8\,\text{m s}^{-2}$ is '9.8 metres per second per second'. It means that the speed is changing by 9.8 metres per second every second.

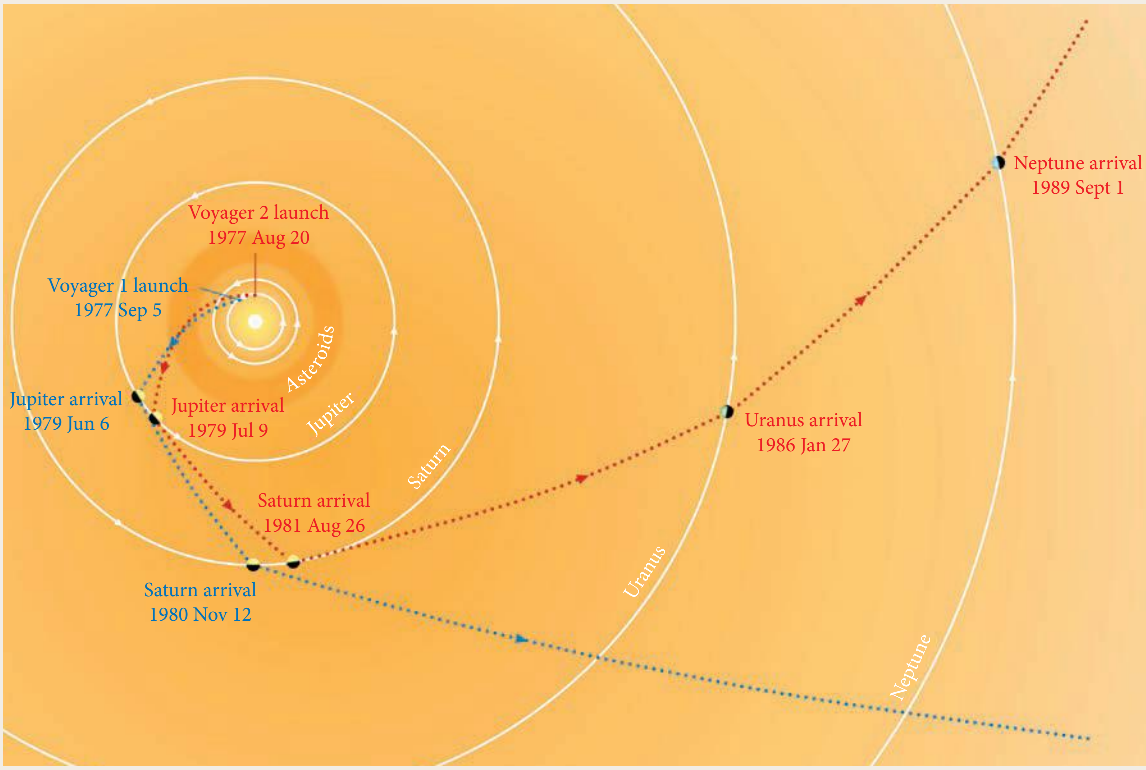
The sign of the values of the variables depends on whether the upwards direction is considered positive. There are occasions when assigning positive values to the downwards direction may eliminate many negative signs in the calculation.

For vertical projectile motion in a straight line, the value of acceleration of the projected object is always taken as being the value of g, $9.8\,\text{m s}^{-2}$ vertically downwards. There is no horizontal acceleration.

Scientific literacy: NASA's Voyager missions – the most distant human-made motion

The mission objective of the Voyager Interstellar Mission (VIM) is to extend the NASA exploration of the solar system beyond the neighbourhood of the outer planets to the outer limits of the Sun's sphere of influence, and possibly beyond. This extended mission is continuing to characterise the outer solar system environment and search for the heliopause boundary, the outer limits of the Sun's magnetic field and outward flow of the solar wind. Penetration of the heliopause boundary between the solar wind and the interstellar medium will allow measurements to be made of the interstellar fields, particles and waves unaffected by the solar wind.

The Voyager Interstellar Mission (VIM) is an extension of the Voyager primary mission that was completed in 1989 with the close flyby of Neptune by the Voyager 2 spacecraft. Neptune was the final outer planet visited by a Voyager spacecraft. Voyager 1 completed its planned close flybys of the Jupiter and Saturn planetary systems while Voyager 2, in addition to its own close flybys of Jupiter and Saturn, completed close flybys of the remaining two gas giants, Uranus and Neptune.



Science Photo Library/Mark Garlick

▲ **Figure 7.20**
Interstellar mission of Voyager 1 and 2

At the start of the VIM, the two Voyager spacecraft had been in flight for more than 12 years, having been launched in August (Voyager 2) and September (Voyager 1), 1977. Voyager 1 was at a distance of approximately 40 AU from the Sun, and Voyager 2 was at a distance of approximately 31 AU.

As of September 2013, Voyager 1 was at a distance of 18.7 billion kilometers (125.3 AU) from the Sun and Voyager 2 at a distance of 15.3 billion kilometers (102.6 AU).

Voyager 1 is escaping the solar system at a speed of about 3.6 AU per year, 35 degrees out of the ecliptic plane to the north, in the general direction of the Solar Apex (the direction of the Sun's motion relative to nearby stars). Voyager 2 is also escaping the solar system at a speed of about 3.3 AU per year, 48 degrees out of the ecliptic plane to the south.

Both Voyagers are headed towards the outer boundary of the solar system in search of the heliopause, the region where the Sun's influence wanes and the beginning of interstellar space can be sensed. The heliopause has never been reached by any spacecraft; the Voyagers may be the first to pass through this region, which is thought to exist somewhere from 13 to 22 billion kilometers from the Sun. This is where the 1.6 million-kilometers-per-hour solar winds slows to about 400 000 kilometers per hour—the first indication that the wind is nearing the heliopause. The Voyagers should cross the heliopause 10 to 20 years after reaching the termination shock. The Voyagers have enough electrical power and thruster fuel to operate at least until 2020. By that time, Voyager 1 will be 19.9 billion kilometers from the Sun and Voyager 2 will be 16.9 billion kilometers away. Eventually, the Voyagers will pass other stars. In about 40000 years, Voyager 1 will drift within 1.6 light years (14.9 trillion kilometer) of AC+79 3888, a star in the constellation of Camelopardalis. In some 296000 years, Voyager 2 will pass 4.3 light years (40 trillion kilometers) from Sirius, the brightest star in the sky. The Voyagers are destined—perhaps eternally—to wander the Milky Way.

NASA Jet Propulsion Laboratory, California Institute of Technology (2013) 'Voyager: The Interstellar Mission: The Mission'. http://voyager.jpl.nasa.gov/mission/interstellar.html

Questions

1 What is the purpose of the Voyager Interstellar Mission? How does this relate to the 'most distant human motion'?

2 Define 'solar wind' and 'heliopause'.

3 Draw a diagram to show the directions of Voyager 1 and Voyager 2 relative to the motion of the ecliptic plane.

4 At the start of VIM, what was the average speed of Voyager 1 since being launched?

5 By what factor has the solar windspeed been reduced by the time it reaches the start of the heliopause?

6 Calculate the speed, in metres per second, at which Voyager 1 is escaping the solar system. Report the answer using scientific notation (see page 408).

7 Explain why the heliosphere is an interesting region for astronomers.

8 How far apart, in kilometres, will Voyager 1 be from Voyager 2 in 2020? Report the answer using scientific notation.

9 Compare the distance of the closest flyby of AC+79 3888 by Voyager 1 with Earth's distance from the Sun.

10 How has your study of motion helped you develop a quantitative understanding of the Voyager Interstellar Mission?

Objects falling directly downwards

When analysing the motion of a falling object, we can make the:

- origin the point at which the objects starts to move.
- downward direction positive.
- variables s, u, v, a and t all positive, so that $a = g = +9.8\,\text{m}\,\text{s}^{-2}$, if in the downwards direction.

  Worked example 7.10 uses the same equations of motion as used previously.

WORKED EXAMPLE 7.10

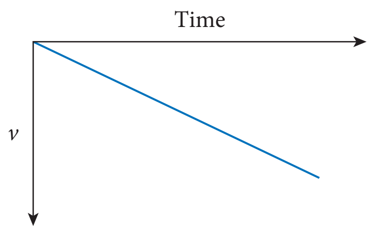
A watch falls from a Sydney Harbour Bridge climber's wrist. The watch falls for 2.5 s before hitting a car below.

a Sketch a velocity versus time graph of this motion. (2 marks)

b With what velocity does the watch hit the car? (3 marks)

c How far did the watch fall? (3 marks)

d Show how these results can be obtained by using only the graph in part a (gradient and area). (3 marks)

Answers	Logic
a 	Draw the correct graph (straight line). 2 marks

b	$u = 0$, $a = 9.8\,\text{m s}^{-2}$, $t = 2.5\,\text{s}$, $v = ?$	Identify the known variables and the variable required.	1 mark
	$v = u + at$	Use the appropriate equation.	1 mark
	$v = 0 + 9.8\,\text{m s}^{-2} \times 2.5\,\text{s}$	Substitute the correct values.	1 mark
	$= 24.5\,\text{m s}^{-1}$ vertically down		
c	$u = 0$, $a = 9.8\,\text{m s}^{-2}$, $t = 2.5\,\text{s}$, $s = ?$	Identify the known variables and the variable required.	1 mark
	$s = ut + \frac{1}{2}at^2$	Use the appropriate equation.	1 mark
	$= 0 + \frac{1}{2} \times 9.8\,\text{m s}^{-2} \times (2.5\,\text{s})^2$	Substitute the correct values.	1 mark
	$= 31$ m down (correct to 2 significant figures)	Calculate the correct answer.	1 mark
d	Gradient $= 9.8\,\text{m s}^{-2} = \frac{v\,\text{m s}^{-1}}{2.5\,\text{s}}$		
	$\Rightarrow v = 9.8\,\text{m s}^{-2} \times 2.5\,\text{s} = 24.5\,\text{m s}^{-1}$ down	Use the gradient formula and substitute the correct values.	1 mark
	Area $= s = \frac{1}{2}(24.5\,\text{m s}^{-1}) \times (2.5\,\text{s})$		
	$\Rightarrow s = 31$ m down	Calculate the correct answer.	1 mark

Try these yourself

A rock dropped from a cliff hits the ocean with a speed of $44.1\,\text{m s}^{-1}$.

a	Sketch the velocity versus time graph for this motion.	(2 marks)
b	For how long did the rock fall?	(3 marks)
c	How high is the cliff?	(3 marks)
d	Show how these results can be obtained by using only the graph in part a (gradient and area).	(3 marks)

EXPERIMENT 7.2

GRAVITATIONAL ACCELERATION

For a falling object not affected significantly by air resistance, the value of the gravitational acceleration, *g*, can be found by collecting first-hand information.

Aim

To find the value of the gravitational acceleration, *g*

Materials

- ruler
- ball bearing
- electronic timer or timing photogate

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
The ball bearing may cause injury if thrown, dropped or stood on.	Never throw ball bearings. Manage the use of the ball bearing carefully. Never leave the ball bearing lying on the ground.

Procedure

1. Set up the electronic timing apparatus.
2. Carefully measure the vertical distance, *s*, that the ball bearing will fall.

3 Using the timing apparatus, measure the time, t, taken for the ball bearing to fall through the known vertical height when released from rest from the upper position.
4 Repeat this several times.
5 Change the height of fall and repeat the procedure.
6 Record sufficient data to plot a graph.

Results

1 Record all raw and derived data in a correctly constructed data table.
2 Plot the data as it is collected.
3 Estimate and record uncertainties in the data.

Analysis of results

1 Plot s versus t_{ave}, showing uncertainty bars.
2 Draw a line of best fit.
3 From the line of best fit, construct a data table of data points, (t_{ave}, s). Add an extra column for $(t_{ave})^2$.
4 Plot s versus $(t_{ave})^2$.
5 Draw a straight line of best fit and calculate the gradient.
6 Show that the equation $s = ut + \frac{1}{2}at^2$ can be used to find the acceleration from the gradient of the s versus $(t_{ave})^2$ graph.
7 Justify the best estimate of the value of the acceleration due to gravity, g, found in this experiment.
8 Use the least and greatest possible values of the gradient of the s versus $(t_{ave})^2$ graph to estimate the uncertainty in the experimental value of g. (Do not use the regression equation from your calculator!)

Discussion

1 Suggest ways in which this experiment could be made to be more accurate.
2 Evaluate the reliability of this procedure by analysing the variation in the separate measurements of time taken by the ball bearing before the average was found.
3 Suggest why a ball-bearing was used rather than a tennis ball or other similar object.

Conclusion

1 Summarise the experiment in one or two sentences.
2 Provide a precise value for g (best estimate ± uncertainty limits).
3 Comment on any difficulties you had when undertaking this experiment. What changes could be made to overcome these difficulties?

Objects projected directly upwards

When given an initial upwards velocity, a projectile will immediately begin to slow down, so $\Delta v < 0$ and $a < 0$, even though the distance from the start is initially increasing $s > 0$.

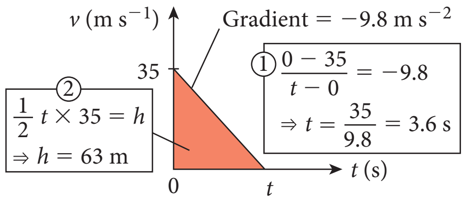
To keep the calculations as simple as possible, we can make the following decisions.

- The origin is the point where the object starts its motion.
- The upward direction is positive.
- Acceleration is downwards, and so is negative: $a = -9.8\,\text{m}\,\text{s}^{-2}$.
- The other variables s, u, v and t are positive if directed upwards and negative if directed downwards.

WORKED EXAMPLE 7.11

A firework ball is launched directly upwards with an initial speed of $35\,\text{m}\,\text{s}^{-1}$. It explodes at the highest position reached.

a Sketch the v versus t graph for the firework ball until it explodes. (2 marks)
b At what height does the firework ball explode? (3 marks)
c How long after being launched does the firework ball explode? (3 marks)

Answers	Logic	
a 	Sketch the graph correctly and label axes.	2 marks
b $u = 35\,\text{m s}^{-1}$, $a = 9.8\,\text{m s}^{-2}$, $v = 0$, $s = ?$	Identify the known variables and the variable needed.	1 mark
$v^2 = u^2 + 2as$	Identify the appropriate equation to use.	1 mark
$v^2 - u^2 = 2as$		
$2as = v^2 - u^2$		
$s = \dfrac{v^2 - u^2}{2a}$	Make *s* the subject.	1 mark
$= \dfrac{(0\,\text{m s}^{-1})^2 - (35\,\text{m s}^{-1})^2}{2(-9.8\,\text{m s}^{-2})}$		
$= 62.5\,\text{m}$		
$= 63\,\text{m}$ (to 2 significant figures; see page 411)		
c $u = 35\,\text{m s}^{-1}$, $a = 9.8\,\text{m s}^{-2}$, $v = 0$, $t = ?$	Identify the known variables and the variable needed.	1 mark
$v = u + at$	Identify the appropriate equation.	1 mark
$at = v - u$		
$t = \dfrac{v - u}{a}$	Make *t* the subject.	1 mark
$= \dfrac{0\,\text{m s}^{-1} - 35\,\text{m s}^{-1}}{-9.8\,\text{m s}^{-2}}$		
$= 3.6$ s (to 2 significant figures)		

Try these yourself

A human cannon ball is launched vertically upwards with a speed of $20\,\text{m s}^{-1}$.

a Sketch a velocity versus time graph for this motion. (2 marks)

b For how long does she stay in the air? (3 marks)

c What is the maximum height she reaches? (3 marks)

Object projected directly upwards then falls back down

In these examples of projectile motion, we can assume that the final velocity $v = -u$ when the object lands back at the launch position. This is due to the symmetry of the motion of the projectile. Remember, for as long as the object is in the air, its acceleration will always be $9.8\,\text{m s}^{-2}$ downwards. This remains so whether the object's velocity is either upwards or downwards. The result of this is that when the velocity is upward, the object's speed is decreasing. It will stop momentarily at the highest point of the flight and then commence its freefall back down again.

WORKED EXAMPLE 7.12

An arrow is fired vertically upwards with an initial speed of $18.8\,\text{m s}^{-1}$. It lands next to where it was launched.

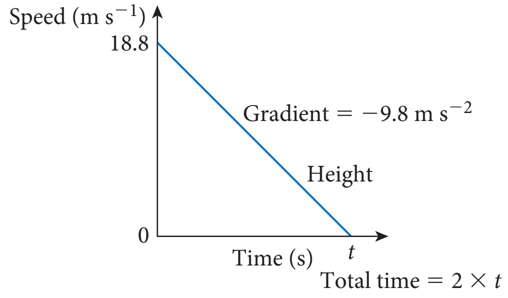
a Sketch a velocity versus time graph for this motion. (3 marks)

b For how long will the arrow remain in flight? (3 marks)

c Find the maximum height reached by the arrow. (3 marks)

Answers

a



Logic

Use the correct axes. 1 mark

Draw a straight line graph. 1 mark

Label the correct data points. 1 mark

b The gradient of the graph is $-9.8\,\text{m s}^{-2}$ from which the time to the top of the flight can be calculated.

Use the gradient equation. 1 mark

Graphical method

By inspection of graphs:

$$\frac{18.8\,\text{m s}^{-1}}{t} = 9.8\,\text{m s}^{-2}$$

Rearrange to make t the subject. 1 mark

$$\Rightarrow \quad t = \frac{18.8\,\text{m s}^{-1}}{8.8\,\text{m s}^{-2}}$$

$$= 1.92\,\text{s}$$

$$\Rightarrow \quad T = 2t$$

$$\Rightarrow \quad T = 3.84\,\text{s}$$

Find the time to the top of the flight and double it. 1 mark

Algebraic method

Identify the known and required variables at the top of the flight. 1 mark

t	u	a	v
?	$18.8\,\text{m s}^{-1}$	$-9.8\,\text{m s}^{-2}$	0

Select the appropriate equation. 1 mark

$$v = u + at$$

$$at = v - u$$

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

Correct answer 1 mark

$$t = \frac{0\,\text{m s}^{-1} - 18.8\,\text{m s}^{-1}}{-9.8\,\text{m s}^{-2}}$$

$$t = 1.92\,\text{s}$$

Total time $= 1.92 \times 2\,\text{s}$

$\approx 3.8\,\text{s}$ (to 2 significant figures)

Note: Both methods give the same answers because they represent the same motion. Graphical solutions are often more obvious. You need to build a wide repertoire of solution methods.

Graphical method

c Distance to top = area above the t axis:	Identify the known and required variables.	1 mark
$s = \frac{1}{2}ut$	Select the appropriate equation.	1 mark
$= \frac{1}{2}(18.8\,\text{m s}^{-1} \times 1.92\,\text{s})$		
$\Rightarrow s = 18\,\text{m}$ (to 2 significant figures)	Correct answer	1 mark

or

Algebraic method

By taking the upwards direction as positive, initial speed will be positive and position above the launch position will be positive. Acceleration being downwards will be negative.

At the maximum height reached, the velocity will be zero.

$u = 18.8\,\text{m s}^{-1}$, $a = -9.8\,\text{m s}^{-2}$, $v = 0$, $s = ?$	Identify the known and required variables.	1 mark
$v^2 = u^2 + 2as$ $2as = v^2 - u^2$ $s = \frac{v^2 - u^2}{2a}$ $s = \frac{(0\,\text{m s}^{-1})^2 - (18.8\,\text{m s}^{-1})^2}{2 \times (-9.8\,\text{m s}^{-2})}$	Select the appropriate equation and make s the subject.	1 mark
$s = 18\,\text{m}$ (to 2 significant figures)	Correct answer	1 mark

Try these yourself

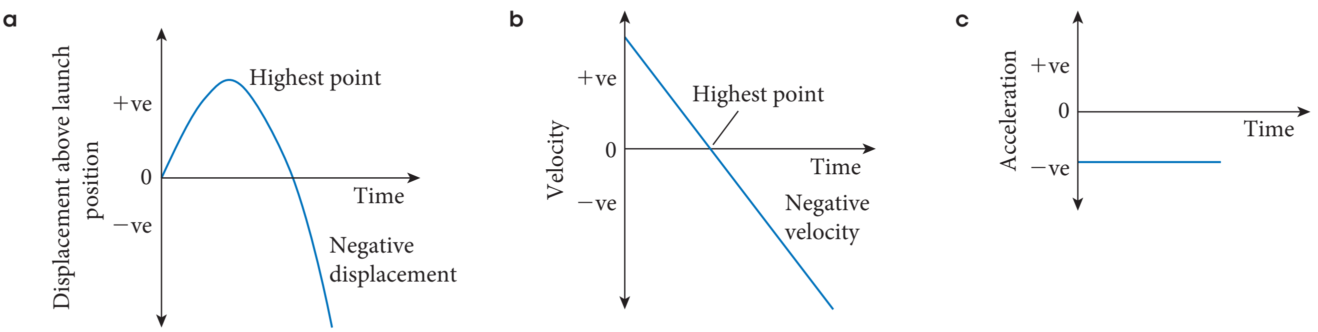
Jan stands exactly 5.0 m vertically below a window. She throws a set of keys vertically upwards so that the keys are stationary when level with the window.

a How long will the keys take to reach the window? (3 marks)

b With what speed should Jan throw the keys? (3 marks)

In cases where the object is thrown upwards from the 'origin' and then continues to fall below this point, its final displacement will have a negative value. This will be the distance below its original launch position. Of course this is only if up is taken as positive and down is taken as negative.

The displacement, velocity and acceleration versus time graphs for such an example are shown in Figure 7.21.



▲ **Figure 7.21**
a) Displacement, b) velocity and c) acceleration for an object launched vertically upwards that falls to a point below its origin or launch position.

EXPERIMENT 7.3

LAUNCH VELOCITY

The time of flight of a projectile can be used to calculate its initial velocity if other variables are known.

Aim

To find the initial vertical velocity of a launched projectile

Materials

- apparatus that will launch an object vertically
- video camera or similar
- ruler or measuring tape

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
Some objects and launch speeds may pose a risk to eyes and faces.	Wear safety glasses. Do not put your face over the launcher.

Procedure

1 Arrange the apparatus so that the object is launched vertically from the edge of a desk or bench top, and lands on the floor beside the desk, as shown in Figure 7.22.

2 Measure s, the displacement of the object from its launch position.

3 Launch the object and record its motion until it hits the ground.

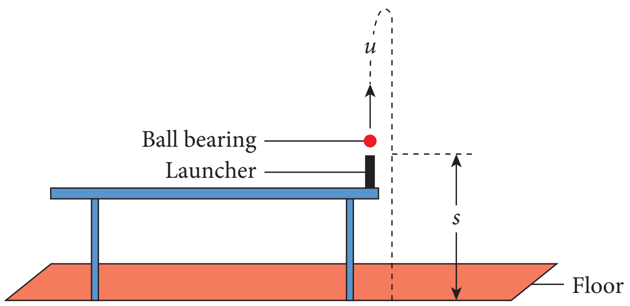


Figure 7.22 Experimental set-up

Results

1 Use the recording to produce s–t, v–t and a–t graphs for the motion.

2 Repeat the experiment for a different launch speed but same final displacement.

Analysis of results

1 Describe the main features of the data collected.

2 Estimate the uncertainty in each of the data points on your graphs.

Discussion

1 Find the accepted value of g and its uncertainty at your latitude.

2 Did your experiment confirm the accepted value of g at your latitude, within the uncertainty of your data?

3 From the data, give a launch value for the launch speed.

Case study

Dr Brian Schmidt – Australian Nobel Prize winner

Dr Brian Schmidt was awarded the 2011 Nobel Prize in Physics for the discovery of the accelerating expansion of the universe. Dr Schmidt and his colleagues, Adam Riess and Saul Perlmutter, who were also awarded the Nobel Prize, used the light from supernovae to measure the distances and speeds of recession of distant galaxies. This relates to Hubble's law, which shows that the distance to far distant galaxies can be shown to be proportional to the speed of recession:

$$v = H_0 d$$

where $H_0 = 70\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$.

Cosmologists wonder about the fate of the universe. Will it go on steadily forever, expand gradually forever, or collapse back on itself into a 'big crunch'? Dr Schmidt's team's discovery came as a surprise – how could the universe be expanding faster and faster? What unknown force was involved here? It is now thought that dark energy accounts for about 70% of the universe.


Newspix/Adam Knott

▲ Figure 7.23
Dr Brian Schmidt

Dr Schmidt is currently leading a team of astronomers on the SkyMapper project. This endeavour is using a telescope owned and operated by the Australian National University in Canberra. The telescope itself sits on a 1100 m high mountain in the Warrumbungle Ranges in northern New South Wales. The project aims to map the entire southern sky using the wide-field telescope and a 268 megapixel camera in wavelengths from the infrared through to the ultraviolet.

Dr Schmidt completed his PhD at Harvard in 1993 at the age of 26. He has been involved in many astronomical fields over the years and continues to be excited by the discoveries being made with new technology and computers that are changing the nature of astronomy. For example, the SkyMapper telescope is operated remotely, with the data being stored at the ANU's supercomputer facility and being made available to the public. Not long ago, observations were painstakingly made and recorded on a glass photographic plate by astronomers who worked all night and slept during the day.

Questions

1 How did Dr Schmidt and his team measure the distances to other galaxies?
2 Use Hubble's law to show that distance *d* is measured in units of megaparsec. What is the length of this unit in km?
3 What is believed to make up 70% of our universe?
4 Why was Dr Schmidt's team's discovery about the expansion of the universe surprising?
5 Why do you think that it is normal now for a team of scientists to be awarded the Nobel Prize rather than an individual?
6 Find out the reasons why the SkyMapper telescope is placed on top of a mountain.
7 Discuss with your classmates whether you think the funding for projects such as SkyMapper is worthwhile. To be able to make an informed decision you will need to undertake some research into the SkyMapper project.

INVESTIGATION 7.1

WATER ROCKET

How high can you make a water rocket rise vertically?

What is your aim?

Your aim should indicate what you intend to measure and what you intend to do with the data. For example, you need to aim to measure the height to which the water rocket rises. But can you also find the launch speed and the effect of air resistance on the rocket?

What will you need?

List the materials you need to conduct this investigation. A water rocket would be a good start, but you will also need time- and distance-measuring systems.

What are the risks?

Construct a table similar to the one below. Identify specific risks involved in the investigation and ways that you will manage the risks to avoid injuries or damage to equipment. Have your teacher approve your risk assessment before you start the investigation.

What are the risks in doing this investigation?	How can you manage these risks to stay safe?

How will you carry out your investigation?

Consider separately the launch system and the measuring system. How will you set them up to achieve your purposes?

It will be useful to do some tests first to consider issues such as consistency of launch and measurement of distance and time.

What results will you collect?

Consider how you will measure the speed at the launch site as well as the height to which the rocket rises. How will you measure time and distance simultaneously? It will not be sufficient to measure only the time at maximum height.

How will you analyse your results?

What graphs will you plot? How could you use these to estimate the launch speed and maximum height?

What have you found?

Display your results.

What do you conclude?

Make sure your conclusion includes quantitative data that relates directly to the measureable quantities you mentioned in your aim. If possible, make an estimate of the uncertainty in your results.

Ideas for improvement or further investigation

Consider the difficulties in getting the data. How could you improve the launch system and the distance and time measures?

QUESTION SET 7.3

1 An object is dropped from a tower 128 m high.
  a What is the speed of the object that falls freely from rest for a distance of 128 m?
  b How long does it take for the object to reach the ground?
  c What is the speed of the object 2.0 s after being released?
  d When is the speed of the object 35 m s^{-1}?
2 A cricket ball thrown vertically upwards remains in the air for 3.0 s.
  a How high does the ball rise?
  b What was its initial velocity?
  c When will the cricket ball have an upwards velocity of 5.0 m s^{-1}?
  d What is the acceleration of the ball at the top of its path?
3 Two objects, A and B, are released from a tower 125 m high. Object A is thrown downwards with an initial speed of 15.0 m s^{-1}, while object B is allowed to fall from rest at the same instant.
  a Calculate the speed of each object on reaching the ground.
  b What is the difference in the time taken for the two objects to reach the ground?
  c How far apart are the two objects after 2.0 s?
4 A parachutist is falling vertically downwards with a constant speed of 4.8 m s^{-1}. When 120 m above the ground, the parachutist drops a small parcel. What is the time difference between the parcel and the parachutist reaching the ground? Ignore air resistance on the parcel.
5 Pelicans tuck in their wings to fall freely when diving for fish. A fish near the surface of the water needs 0.10 s to take evasive action. A pelican 25.0 m above the water starts its dive. The fish first notices the pelican when the pelican is 5.0 m above the water.

  Does the pelican go hungry, or does it catch its prey? Use graphs and calculations to support your answer.

CHAPTER SUMMARY

- Movement is the change of position in a time interval.
- The motion of any object can be modelled as the motion of a 'point mass'.
- All motion is referred to a fixed position, the origin.
- Distance, *d*, is the length travelled, without direction.
- Displacement, $\vec{s}$, is the change in position from one point to another, including direction.
- Speed, *v*, is the magnitude of velocity.
- Velocity, $\vec{v}$, is the time rate of change of position.
- Acceleration, *a*, is the time rate of change of velocity.
- Instantaneous speed, v_{inst}, of an object is the speed at an instant in time.
- Average speed , v_{ave}, is that one constant speed that would allow an object to move with different instantaneous speeds over a time interval to cover the same distance in the same time.

- Instantaneous and average velocity are vectors. They have similar meanings to speed but include direction associated with the motion.
- Instantaneous speed can be found graphically as the value of the gradient of the distance versus time graph at a particular time.
- Instantaneous acceleration can be found graphically as the value of the gradient of the velocity (or speed) versus time graph at a particular time.
- For uniformly accelerated motion:

  $v = u + at$

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

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

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

  where s = displacement, t = time interval, u = initial velocity, v = final velocity, a = acceleration.
- Graphical analysis of motion:

  v = gradient of the displacement versus time graph; $\frac{\text{distance interval}}{\text{time interval}}$

  a = gradient of the velocity versus time graph; $\frac{\text{velocity interval}}{\text{time interval}}$

  s = area under the velocity versus time graph; s is a distance interval

  Δv = area under the acceleration versus time graph
- For vertical motion near Earth, take the acceleration due to gravity, g, to be constant: 9.8 m s^{-2}
- Graphical analysis and algebraic equations for motion under constant acceleration are equivalent representations of projectiles near Earth. Care must be taken when assigning positive and negative values.

CHAPTER GLOSSARY

acceleration, *a* time rate of change of speed (or velocity)

average speed, v_{ave} speed which, if maintained for the entire journey, results in the same distance travelled in the same time

braking distance distance travelled while braking

centre of mass the position in an object at which we measure its position when we model the object as a point-like particle

component the projection of a vector quantity along an axis

deceleration negative acceleration or acceleration in the negative direction when speed is in the positive direction

displacement, $\vec{s}$ change in position

displacement interval, *s* change in displacement

distance, *d* length travelled by an object during its motion

distance interval, Δd or *s* change in the distance

frame of reference settings from which measurements are taken

gravitational acceleration, *g* acceleration of objects due to the gravitational force applied

gravitational force force by one mass on another mass

gravity common name given to the force applied by the very large mass of Earth on relatively smaller masses

instantaneous speed, v_{inst} speed of an object at a moment in time

origin point of reference for measurements of position

reaction distance distance travelled from the time the driver notices a situation and the time the brakes are first applied

reaction time the time taken for a driver to react and apply the brakes from the time a situation is noticed

speed the distance covered in a given time interval

stopping distance the total distance covered from the recognition of the hazard to coming to a stop; it is the sum of the reaction distance and the braking distance

stopping time the total time taken from the recognition of the hazard to coming to a stop; it is the sum of the reaction time and the braking time

time interval, *t* or Δt change in time

vector a quantity that has both magnitude and direction

velocity $\vec{v}$ speed with an associated direction

CHAPTER REVIEW QUESTIONS

Remembering

1 Write down the symbols for acceleration, initial velocity, final velocity, time interval and displacement.

2 Describe the difference between:

  a distance and displacement.

  b speed and velocity.

3 What is the difference between instantaneous and average:

  a speed?

  b velocity?

  c acceleration?

4 Draw vector diagrams to show change of:

  a displacement.

  b velocity.

5 For a velocity versus time graph, what quantity is found by calculating the:

  a area under the graph?

  b gradient?

Understanding

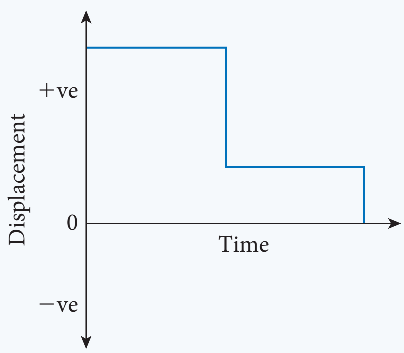
6 Consider how the speed of an object that is thrown upwards changes over time.

  a Describe this in words.

  b Describe it on a *v–t* graph.

7 Explain the positive horizontal line for the acceleration versus time graph shown in Figure 7.19.

8 Explain why Figure 7.24, a position versus time graph, cannot be a completely true graph of an object's actual motion.



◀ Figure 7.24

9 Show that the unit of the area under a velocity versus time graph is the same as the unit of displacement.

Applying

10 In a 100 m sprint race, the winning time is 10.6 s.

  a What was the winner's average speed?

  b Do you think that the runner's average speed was the same as their instantaneous speed during the race? Explain your reasoning.

11 A robot takes three paces forwards and then two paces back, taking 6.0 s for this motion. Use calculations to explain why the robot's average speed is not the same as its average velocity.

12 Find the change in displacement and velocity for the following movements.

  a 100 km, N to 240 km, W in 2.0 h (m s^{-1})

  b 250 km, N45°E to 550 km, N45°W in 88 min (km h^{-1})

  c 350 m, S35°E to 475 m, S20°E in 17.5 min (m s^{-1})

13 Draw the following to scale.

a The change of displacement of a body that moves from 10 m N25°W to 20 m N25°E

b The change of velocity of a body that has an initial velocity of 20 $m\,s^{-1}$, S40°W, and a final velocity of 20 $m\,s^{-1}$, S65°W

14 A fireworks rocket is shot vertically upwards from its launcher. It explodes at its maximum height exactly 3.5 s after launch.

a Sketch the rocket's velocity versus time graph.

From the graph find:

b the initial velocity, *u*, of the rocket.

c the maximum height reached by the rocket.

15 A diver jumps from the 10-metre platform with a vertically upwards speed of 3.0 $m\,s^{-1}$.

a Sketch the diver's velocity versus time graph.

b Use the graph to find the:

i maximum height above the platform of the diver.

ii distance travelled by the diver for the entire dive.

iii diver's displacement.

iv time the diver has to execute his manoeuvres before hitting the water.

16 As a blue car moving at a constant 18 $m\,s^{-1}$ passes a stationary red car, the red car begins to move with a constant acceleration of 3.0 $m\,s^{-2}$.

a Show the motion of the two cars on a velocity versus time graph.

b From the graph find the time when the two cars are next to each other again.

c Check your answer to part b using appropriate equations of motion.

Analysing

17 A builder drops a hammer from scaffolding 25 m above the footpath.

a How fast will the hammer be moving just before it hits the ground?

b How long will it take the hammer to fall?

18 Discuss why using the ground around us is a useful frame of reference for analysing motion even though we know that Earth is rotating on its axis and revolving around the Sun.

19 Sketch a velocity versus time graph that shows the motion of a car that starts from rest and reaches a speed of 15 $m\,s^{-1}$ after the first 5.0 s of motion. It then maintains this speed for 10 s before the brakes are applied and the car stops in 2.0 s. Add appropriate scales to the axes.

20 A rock is thrown vertically upwards with a speed of 8.0 $m\,s^{-1}$, from the edge of a 30.0 m high cliff. From what height would another rock need to be dropped so that the two rocks hit the water below the cliff with the same speed?

21 On page 224, the following is said: '… if the braking distance for a car travelling at 40 $km\,h^{-1}$ is 15 m, the braking distance at 80 $km\,h^{-1}$ increases by 4 times to 60 m.' Use a graphical approach to find the:

a time taken to come to a stop.

b assumed acceleration used in these cases.

c braking time and braking distance when the acceleration is twice the 'comfortable' acceleration when coming to a stop.

22 In an orienteering event, a runner moves from a checkpoint at 200 m, N30°E from the start to another checkpoint at 400 m, S60°E. What is the change of displacement of the runner?

Reflecting

23 Give a qualitative account of how air resistance on a projectile changes its range and maximum height reached.

24 Given that the acceleration of any free body near Earth is vertically downwards, explain why a bullet fired horizontally over level ground will land in the same time as a bullet dropped from the same height. (Hint: Search for *Mythbuster* episode 'Dropped versus Fired Bullet'.)

25 The analysis of the motion of objects moving in a straight line has been done using both graphical and algebraic methods in this chapter. Discuss why such techniques are referred to as 'modelling' the motion of objects.

26 Summarise the main points you have learned in this chapter about linear motion.

Shutterstock.com/Neale Cousland

CHAPTER 8
FORCE

By the end of this chapter you will have covered the following material.

Science Understanding

- Vertical motion is analysed by assuming the acceleration due to gravity is constant near Earth's surface (ACSPH062)
- Newton's Three Laws of Motion describe the relationship between the force or forces acting on an object, modelled as a point mass, and the motion of the object due to the application of the force or forces (ACSPH063)

Introduction

Forces make things move. Forces are external actions applied to objects. Force is not contained within an object. Force can be applied by direct physical contact, **contact force**, or over a distance, **non-contact force**, including through a vacuum. The modern concept of forces and how they affect motion was adopted as recently as a few hundred years ago when Sir Isaac Newton (1642–1727) revolutionised the views of force and motion. He also developed the theory of universal gravitation, published in 1687.

We are able to apply models for force that explain our observations of everyday examples of force and motion. We know now that these models may not necessarily be 'true'. For example, we know that when force is applied to objects moving at close to the speed of light, these models need to be modified. For ordinary motion with which we are usually involved, Newton's and Galileo's models of force will be considered.

Force: actions by one thing on another

The state of motion of an object is affected by forces acting on it. A force is always applied by something (the agent) on something else (the receiver). We use this idea of agent–receiver because students often think, wrongly, that forces are inside objects and that objects somehow contain a single force that cannot be changed.

When objects touch, they apply contact forces. Objects that are not touching can affect each other. This is called 'action-at-a-distance'. The three most obvious non-contact forces are the gravitational, electrostatic and magnetic forces. Every moment, each of us experiences gravitational force – gravity – because the very large **mass** of Earth acts as the agent on our very much smaller mass.

We shall also come to see that our mass acts as the agent on Earth's mass. It depends on the point of view. If we want to know how Earth's mass affects us, we consider the mass of Earth to be the agent. If we want to know how we affect Earth, our mass becomes the agent acting on Earth's mass. Agent and receiver are defined by the object of interest for analysis.

Early ideas about force

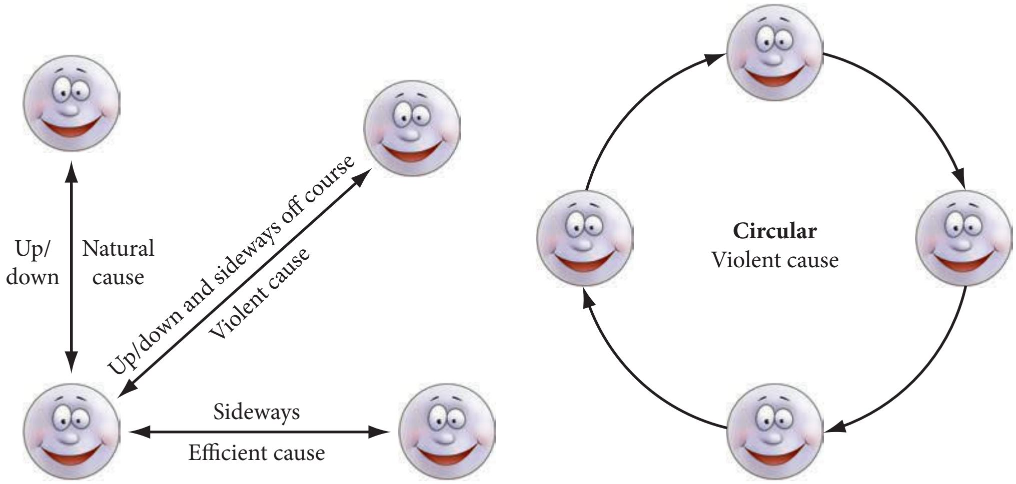
Our modern concept of force was developed over thousands of years. Careful thought, observation and measurements led to theories of what force was and why things moved.

Aristotle

For nearly 2000 years, the Greek philosopher Aristotle's (384–322 BCE) writings were taught in schools and universities. Aristotle's ideas were debated and refined during this time, but they generally stood up well to this scrutiny. Their significance was such that, eventually, they became fundamental to the understanding of the natural world.

The **Aristotelian** physics of motion included **efficient cause**, the external force that was required to make things move in a horizontal direction, and the **violent force**, which caused things to be moved off their natural course and to move in circles. Objects underwent **natural motion**; that is, the motion that they naturally do of themselves due to their **natural cause**. Smoke, for example, rose because it was made from air, so it naturally moved towards the air. A rock, on the other hand, would drop because it was made of earth and that is where it naturally moved.

Aristotle taught that if an object has a force applied to it in one direction and there was a smaller resistive force opposing the motion, the object would continue to move at a constant speed. People still use the idea that a *net force* on an object would result in the object moving with a constant speed to explain motion but they are wrong to do so.



◀ **Figure 8.1**
Aristotle's ideas about motion. Some things naturally go up, others naturally go down. Efficient force makes things move horizontally. Violent force makes things move off their natural course and to move in circles. Beyond the Moon, things were different. These ideas are no longer considered to be correct.

Towards modern thinking

Significant criticism of Aristotle's views on force and motion began to gain momentum during the 15th and 16th centuries. Nicolaus Copernicus (1473–1543), Johannes Kepler (1571–1630), René Descartes (1596–1650), Galileo Galilei (1564–1642) and Isaac Newton were some of the more notable astronomers and mathematicians to cast doubt on Aristotle's views.

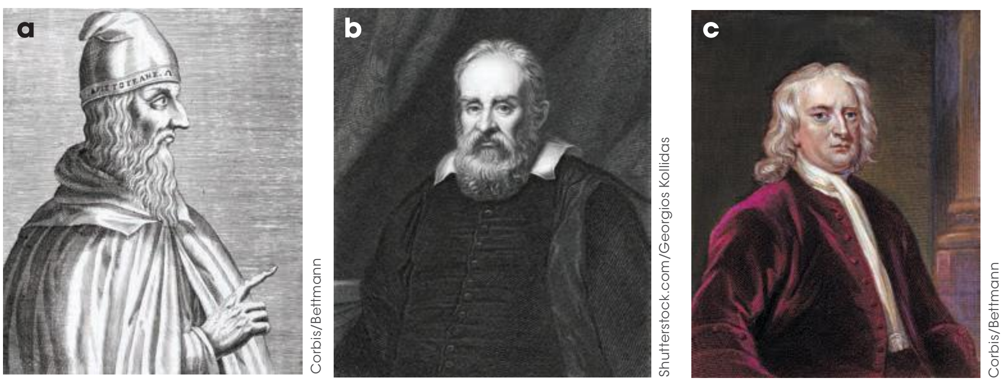
The Aristotelian natural motion of objects was replaced with the newer concept of forces acting externally on objects, whether the object was on Earth or in the heavens. Forces were no longer contained within objects. **Non-zero net forces** resulted in acceleration rather than motion with constant speed.

Galileo

Galileo was an Italian astronomer, mathematician and scientist. He studied the motion of objects, including balls rolling down an inclined plane. He was able to show that the speed of falling objects increased by the same amount in equal time intervals. This acceleration was constant.

Newton

Newton spent much of his life at Cambridge University in England. Born in the year in which Galileo died, Newton was able to pull together the work of earlier scientists. His three laws of motion, along with his work on the importance of gravity in controlling the motion of the planets, are used today for explaining common examples of motion.



◀ **Figure 8.2**
a) Aristotl; b) Galileo; c) Newton

Contact forces

Everyday pushes and pulls are examples of contact forces. In Figure 8.3, a pair of **balanced forces** act on a single object. The two forces are equal in magnitude and opposite in direction. They add up to zero force applied to the object. The sum of forces applied to a single object is the **net force** on the object. The net force on object A is written as $\Sigma\vec{F}$ (on A). Net force means the sum of all the agents on the one receiver.

NEWTON'S LIFE
This site has interesting information on the life of Sir Isaac Newton.

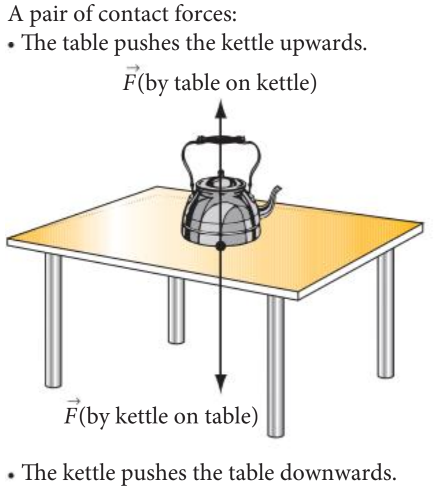


Figure 8.3 ▲ The kettle contacts the table, so the kettle acts on the table: $\vec{F}$(by kettle on table). Clearly, the table contacts the kettle, so the table acts on the kettle, too: $\vec{F}$(by table on kettle)

Such a balanced pair of forces does not result in motion as the kettle is not originally moving. In every situation, an object A is applying a force to another object B. This is written as $\vec{F}$(by A on B). At the same time, object B will be applying a force on object A, so that $\vec{F}$(by B on A) is equal in magnitude but opposite in direction to $\vec{F}$(by A on B).

In Figure 8.3, $\vec{F}$(by kettle on table) is equal in magnitude and opposite in direction to $\vec{F}$(by table on kettle). As the kettle is in contact with the table, this is an example of a contact pair of forces. The kettle is acting as the agent by applying a force on the table, which is the receiver of this force. The other force in this pair of forces is the table, now the agent, applying a force on the kettle, now the receiver: $\vec{F}$(by table on kettle).

Other contact forces include **friction**. Friction occurs wherever two objects move over or slide over each other. Air friction occurs in the interaction between a cyclist and the air. It is why aircraft will climb to high altitudes where the air is less dense and causes less resistance to the aircraft's motion. Again, we consider that $\vec{F}$(by air on cyclist) is equal in magnitude and opposite in direction to $\vec{F}$(by cyclist on air).

Forces act on objects

It is never correct to state that an object 'has a force'. The object applies a force to another object.

When object A acts on object B, we will describe this as $\vec{F}$(by A on B) throughout this book. Figure 8.4 shows an example in which two people are pushing on each other.

An alternative perception of the way forces act is to consider that object B is acted upon by object A. We could describe this as $\vec{F}$(on B by A). Some people prefer to think of forces this way. However, it does not really matter which way the force is written, as long as the naming convention is used consistently. There are two important points underlying the naming convention:

1 *Forces act externally on objects.*

2 There is always an *agent* that acts on a *receiver*.

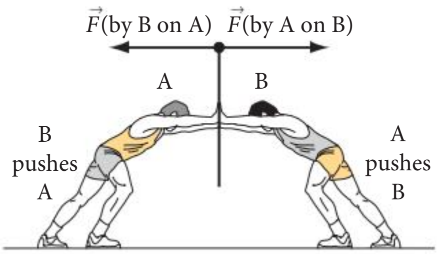


Figure 8.4 ▲ Person A contacts person B, so A acts on B: $\vec{F}$(by A on B). Notice that, at the same time, B contacts A, so B acts on A: $\vec{F}$(by B on A).

Non-contact (action-at-a-distance) forces

Some pushes and pulls occur when the agent and the receiver are not in contact. A magnet can pull another magnet towards it from a distance. It can also push another magnet from some distance away. The magnetic force is a non-contact force (action-at-a-distance force). Again, whenever a force is being applied by object A on object B, then object B is also applying a force on object A.

Electrostatic forces occur when two charged objects are placed near each other. A typical example is the charge between a plastic comb held near, but not touching, dry hair. The comb acts as the agent that acts on the hair, which is the receiver: $\vec{F}$(by comb on hair). Simultaneously, the hair acts as the agent that acts on the comb, which has become the receiver from this viewpoint: $\vec{F}$(by hair on comb).

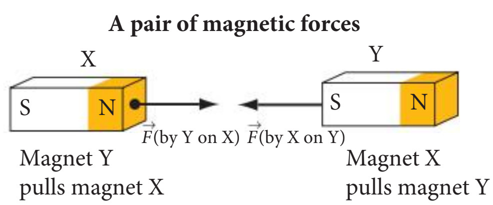


Figure 8.5 ▶ The magnetic force by magnet X on magnet Y acts on Y from a distance: $\vec{F}$(by X on Y) is a non-contact force. Similarly, the magnetic force by magnet Y on magnet X acts on X from a distance: $\vec{F}$(by Y on X) is also a non-contact force.

WOW

Non-contact forces and trains

The fastest trains in the world use non-contact forces to levitate above the track. Superconducting magnets and electromagnetic drive systems means that there is no contact between the train and track. Speeds of more than 400 $km\,h^{-1}$ are regularly achieved in these 'maglev' trains between Shanghai and the city's airport. Such a train could, in theory, travel between Sydney and Melbourne in 2 hours. Presently, this is a 10-hour car trip or a 1-hour plane flight.

MAGLEV TRAINS

List the reasons why maglev trains will eventually replace conventional trains.

Describing forces

A force is not some property of an object. A force is always applied by one object on another. We shall develop a way of modelling the action of one object on another so as to avoid any potential confusion.

Drawing forces

Arrows can be used to represent forces in diagrams. Such arrows have been used in Figures 8.3–8.5. The tail of the arrow always starts on the point of application of the force, with the arrowhead pointing away from the point of application. The length of the arrow on the diagram is used to depict the relative size or **magnitude** of the force.

WORKED EXAMPLE 8.1

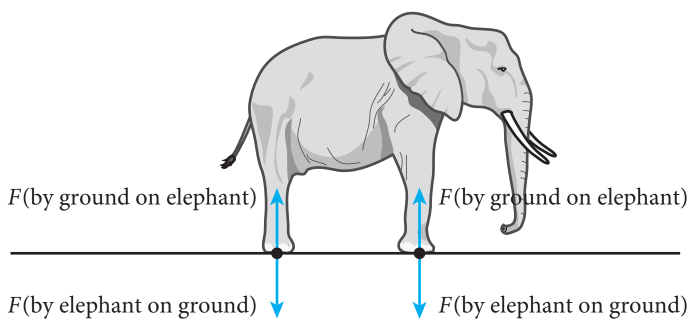
For the following situations, draw a diagram with vector arrows to show the action–reaction (Newton 3) pairs of forces. Label the forces with the correct nomenclature, F(by A on B).

a Contact forces for an elephant standing on the ground (3 marks)

b A person walking a dog on a taut lead (forces involving the lead) (3 marks)

Answers

a

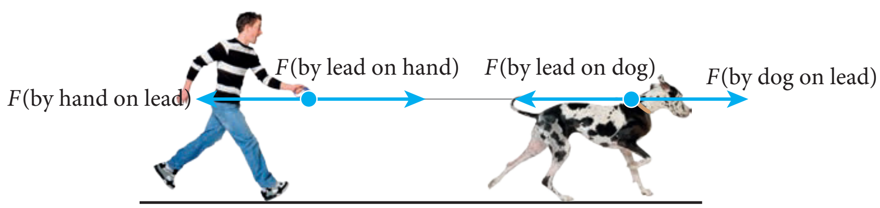


Logic

Draw arrows to scale (approx.) and in right directions as pairs. 2 marks

Label forces correctly. 1 mark

b



Draw arrows to scale (approx.) and in right directions as pairs. 2 marks

Label forces correctly. 1 mark

Try this yourself (2 marks)

Draw a diagram to show all the pairs of forces, including friction (water resistance) acting when a single line from a tug boat pulls on a ship. Label the forces with the correct nomenclature, $\vec{F}$(by A on B). Assume all forces act horizontally.

QUESTION SET 8.1

Remembering

1 How did Aristotle explain the motion of smoke rising from a fire?

2 What effect did Galileo find that the mass of a falling object has on its rate of acceleration?

3 A child pulls their toy truck along the ground by pulling on a string. Identify the agent and the receiver of the force at the point where the string is attached:

a to the truck.

b at the child's hand.

Understanding

4 Imagine that Galileo dropped two balls with different masses from the Leaning Tower of Pisa. What would he have been trying to demonstrate?

5 Why is the notation $\vec{F}$ (by A on B) used rather than F when describing a force on an object?

Applying

6 Using the correct notations, draw a diagram showing the forces acting on a dancer:

a while pushing on the floor.

b when in the air.

7 Magnet X is brought close to magnet Y, as shown in Figure 8.6. Magnet Y then begins to move. Copy Figure 8.6 and on the diagram add arrows showing the forces acting on the magnets. Label the arrows with the correct notations for the forces using arrow conventions.

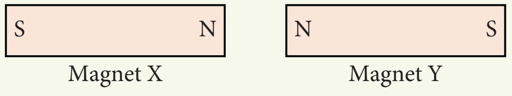


Figure 8.6 ▲

Analysing

8 Outline major differences between the explanations of Aristotle and modern day explanations of the motion of objects. Refer particularly to differences between explanations of the continued motion of a pushed object.

9 Two blocks, A and B, are touching each other. A third object, C, applies a force to the right as shown (Figure 8.7).

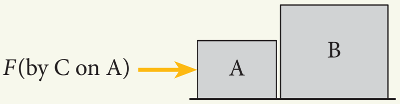


Figure 8.7 ▲
Force applied to two blocks in contact when a third object applies an external force

a Explain why the acceleration of the two blocks, block A and block B, is the same.

b Redraw Figure 8.7 to show:

i F(by A on B).

ii F(by B on A).

c Write an equation for the net force on A.

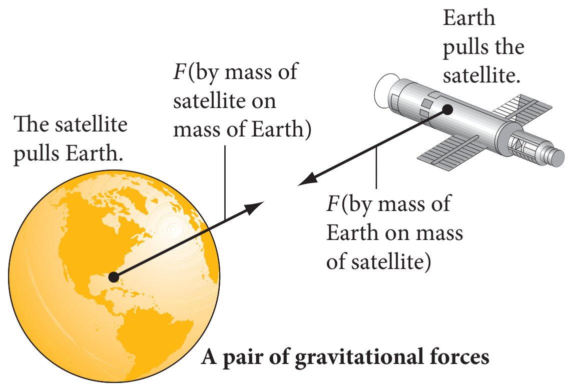
Reflecting

10 Summarise your understanding of forces. What has changed in your thinking about forces?

Gravitational effects on mass

A mass may exert a force on another mass even when they are not in contact. Mass is the amount of matter contained in an object. Ultimately, mass relates to the number and type of atoms in the object.

The attractive force by one mass on another is called the gravitational force. When Earth's mass attracts our mass, we tend to shorten this term to the one word, gravity. Gravity is the force that keeps the Moon and artificial satellites in orbit around Earth. If one mass can exert a force of attraction on another mass, it is correct to say that our mass attracts the mass of Earth and that the mass of a satellite attracts the mass of Earth (Figure 8.8).



▲ Figure 8.8
The gravitational force by the mass of Earth on the mass of the satellite acts on the satellite from a distance. It is a non-contact force. Similarly, the gravitational force by the mass of the satellite on the mass of Earth is also a non-contact force.

Weight

Weight is the force applied to a mass by another mass. On Earth, the term 'weight' is used to mean the gravitational force applied by the mass of Earth on a much smaller mass such as a person. If an object with a particular mass were to be taken to another planet or the Moon, the amount of matter in the object would not change, but the gravitational force applied by the planet or the Moon on that mass would be different. This results in the object having a gravitational force applied to it that is different from the force that would be applied to it on Earth. Hence, it would have a different *weight force* applied to it. Weight is a force, so it has the unit newton, N.

A person who says they 'have a weight of 60 kg' is confusing two concepts. First, the weight unit referred to is actually the unit of mass (kg). Weight is a force, so its unit is the newton (N). Second, people do not possess 'weight' as they do not contain 'force'. Earth applies a gravitational force on a 60 kg person of about 588 N.

Typical bathroom scales have been calibrated to read the equivalent mass of a person when they are standing on Earth, which is where the scales were intended to be used. Taken to the Moon, these scales, which operate by measuring the compression of a spring, would show that a person's mass is about one-sixth that on Earth. Clearly, a person does not lose five-sixths of their matter just by travelling to the Moon.

WOW

Extreme skydiving

In October 2012, Felix Baumgartner, an Austrian daredevil, jumped from a balloon at more than 30 km above Earth. Due to the low density of the atmosphere, he wore a space suit. He fell faster than the speed of sound before slowing down and landing safely. For part of the fall he travelled at terminal speed, where the upwards and downwards forces balanced out. The whole fall took about 15 minutes.

AAP Images/EPA

▲ Figure 8.9
Felix Baumgartner jumping from the balloon more than 30 km above Earth

BAUMGARTNER AND THE SPEED OF SOUND

Did Baumgartner break the speed of sound?

TERMINAL VELOCITY

Find out how a cat falling out of window can land on all fours.

Terminal speed

When the forces on a falling body are balanced, the net force is zero. The acceleration becomes zero. The falling body then travels at constant speed. A skydiver, if rolled up into a ball, would fall to Earth at more than 300 km h^{-1} (83 m s^{-1}). With arms and legs spread out, this speed is reduced to around 200 km h^{-1}, or 55 m s^{-1}. A skydiver who jumps from a plane at an altitude of 3000 m can freefall for about 40–50 seconds before opening the parachute and slowing to a more sedate 40 km h^{-1} for a safe landing.

Gravitational effects near Earth

These concepts will be explored more fully in Unit 3.

The mass of Earth, approximately 6.0×10^{24} kg, gives rise to the gravitational force that acts on any nearby object. The gravitational force applies 9.8 N for every one kilogram of mass near the surface of Earth. In a 10 kg mass, all the 1 kg pieces are connected. Each 1 kg piece experiences the same size and direction of force, 9.8 N downwards.

As the 10 kg mass does not disintegrate as it falls, each kilogram must fall at the same rate. The magnitude of the gravitational force is proportional to the mass to which it is being applied. Therefore, any mass, regardless of its size, will accelerate downwards at the same rate.

'Near Earth'

The strength of the gravitational force on a mass decreases as the distance from the centre of Earth increases. Nevertheless, up to at least as high as the tallest building, the strength of the gravitational force can be considered to be constant. For a 1.0 kg mass this value is 9.8 N. For a 10 kg mass this force is $10 \times 9.8 = 98$ N.

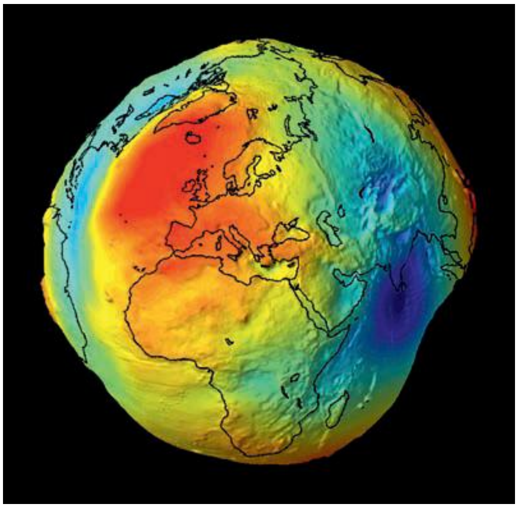
The value of 9.8 N on each 1.0 kg is not exactly the same in all places on Earth's surface for a number of reasons.

- Earth is not a perfect sphere, resulting in the surface at the poles being several kilometres closer to the centre of Earth than the equator.
- Earth's rotational speed causes the measured value of the gravitational force to be slightly less at the equator than at the poles.
- Landforms such as mountains make the surface of Earth irregular, varying the distance from the centre.
- Earth's crust varies in density due to different thicknesses, rock structures and mineral deposits.

A map of the variation in the gravitational force around Earth's surface is called a geoid. Figure 8.10 shows this variation as a three-dimensional interpretation of the globe. The data used to produce this representation come from the GRACE satellites that orbit Earth and map the slight variations in the gravitational field.

GRACE PROJECT

This NASA website has all the background information on the GRACE mission.



Science Photo Library/European Space Agency

Figure 8.10 ▲ A geoid

WORKED EXAMPLE 8.2

An astronaut on the Moon weighs himself on a set of bathroom scales. He declares 'I've lost weight!'. Discuss his statement. (2 marks)

Answer	Logic	
The astronaut's weight has decreased. The force by the mass of the Moon on the astronaut is less than the force applied by the mass of Earth on the astronaut. However, the astronaut did not lose mass. This would involve his body having less matter in it.	Comparison between mass and weight is correct.	2 marks

Try these yourself

1 While moving at constant speed, on a direct trip from Earth to the Moon, an astronaut stands on a set of bathroom scales. At one point the scales read zero. In what sense is the astronaut 'weightless'? (2 marks)

2 Are you really 'weightless' in freefall? Explain. (2 marks)

QUESTION SET 8.2

Remembering

1 What does the mass of an object measure?

2 What is the key difference between 'weight' and 'mass'?

3 Define 'gravity' in terms of 'gravitational force'.

Understanding

4 Which applies the greater gravitational force: Earth on you or you on Earth? Explain your answer.

5 Qualitatively, how would your weight be different if you weighed yourself:

a 100 km above Earth?

b 100 km above the Moon?

c near the equator if Earth's rotation were 10 h?

Applying

6 In a video, when a skydiver opened her parachute, she appeared to fly up, even though her acceleration decreased.

a Explain why.

b How does the skydiver manage to get to the ground safely?

7 The slogan 'Lose weight – go to the Moon' appears in a weight-loss advertisement. Why is this both true and amusing?

Analysing

8 Explain why all masses fall at the same rate near Earth.

9 Why is a beam balance more accurate than a weighing scale that uses a compressed spring (Figure 8.11)?

Reflecting

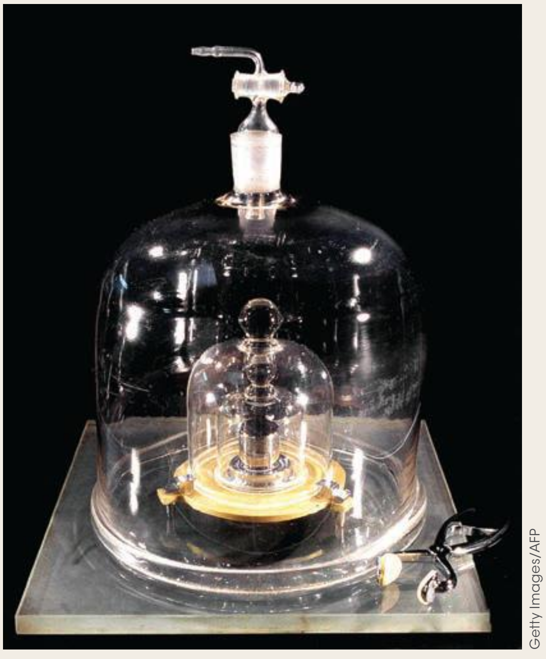
10 How has your understanding of mass, weight and the way different masses fall near Earth developed so far?

▲ Figure 8.11
A mechanical balance is used to measure the mass of materials accurately.

Scientific literacy: Weight of the world

Once a year, three officials bearing three separate keys meet at the bottom of a stairwell at the International Bureau of Weights and Measures, in Sèvres, France. There they unlock a vault to check that a plum-size cylinder of platinum-iridium alloy is exactly where it should be. Then they close the vault and leave the cylinder to sit alone, under three concentric bell jars, as it has for most of the past 125 years.

This lonely cylinder is the international prototype of the kilogram, known colloquially as Le Grand K, and it is the last remaining physical object to define a unit of measure. It's a quaint throwback to a time when people compared the ocean's depth to the span of a man's outstretched arms and the second to a tiny fraction of a year. Now we fix our rulers to the speed of light and our clocks to a spectral property of caesium. By thus linking measurement to a fundamental and unchanging phenomena, scientists have paved the way for GPS satellites, gravity-wave detectors and many other precision technologies that simply wouldn't have been possible before.



▲ **Figure 8.12**
Le Grand K

The trouble posed by the master kilogram is apparent in the many friction-filled steps by which it calibrates other masses. Once every few decades, a scientist plucks the cylinder from its perch with chamois-leather-padded pincers, rubs its surface with a cloth soaked in alcohol and ether, and steam-cleans it. Then he puts the prototype in a precise balance that compares it to the bureau's official copies, which are in turn compared to copies kept by member countries. And thus, the prototype's mass trickles down to set the standard for the rest of the world.

The system has been far from seamless. When the cylinder was last removed from the vault in 1988, the bureau's metrologists were disappointed to discover that its mass and those of its official copies had drifted apart by as much as 50 micrograms (μg) since 1889. That discrepancy is tiny – comparable to the mass of a small grain of sugar – but it confirmed a troubling instability. All that metrologists can say is that the master kilogram seems to have lost as much as 50 μg over the course of a century relative to its siblings. But the actual drift could be up or down, and it might even be a lot more than 50μg, because the prototype and its metallurgically identical copies could all be changing as an ensemble.

'It's a bit ridiculous in this day and age, because it's not just the mass that depends on the prototype. It's all energy, all force, all units that are linked in any way to the kilogram,' retired metrologist Terry Quinn explains one grey afternoon at the bureau, about a week before an international committee was to convene to decide the fate of the kilogram.

A former director of the bureau (often referred to by its French acronym, BIPM), Quinn has been campaigning since the early 1990s to peg the kilogram to an unchanging aspect of nature. This would be a boon to scientists who depend on stable units to perform long-term measurements, he says.

Courtland, R. (2012) 'The kilogram, reinvented'. *IEEE Spectrum,* 30 April.

Questions

1 What is the colloquial name for the prototype kilogram?

2 Why is it important for the prototype kilogram to be handled so rarely?

3 The standard kilogram has changed by as much as 50 μg relative to copies made of it.

   a How could this happen?

   b Why is it useful to know what causes this discrepancy?

4 Explain why other units are linked to the unit of mass, the kilogram.

5 Analyse the importance of having an unchanging, defined measure of the unit of mass.

6 What would have been the value of having a known mass to represent the kilogram in bygone days? What could the modern-day alternative to Le Grand K be? Explain the advantages that this could bring.

Newton's laws

Newton's laws of motion are important explanatory tools to help us understand motion. They are used to help organise our thinking. They cannot be properly understood simply by accepting them as 'fact'. Many people still organise their thoughts, deep down, in Aristotelian terms. So, Newton's laws are not necessarily 'obvious' to everyone. They cannot be understood simply by remembering a few formulas. A deeper level of 'knowing' is needed. In this section, your thinking and understanding of the nature and causes of motion, and their explanation, will be challenged. Think carefully about how Newton was able to so neatly sum up so much physics within these laws.

MEASURING GRAVITATIONAL AND INERTIAL MASS

This site shows the different methods of measuring the mass of objects. Relate this to what you have read in this chapter.

Newton's first law: the law of inertia

Newton's version of the first law was developed from the work of Galileo. In translation, it reads: 'Every body continues in its state of rest, or of uniform motion in a right (straight) line, unless it is compelled to change that state by forces acting on it'.

> **Newton's first law: If the net force acting on an object is zero, then its velocity remains constant.**

The term **inertia** is used to describe the tendency of an object to remain at rest or continue to move at constant velocity. It is a kind of resistance to the change of motion. Inertia and inertial mass are the same thing. When a net force is applied to an object, the object will more or less resist the change depending on its **inertial mass**. This is demonstrated by the acceleration of the object.

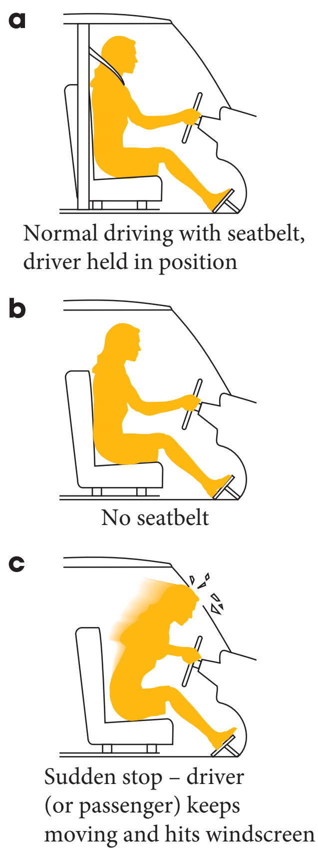
> **Inertia, or inertial mass, of an object is measured by applying a force to it and measuring the acceleration.**

Before Galileo and Newton, it was believed that a force must be maintained on an object to keep it in a state of motion. Forces that oppose motion – friction and air resistance – were not considered. Many still hold this belief, and if friction as a force is ignored, it is easy to see why. To keep a car moving along level ground at constant speed, a constant force must be applied. To keep a bicycle moving at constant speed, a gentle yet continuous pedalling effort is needed. An initial push will result in motion, which soon stops. Both Galileo and Newton recognised that, if there were no **friction forces** opposing the motion, the same initial push would result in continuous and constant motion.

Examples of Newton's first law

An ice skater will continue to move over the ice until a force causes the skater to stop. This force could be exerted by the wall of the ice rink on the skater when the skater runs into it. It could be the force by another skater on the person as they collide. It could also be the friction force by the ice on the blades when the skater digs the blades of the skates into the ice in an effort to slow down.

When travelling in a car, a seatbelt is worn for safety. In the event of a sudden stop, the seatbelt keeps the occupant connected to the car so that they come to a stop with the car. If this were not the case, the occupant would continue with their motion at constant velocity. At 60 km h^{-1}, an unrestrained person will continue towards, and hit the windscreen. The force applied by the windscreen to the head will cause the skull to be crushed by about 3.5 cm.



▲ **Figure 8.13** a) A driver wearing a seatbelt stops with the car. b) and c) An unrestrained driver continues on at the same speed as the car just before the crash begins. The collision with the windscreen can be disastrous.

WOW

The consequences of brain injury

The front of the brain contains the part that helps you control your emotions and behaviour. If this is damaged by being smashed into a car windscreen or dashboard, you are likely to suffer permanent brain damage. This shows up as mood swings, self-centredness and inflexible thinking. You will tire more quickly from mental exertion, have difficulty solving problems and experience short-term memory loss. Frequently, people with this form of acquired brain injury become socially isolated because they can no longer maintain relationships or hold down a job.

BRAIN INJURIES FROM CAR CRASHES

About 7000 Australians suffer from traumatic brain injury as a result of car crashes every year. This site contains further reading, including the stories from survivors.

When you are in a car travelling fast around a corner, you may feel as if you are being thrown sideways out of the car. This is what you feel, but Newton's explanation is interested in the forces pushing you off a straight line, not how you feel about it. Newton's first law says that you will tend to keep moving with constant velocity; that is, straight ahead. However, the car seat pushes you off this straight-line motion. You will feel pushed off your line, which is indeed what is happening. The car itself can only go around the corner because the sideways force by the road on the tyres (friction) causes the tyres to change direction – the car changes velocity. If the friction was not sufficient, the car would continue in a more nearly straight line. The car could slide off the road or fail to take the bend. The friction force depends on complex interactions between tyres, road surface and speed.

Net force

Recall that the sum of all the forces acting on an object at the same time is called the net force. The net force on object A is written in vector form, with an arrow on top, as $\Sigma\vec{F}(\text{on A})$. This sum must take into account the directions of all the forces applied to A. The net force has a direction as well as a magnitude. Net force is a geometric sum of all forces.

A vector is a quantity that has both magnitude and direction. Force is a vector. If we only consider the magnitude of a force, we cannot predict its effect – we also need to know its direction. The effect of a force of 100 N up on an object is different from the effect of a force of 100 N down.

If the forces acting on A are balanced, there will be no change to the state of motion of object A. Object A will either continue with constant velocity or it will remain at rest. Remaining at rest is simply a special case of maintaining constant (zero) velocity.

Newton's second law

A non-zero (unbalanced) net force acting on object A that has a mass *m* will cause the object to accelerate at a rate *a*. The rate of acceleration produced is proportional to the net force. This can be represented as:

$$\vec{a}_{\text{A}} \propto \Sigma\vec{F}(\text{on A})$$

For a given particular force, if the mass is large, then the acceleration is small. Conversely, if the mass is small, the acceleration is large. That is:

$$\vec{a} \propto \frac{1}{m}$$

Combining these observations gives:

$$\vec{a} \propto \frac{\Sigma\vec{F}(\text{on A})}{m}$$

$$\Rightarrow \vec{a} = k\frac{\Sigma\vec{F}(\text{on A})}{m} \quad \text{where } k \text{ is the constant of proportionality.}$$

The value of $k = 1$, when SI units are used. A 1 N net force acting on a 1 kg mass results in an acceleration of the mass of 1 m s^{-2}. Further, the acceleration occurs in the same direction as the net force.

For balanced forces:

$$\Sigma\vec{F}(\text{on A}) = 0$$

If the forces on A are unbalanced, then the state of motion of A will change. Either the speed or direction will change, or both speed and direction will change.

For non-zero net forces:

$$\Sigma\vec{F}(\text{on A}) = m_A\vec{a}_A$$

Along a straight line the direction is either positive or negative. In this case, we do not need to use the arrow to represent direction, so we can write Newton's first law as:

$$\text{for non-zero net forces, } \Sigma F(\text{on A}) > 0 \text{ or } \Sigma F(\text{on A}) < 0$$

Newton's second law in its more familiar form comes from making force the subject of the equation:

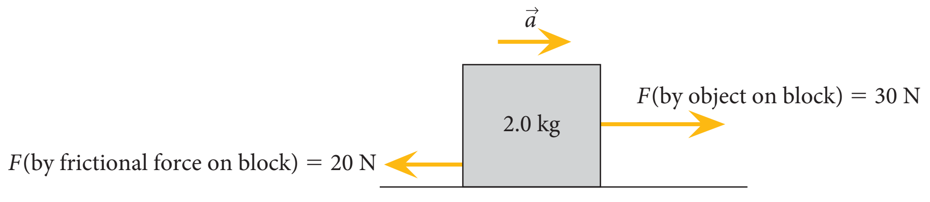
$$\Sigma\vec{F}(\text{on A}) = m\vec{a}$$

The acceleration ($\vec{a}$) of an object is dependent on the sum of all forces acting on the object ($\vec{F}$(on object)) and the mass (m) of the object: $\vec{a} = \dfrac{\Sigma\vec{F}(\text{on object})}{m}$.

WORKED EXAMPLE 8.3

1 A force of 30 N to the right is applied (by some other object) to a 2.0 kg block. A frictional force (by a surface on the block) of 20 N opposes the motion. This is shown in Figure 8.14.

What is the acceleration of the block? (4 marks)



▲ Figure 8.14

2 A car with a mass of 1000 kg is travelling at a speed of 20 m s^{-1}. Calculate the average net force applied to the car if it is to stop in a distance of 25 m. (4 marks)

3 An aeroplane with a mass of 8.0×10^4 kg accelerates from rest to take-off speed of 75 m s^{-1} in 25 seconds. What is the average net force on the aeroplane during this time? (4 marks)

Answers	Logic	
1 $a_{block} = \dfrac{\Sigma\vec{F}(\text{on block})}{m_{block}}$	Select the correct formula.	1 mark
$= \dfrac{30\text{ N} - 20\text{ N}}{2.0\text{ kg}}$	Substitute correct values.	1 mark
$= 5\text{ m s}^{-2}$	Calculate the answer.	1 mark
The resultant of the forces is 10 N to the right, so the object will accelerate at 5.0 m s^{-2} to the right.	Give correct answer and direction.	1 mark

2 $s = 25\,\text{m},\ u = 20\,\text{m s}^{-1},\ v = 0,\ a = ?$ — Select the correct equation. — 1 mark

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

$2as = v^2 - u^2$

$$a = \frac{v^2 - u^2}{2s}$$

Rearrange the equation to make a the subject. — 1 mark

$$= \frac{(0\,\text{m s}^{-1})^2 - (20\,\text{m s}^{-1})^2}{2(25\,\text{m})}$$

$$= -8.0\,\text{m s}^{-2}$$

Now, $\Sigma\vec{F}(\text{on car}) = 1000\,\text{kg} \times -8\,\text{ms}^{-2}$

$= -8000\,\text{N}$ — Calculate the answer. — 1 mark

Therefore, the average braking force is 8000 N in a direction opposite to the direction of motion of the car. — Find the correct force. — 1 mark

3 $v = 75\,\text{m s}^{-1},\ u = 0,\ t = 25\,\text{s},\ a = ?$ — Select the correct equation. — 1 mark

$v = u + at$

$at = v - u$

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

Substitute the correct values. — 1 mark

$$= \frac{75\,\text{m s}^{-1} - 0\,\text{m s}^{-1}}{25\,\text{s}}$$

$$= 3.0\,\text{m s}^{-2}$$

Now, $\Sigma\vec{F}(\text{on plane}) = 8.0 \times 10^4\,\text{kg} \times 3.0\,\text{m s}^{-2}$

$= 2.4 \times 10^5\,\text{N}$ — Calculate the answer. — 1 mark

in the direction of the plane's acceleration. — Find the correct force. — 1 mark

Note: You could do this example graphically by drawing a v–t graph to find the acceleration.

Try these yourself

1 A tennis ball of mass 56.0 g is struck by a racquet and accelerated from rest to $50\,\text{m s}^{-1}$ in $6.0 \times 10^{-3}\,\text{s}$. What is the net force applied to the ball by the racquet? (4 marks)

2 A car with a mass of $1.2 \times 10^3\,\text{kg}$ accelerates from rest to $25\,\text{m s}^{-1}$ in 5.0 s. If the same car has a load of 400 kg added to it, how long, under the same net force, would it take to accelerate from rest to $25\,\text{m s}^{-1}$? (Hint: Sketch a v–t graph to find acceleration as the gradient.) (4 marks)

Newton's third law

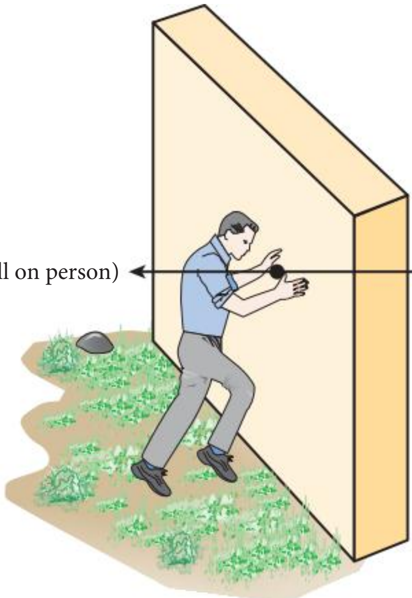
Forces always act in pairs. When two objects, A and B, interact, each object applies a force on the other. Newton showed that this is an obvious consequence of object A acting as agent on object B, $\Sigma\vec{F}(\text{by A on B})$ (agent). Simultaneously, object B, the receiver, acts as agent on A (reaction), $\Sigma\vec{F}(\text{by B on A})$.

The two forces, $\Sigma\vec{F}(\text{by A on B})$ and $\Sigma\vec{F}(\text{by B on A})$ are:

- equal in magnitude
- opposite in direction

AND

- they act on different objects.



◀ Figure 8.15
The pair of action–reaction forces when a person pushes on a wall. The forces act on different things.

When a person pushes on a wall, the wall pushes back on the person. The two forces are equal in magnitude but opposite in direction.

Action–reaction pairs of forces always act on different objects. If object A acts on object B, then object B must also act on object A. As shown in the example in Figure 8.15, $\Sigma\vec{F}$(by A on B) and $\Sigma\vec{F}$(by B on A) are equal in magnitude, but the receiver objects are different.

Two important forces associated with movement on Earth are the gravitational force and the electrostatic force. Gravitational force is the force applied by one mass on another. Electrostatic force is the force applied by one charge on another. When you stand on a surface such as the ground, the electrostatic force pushes up on you from the surface. This is the force $\vec{F}$(by surface on you). The gravitational force by the mass of the Earth pulls down on you. This is the weight force, $\vec{F}$(by mass of Earth on you). These two forces act on you, so they can be used to work out the net force on you: $\vec{F}_{net}$(on you). As the the forces act on different bodies they cannot be a pair of forces in the sense of Newton's third law.

Net force, $\Sigma\vec{F}$ (on receiver)

Which forces do we consider when we find the net force on an object? Because each force in a Newton 3 (action–reaction) pair of forces acts on a different object, these forces cannot be added together. They relate to the net force on different objects. When finding the net force acting on object A, use only the forces acting *on* A: $\vec{F}$(by B on A), $\vec{F}$(by C on A) etc.

The pair of action–**reaction forces** is not part of the calculation of the net force that acts on a particular object because one part of the pair is not acting on the object in question.

Newton's third law pair of forces

$\vec{F}$(by A on B) and $\vec{F}$(by B on A) are:

i equal in magnitude
ii opposite in direction
iii have the same fundamental nature

AND

iv each force acts on a different object.

All four criteria must be satisfied.

They cannot be added to make a net force.

WORKED EXAMPLE 8.4

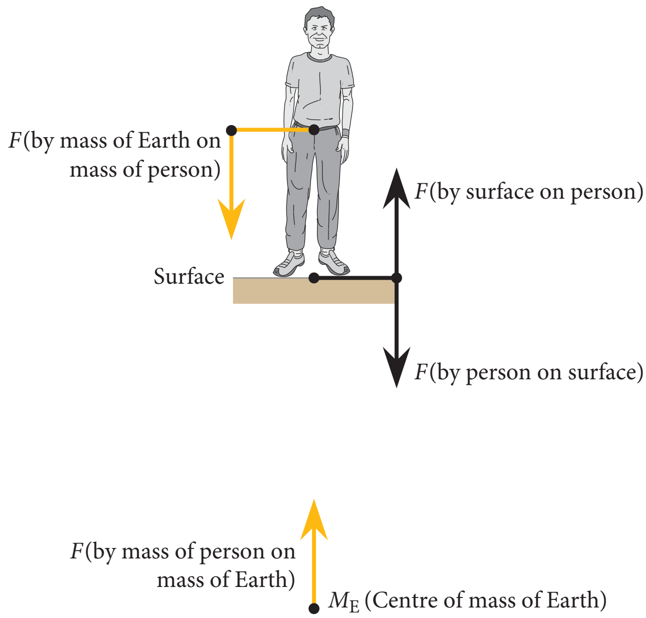
When you jump up you have to push down on a surface. Your downwards force causes you to go up. How is this possible? (3 marks)

Answer

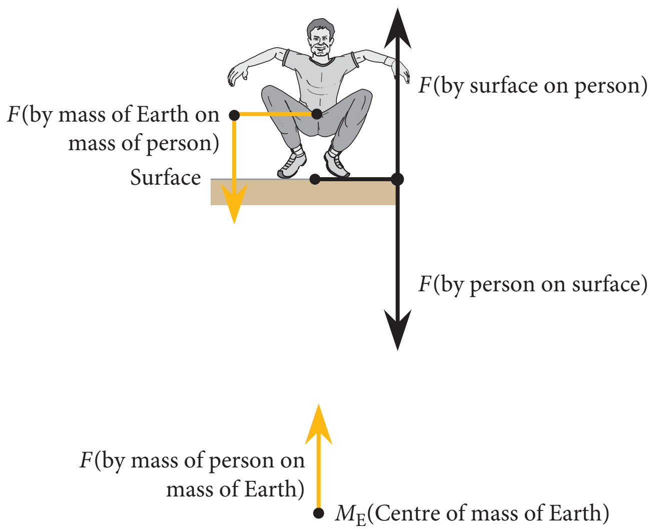
When you stand still, the net force on you is zero. To jump up, the net upwards force must be upwards:

$$\Sigma\vec{F}(\text{on you}) = \vec{F}(\text{by surface on you}) - \vec{F}(\text{by Earth's gravity on you}) > 0$$

Figure 8.16 shows the forces when standing and Figure 8.17 shows the forces when pushing down in order to go up.



▲ **Figure 8.16**
Standing still: $\Sigma\vec{F}(\text{on you}) = 0$



▲ **Figure 8.17**
When you push down, the reaction by the surface on you increases: $\Sigma\vec{F}(\text{on you}) > 0$

Logic

Find the correct net force.	1 mark
Find the correct non-zero net force.	1 mark
Make the correct conclusion.	1 mark

Newton's third law means that no matter whether you are standing or pushing down hard on the surface:

$\vec{F}$(by you on surface) and $\vec{F}$(by surface on you) are equal

When you push down, you increase the force by you on the surface. This, in turn, means that the force by the surface on you increases. It becomes greater than the weight force on you.

$\vec{F}$(by surface on you) > $\vec{F}$(by Earth's gravity on you)

Hence, the net upwards force on you becomes greater than zero:

$\Sigma\vec{F}$(on you)>0

Try this yourself

Explain how it is possible for a space rocket to rise vertically from the launch pad by sending exhaust gases downwards. (3 marks)

QUESTION SET 8.3

Remembering

1 Write:
- a Newton's law of inertia.
- b Newton's second law.
- c Newton's third law. (Refer to the four criteria needed to define this law correctly. Do not use the words 'action' or 'reaction'.)

Understanding

2 Is a force caused by an acceleration? Explain with reference to Newton's second law.

3 What is the magnitude of the net force acting on an aeroplane flying horizontally at a constant speed of 400 km h^{-1}?

4 All forces come in pairs. Explain this in terms of agent and receiver.

5 An elephant pulls on a rope with a force of 500 N. With what force does the rope pull on the elephant?

6 Explain the following with reference to Newton's first law.
- a An object at rest does not move.
- b A person standing on a bus stumbles forwards when the bus stops suddenly.
- c A seated person not wearing a seatbelt is not 'thrown forwards' towards the windscreen in a car crash but they do hit the windscreen.

Applying

7 Two people have a tug-of-war that results in a tie. Sketch a diagram to show all the forces acting on:
- a each person.
- b the rope.

8 When a person leans against a pole, the pole pushes on the person.
- a Explain why the person may not get pushed over by this force from the pole.
- b Explain how the person may be pushed over by the pole.

Analysing

9 A child is pulling a toy truck behind her over level ground at constant speed. How do you know that all the forces acting on the toy truck are in balance?

Reflecting

10 Consider the ways in which Newton's first law is connected with Newton's second law and discuss your thoughts in your group.

Scalar and vector quantities

Some quantities with which we are familiar do not have a direction associated with them. Temperature, mass, speed and volume are examples. Such quantities only have magnitude, or size. They are known as **scalar quantities**. Scalar quantities can be added and subtracted on a number line with positive and negative values.

The analysis of motion and forces in a straight line can be undertaken using a number line. In these cases they can be considered to be scalar. However, when we extend the analysis of motion and forces to two- and three-dimensional motion, a geometric approach is required. We have already seen that distance is scalar, but displacement is a vector quantity. Speed is scalar but velocity is vector. Acceleration can be treated as scalar for straight-line motion but must be treated as a vector when the motion is on a plane or within a three-dimensional space.

Vector quantities

Any quantity or measurement that has both magnitude and direction is a **vector quantity**. Displacement, velocity, force and acceleration are examples in which a direction is associated with the quantity. A plane may be flying at $500\,km\,h^{-1}$ (its speed), but its velocity requires a direction, for example $500\,km\,h^{-1}$ west.

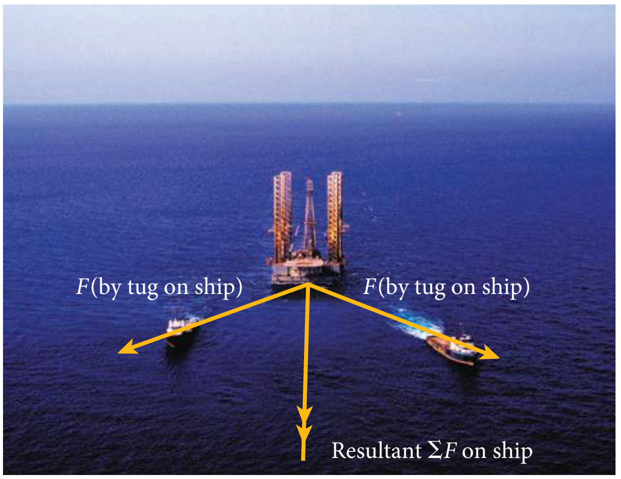
When vector quantities are added together, the resultant, that is the sum of all the vector quantities, will not be the simple arithmetical sum of all the contributing vector quantities. It is a geometrical sum.

Force as a vector

When a force is applied to an object it has a magnitude (the size of the force) as well as a direction. Weight force is applied in a vertically downwards direction. A tug boat pulling a ship applies the force to the ship through a rope. The direction of the applied force is the same as the direction made by the rope.

The ship shown in Figure 8.18 has forces by two tug boats being applied to it. The **resultant force** on the ship is found by adding these forces in a geometric fashion, along with any friction forces involved. The acceleration of the ship is in the direction of the sum of all the forces acting on it, $\Sigma\vec{F}$(on ship). In Figure 8.18, only the forces on the ship applied by the tug boats are shown. The force of friction, in this case being applied by the water against the ship, opposes the relative motion. It is therefore in a direction opposite to that of the motion of the ship. If the ship is moving with constant velocity, the net force on the ship, $\Sigma\vec{F}$(on ship) $= 0$. The direction of motion of the ship is found from the sum of the applied forces acting on the ship. If the ship is not moving relative to the water, no friction force by the water on the ship will be acting.

Figure 8.18 ▼ The combined effect of the tugboats on the ship lies between the two towropes. Friction is not considered, but will act along the line of the resultant vector.



Addition of forces

When forces are acting along a straight line, or in one dimension, using a number line system with positive values for one direction and negative values for the opposite direction is sufficient. However, in two-dimensional cases in which the forces are acting in a plane, the addition of forces can be analysed by using a scale diagram or by drawing a geometric sketch and using trigonometric calculations.

Free body diagrams are used in this chapter to show the magnitude and direction of the forces acting on a body using arrows.

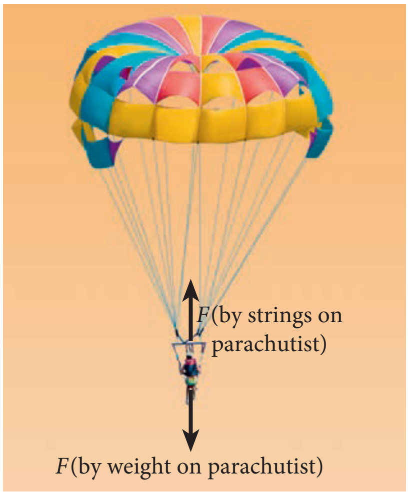
Forces acting along a straight line

Along a straight line, direction can be assigned as either positive or negative. For example, any force acting to the right can be assigned a positive value, while forces acting to the left can be negative. The net force will either be positive or negative.

In Figure 8.19, the sum of all the forces by the strings on the parachutist are shown acting upwards while the gravitational force (weight) acts downwards along the same line. The net force on the parachutist is zero when travelling at terminal (constant) speed.

In the example shown in Figure 8.20, $\Sigma\vec{F}(\text{on box}) = +4.0 + 3.0 = +7.0\,\text{N}$ (positive direction to the right).

In Figure 8.21, the forces act in opposite directions, so $\Sigma\vec{F} = +4.0 + (-3.0) = 1.0\,\text{N}$ (positive direction to the right).



▲ **Figure 8.19** Forces on parachutist. The sum of the forces by the strings on the parachutist are shown as a single, upwards force.

▲ **Figure 8.20** Addition of forces in one direction: 7 N to right

▲ **Figure 8.21** Addition of forces in opposite directions: 1 N to right

Forces acting in a two-dimensional plane

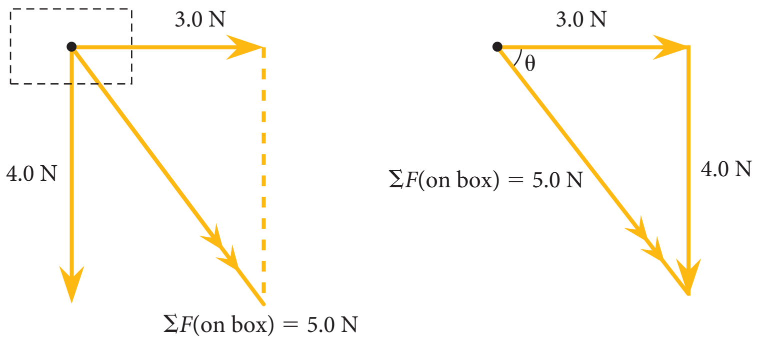
In Figure 8.22, the forces add geometrically to form a right-angled triangle. Using Pythagoras' theorem and trigonometry:

$$\Sigma\vec{F}(\text{on box}) = \sqrt{(3.0\text{ N})^2 + (4.0\text{ N})^2}$$
$$= 5.0\,\text{N}$$
$$\theta = \tan^{-1}\left(\frac{4.0\text{ N}}{3.0\text{ N}}\right)$$
$$= 53°$$

Notice that the magnitude of the net force is not the arithmetic sum of 7.0 N. The sum is not in a straight line as in Figure 8.20.

In a two-dimensional analysis of forces, the sum or difference is calculated geometrically.

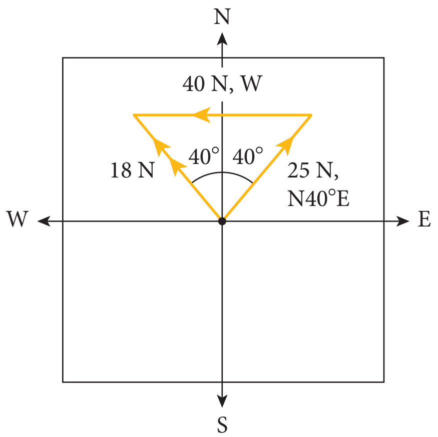
In general, two equivalent solution methods can be employed: scale drawing or trigonometric analysis.



◀ **Figure 8.22** Geometric addition of forces at right angles

Scale drawing involves careful construction of the geometric solution. A simple scale drawing can be used to find the vector sum. The force arrow lengths are drawn to a suitable scale, for example 1 cm represents 10 N. The direction of the arrows is drawn using a protractor. The length of the resultant force can be measured directly from the drawing and converted using the scale factor. The angle is measured with a protractor. Figure 8.23 shows an example of the addition of two forces acting at different angles on a block, as seen from above. The scale diagram gives a resultant force of 18 N in a direction N40°W.

This example can also be analysed using trigonometry.



▲ **Figure 8.23** An example of using a scale diagram to find the vector sum of two forces

EXPERIMENT 8.1

ADDITION OF FORCES IN A PLANE

Three forces applied to a stationary ring along different lines in a plane should add to zero. However, measurement uncertainties may lead to a measurement result that is not equal to zero. A quantitative method of determining the uncertainty in the measurement result can be used to show whether this result is within the limits of accuracy of the measurement system.

Aim

To show that the net force on a stationary, non-accelerating object is zero, within uncertainty limits

Materials

- 3 retort stands or wooden board and nails
- string
- protractor
- 3 spring balances
- small metal ring

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
The spring may flick back or flick an object into a person's eye.	Wear safety glasses when working with springs.

Procedure

1. Tie a piece of string to each of the 3 spring balances.
2. Tie the other end of each string to the metal ring.
3. Pull the spring balances and the string out from the ring at different angles and secure the ends of the balances to nails on a wooden board or the retort stands.
4. Adjust the tensions in the strings so that all 3 balances are reading within their limits.

Results

1. Record the reading on each spring balance as accurately as possible. Estimate the uncertainty in each reading.
2. Using a protractor, measure the angles made by the strings relative to each other. Make one string the reference string against which the other two angles are measured. Estimate the uncertainty in these measurements.

Analysis of results

1. Use a carefully drawn scale diagram to show the measured forces acting on the metal ring.
2. Produce a careful, scale drawing of the vector sum of the three measured forces acting on the ring.
3. Use the estimates of uncertainty in the measured spring balance forces and the measured angles to produce scale drawings that enable you to calculate the maximum and minimum values of the resultant force.

Discussion

1. Provide a value for the resultant force with uncertainty limits.
2. Was the net force on the ring zero, within the uncertainty limits? Explain.

Taking it further

How could the precision of this experiment be improved?

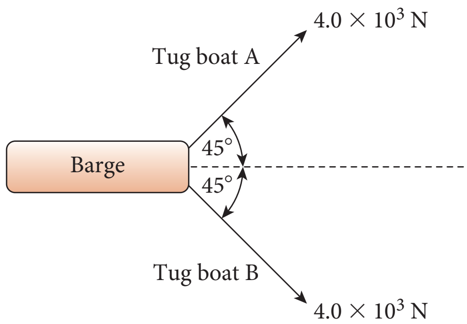
WOW

Aircraft shapes and forces

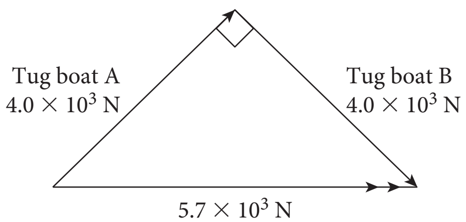
Despite being designed to fly relatively fast, the shape of modern airliners is very different from the shape of fighter aircraft. Airliners are not designed to travel faster than sound through air (1236 $km\,h^{-1}$ at 20°C but slower for colder temperatures). They have rounded, blunt noses and thick fuselages in order to carry as many fare-paying passengers as possible. Like the old supersonic Concorde, fighter planes have sharp noses and edges. The difference is rather like the difference between cutting soft butter with a hot blunt knife and slicing a cake with a sharp knife. Subsonic aircraft push the air around them, whereas the air molecules do not have time to get away from the supersonic aircraft and are bounced off the surface of these planes. This changes the way in which forces are applied to the wings of the plane and so the designs are very different.

WORKED EXAMPLE 8.5

Two tug boats apply forces on a barge through ropes as shown in Figure 8.24. What is the resultant of these two forces? (3 marks)



▲ Figure 8.24

Answer	Logic	
The two forces are acting at 90° to each other. They form a right-angled triangle:	Construct a vector scale diagram.	1 mark
 Scale: 1 cm represents 1.0×10^3 N	Show correct measurement. Select appropriate scale and use correctly.	1 mark
Note: Pythagoras' theorem can also be used. Resultant $= \sqrt{(4.0 \times 10^3\ \text{N})^2 + (4.0 \times 10^3\ \text{N})^2}$ $= 5.7 \times 10^3$ N to the east as shown.	Calculate the correct answer and direction.	1 mark

Try these yourself

1 An asteroid with a mass of 5.0×10^4 kg has a force of 250 N to the north and another force 400 N to the east acting on it. Calculate the acceleration of the asteroid due to these two forces. Use a scale drawing. (3 marks)

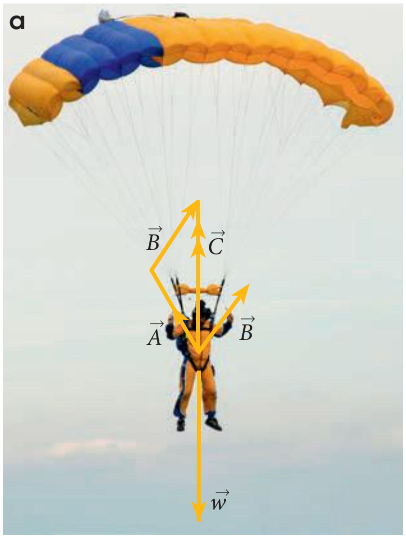
2 Two forces are applied to an alien spaceship, P, that is flying horizontally just above Earth's surface at a speed of 50 m s^{-1} directly west. One force is 6.0×10^4 N, S25°E. The other force is 8.0×10^4 N to the west.
Use a scale drawing to find the direction of the net force on the spaceship. (3 marks)

Subtraction of forces – addition of the negative!

Subtracting vector $\vec{B}$ from vector $\vec{A}$, that is $\vec{A} - \vec{B}$, is the same as adding the negative of $\vec{B}$. That is, $\vec{A} - \vec{B} = \vec{A} + (-\vec{B})$. The negative of a vector is simply the vector with the direction reversed.

Along a straight line, two forces applied to the same object can push in the same direction or they can oppose each other. The net force on the object is either the addition or the subtraction of the two forces. For example, the net force on the parachutist in Figure 8.25 is the sum of the *downwards* weight force on the parachutist and the *upwards* force applied by the strings of the parachute that are attached to the parachutist's hands. Overall, the upwards force is caused by the air friction, applied mainly to the parachute. Notice that **vector subtraction** is done by adding vectors, as long as the direction of the vectors is clearly noted.

The air friction on the parachutist is actually the vector sum of all the upwards forces on the parachutist. If we assume that there are two strings attached, then the net force on the parachutist comprises the upwards force by strings on parachutist (the vector sum) plus the downwards weight force applied to the parachutist. The net force on the parachutist is shown more realistically in Figure 8.25.



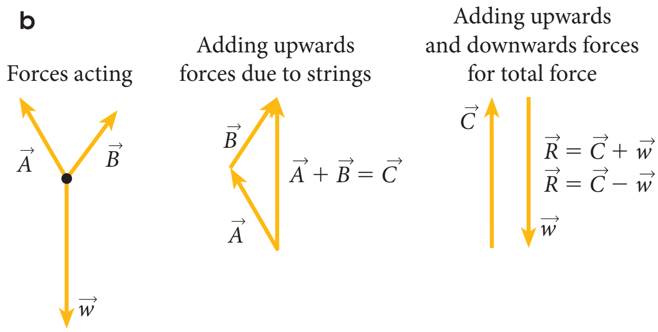


Figure 8.25 The vector sum of three forces. Note that the subtraction of the weight force is really the addition of a force in the negative sense of the other forces. Realistically, the three vectors should join up because the parachutist is likely to be travelling at constant velocity, for which $\Sigma\vec{F}\text{(on parachutist)} = \vec{C} + \vec{w} = (\vec{A}+\vec{B}) + \vec{w} = 0$.

Summary: vector addition and subtraction (addition of the negative)

Figure 8.26 shows the sum and difference of two vector forces that act on an object, P. The forces are $\vec{F}$(by A on P) and $\vec{F}$(by B on P).

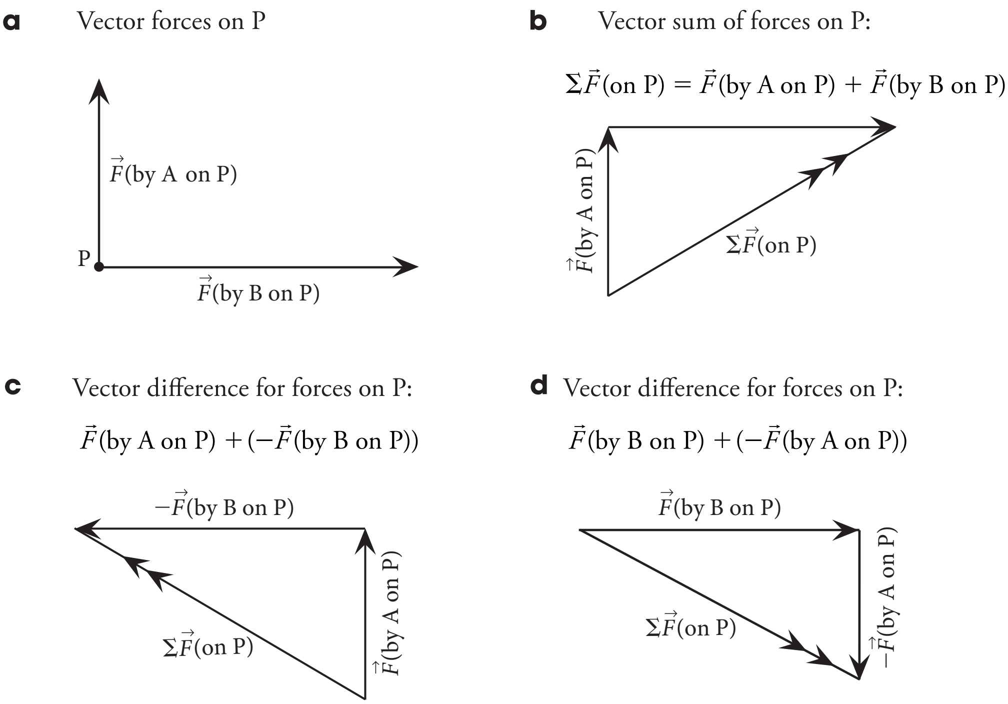


Figure 8.26 Vector sum and vector subtraction. Both are geometric additions taking account of the sense or direction of each vector. There are two possible subtraction solutions for any two vectors. a) Vectors applied at P; b) Vector sum; c) one possible subtraction vector; d) second possible subtraction vector

Case study

Road safety research – Gordon Trinca

The hospital Emergency Department on Saturday night was overflowing. There were the usual facial injuries from drunken brawls, the odd broken limb, a disturbed person demanding attention. But the worst were the car accident victims. Rushed to hospital, they required immediate attention. Were they bleeding internally? Were their brains traumatised? How much patching up was needed to give them life again, even if it meant a wheelchair and permanent care? How were they coping with the knowledge of deaths and injuries observed at the crash site?

Surgical patients were queued for assessment and then the operating theatre. But there were not enough surgeons on duty. Some victims bled out on the trolley waiting for help. Others suffered lifelong damage, which could have been avoided, before they were assessed. Early diagnosis and treatment was needed. But this was the 1960s and no one seemed to be aware of the human and economic costs of road trauma.

Gordon Trinca was an active member and, later, chairman of the Royal Australasian College of Surgeons' influential Road Trauma Committee. Fed up with stitching up road accident victims, Trinca insisted on surgeons being at the hospital when road trauma patients arrived. Later research was to show conclusively how important early diagnosis and treatment was for the recovery of trauma victims.

Trinca was pivotal to the introduction of road safety measures, such as compulsory seatbelt legislation in December 1970, which was a world first. His work was supported by Harry Gordon, a distinguished local journalist, who led the most successful newspaper campaign of the 20th century: 'Declare War on 1034'. This was named after the 1034 people who died on Victorian roads in 1970. Today, with far more road users, the number of people killed on Victorian roads has fallen from 3 a day to 1 a day. Across Australia, road fatalities per 100 000 people have dropped from 30 in 1970 to 6 in 2011. Worldwide, campaigns have reduced road trauma and increased the number of positive outcomes for trauma victims through stricter seatbelt laws, alcohol and other drug limits, restrictions on mobile phone use and safer vehicles.

Over the last 30 years, Australia has developed significant expertise in road safety research. Organisations such as the Monash University Accident Research Centre (MUARC) in Victoria, the Curtin-Monash combined C-MUARC in Western Australia and Queensland's Centre for Accident Research & Road Safety – Queensland (CARRS-Q) investigate and report on road use and safety. Road safety is increasingly viewed from multiple perspectives. Safe driving, cycling and walking all relate to interactions between humans and machines, mediated by the environment. For example, road workers are more likely to be injured at night than during the day.

Road safety research typically involves teams with expertise in diverse disciplines such as physics, biology, psychology and sociology. They investigate such things as crash data, safety standards, injury analysis, and human and environmental factors that affect road safety. They use the data to design safer vehicles and road surfaces as well as advising on legislation and regulations. Physicists have shown that crumple zones reduce the final effect on vehicle occupants involved in an accident. Braking distances more than double as speed doubles.

The ultimate goal is to reduce road crash deaths and injuries to zero.

Read about Professor Joanne Wood's (School of Optometry and Vision Science at Queensland University of Technology) work on blur and biomotion in Chapter 11, page 390.

Questions

1 a What was the problem to which Trinca and Gordon applied themselves?
  b What was their ultimate purpose?
2 a Find out what the term 'evidence-based lobbying' means?
  b How does the work of one road safety research unit mentioned in the article assist in evidence-based lobbying? Give two examples.
3 Why does legislation restrict drivers' alcohol levels and use of mobile phones? Discuss in terms of reaction distance.
4 Explain how braking distance and energy are related. Use a graph to support your answer.
5 Use the concepts of force and momentum to explain how crumple zones reduce the final effect on vehicle occupants involved in an accident. Illustrate your answer with reference to a specific set of data.

WORKED EXAMPLE 8.6

Draw arrows on a diagram to show the forces applied by external agents on a car when it is:

a stationary. (1 mark)

b moving along a straight stretch of road at a constant 60 km h^{-1}. (1 mark)

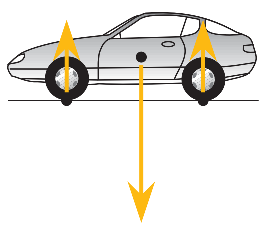
c gaining speed (positive acceleration) in a straight line. (1 mark)

d reducing speed by braking (negative acceleration) in a straight line. (1 mark)

The tail of the arrow should be placed on the point where the force acts and the arrowhead should point directly away from the point of action to show the direction of the action. The length of the arrow should represent the magnitude of the force. As this is a qualitative question, the lengths of the arrows simply represent whether the forces are equal, larger or smaller than other forces. The forces should be labelled in the form $\vec{F}$(by A on B) .

Answers	Logic	
The contact forces applied by the road on the tyres: • friction forces in the forwards direction, $\vec{F}$(by road on tyres) • the contact friction forces applied by air resistance to movement in axle/wheel connections etc., $\vec{F}$(by air on car) and the non-contact gravitational force on the mass of the car, $\vec{F}$(by weight on car), are shown in Figures 8.27–8.30.		

a



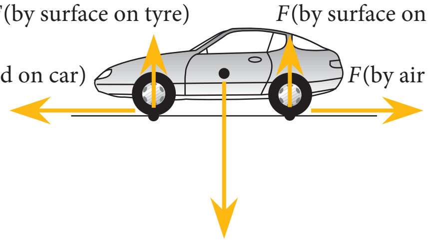
▲ **Figure 8.27**
The downwards force by weight on the car must be equal to an upwards force by the road on the car:

$\vec{F}$(by road on car) − $\vec{F}$(by weight on car) = 0

$\vec{F}$(by road on car) = $\vec{F}$(by weight on car)

Show arrows correctly. — 1 mark

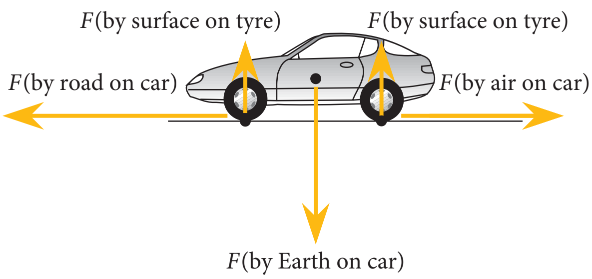
b



▲ **Figure 8.28**
The combined friction forces on the car must be equal to the net forwards forces on the car (Newton 1). The car is not travelling vertically, thus:
$\vec{F}$(by road on car) = $\vec{F}$(by weight on car).

Show arrows correctly. — 1 mark

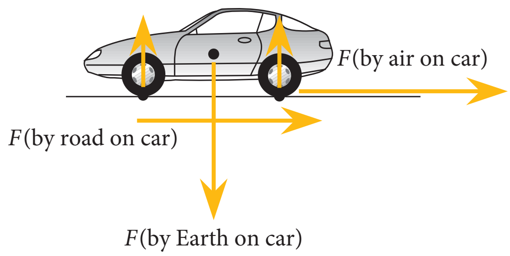
c



Show arrows correctly. 1 mark

▲ **Figure 8.29**
The combined forwards friction forces on the car must be greater than the combined backwards friction forces on the car: $\overline{F}$(forwards friction force on car) > $\overline{F}$(by opposing friction on car)·

d



Show arrows correctly. 1 mark

▲ **Figure 8.30**
The combined backwards friction forces on the car cause it to slow down (negative acceleration)

Try these yourself

Draw arrows on a diagram to show the forces acting on the following objects.

a A brick is first dropped from a cliff. (2 marks)

b A bird is flying through the air with increasing speed (i.e. positive acceleration). (2 marks)

Contact forces: The friction force and the normal force

When objects are moving across surfaces, they are simultaneously pulled down into the surface by the gravitational force (weight force), pushed up away from the surface by an electrostatic force (**normal force**) and pushed or pulled along the line of the surface by electrostatic force (friction force). Other external forces may also be applied perpendicular and/or parallel to the surface.

In Worked example 8.6 we saw that the contact force by the road acting on the tyres of the car has two components: the friction force and the normal force. The friction force is the **component of the force** by the road on the tyres that acts parallel to the road. The normal force is the component of the force by the road on the tyres that acts perpendicular to the road. Both are aspects of a single force – the electrostatic force.

Normal force

The normal force is the electrostatic force that prevents one surface from moving into another surface. It always acts perpendicular to a surface. A cup sitting on a table is subject to two forces that both act at right angles to the table. The gravitational force acts on the cup downwards. It is the gravitational force that would cause the cup to move into the table surface (Figure 8.31(a)).

However, the electrostatic normal force opposes this in the upwards direction. The weight force, F(by Earth's mass on cup) is equal to, but in the opposite direction from, the normal force, F(by surface on cup). As the two forces are acting on the same object they can be used to find the net force on the cup, F_{net}(on cup):

$$F_{net}(\text{on cup}) = F(\text{by surface on cup}) - F(\text{by Earth's mass on cup})$$

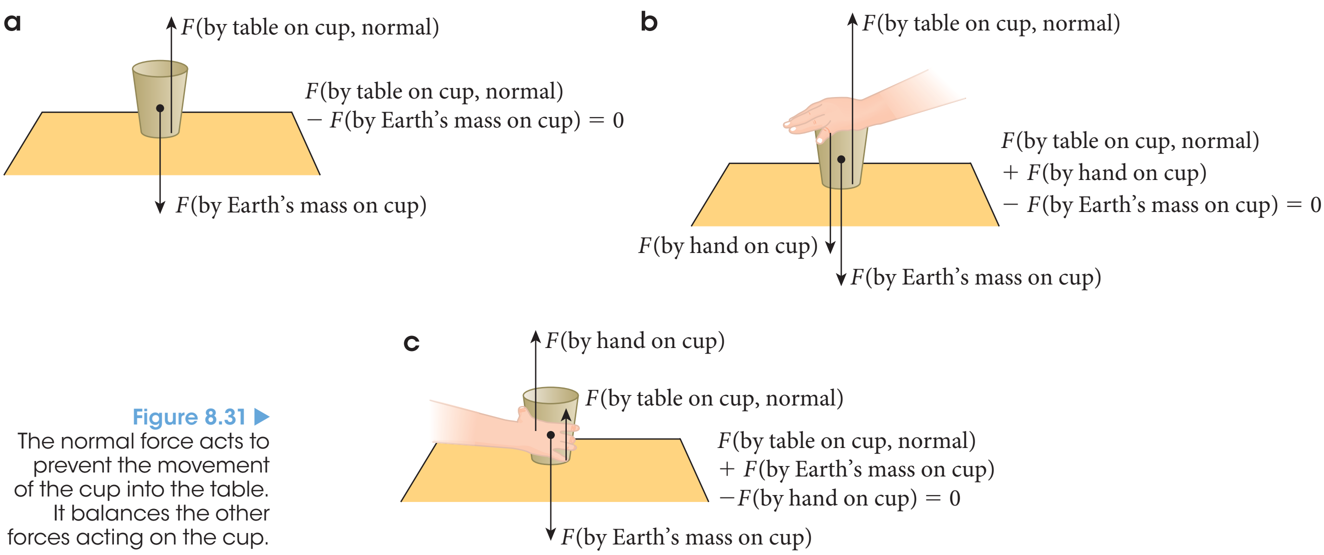


Figure 8.31 ▶ The normal force acts to prevent the movement of the cup into the table. It balances the other forces acting on the cup.

If these are the only two forces applied to the cup, and the cup is not accelerating, then the net force on the cup is zero:

$$F(\text{by surface on cup}) - F(\text{by Earth's mass on cup}) = 0$$

$$F(\text{by surface on cup}) = F(\text{by Earth's mass on cup})$$

If you push down on the cup in the direction of the weight force, the normal force on the cup will increase up to some maximum value beyond which either the cup or the surface breaks (Figure 8.31(b)).

Up until this maximum or breaking point, the cup is not accelerating, so the net force on the cup is still zero:

$$\Rightarrow F_{\text{net}}(\text{on cup}) = F(\text{by surface on cup}) - [F(\text{by you on cup}) + F(\text{by Earth's mass on cup})] = 0$$

$$\Rightarrow F(\text{by surface on cup}) = F(\text{by you on cup}) + F(\text{by Earth's mass on cup})$$

$$\Rightarrow N = F(\text{by you on cup}) + mg$$

Notice that the normal force in this case is *greater than* the weight force.

If you lift the cup up with a force opposite to the weight force, you can reduce the normal force to zero at the point at which the cup and surface are no longer in contact (Figure 8.31(c)):

$$\Rightarrow F_{\text{net}}(\text{on cup}) = [F(\text{by surface on cup}) + [F(\text{by you on cup})] - F(\text{by Earth's mass on cup})$$

$$\Rightarrow F(\text{by surface on cup}) = F(\text{by Earth's mass on cup}) - F(\text{by you on cup})$$

$$\Rightarrow N = mg - F(\text{by you on cup})$$

$$\Rightarrow N = 0, \text{ when } F(\text{by you on cup}) = mg$$

Notice that the normal force in this case is *less than* the weight force.

It should be obvious from this discussion that the normal force and the weight force act on the same object and do not always have the same value. Thus, the normal force and the weight force are NOT a pair of action–reaction forces in the sense of Newton's third law.

The normal force and the weight force act on the *same* object. They can be used to help find the net force on the object. They are NOT a Newton 3 pair of action–reaction forces.

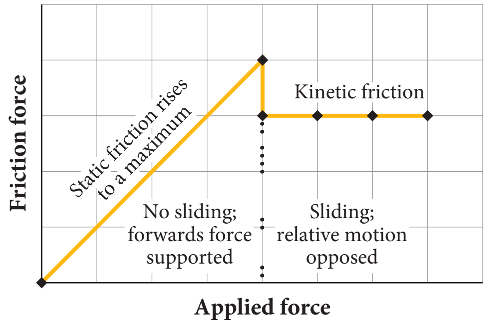
Friction force

The friction force is the electrostatic force that prevents one surface from sliding over another surface with which it is in contact. It always acts parallel to a surface. Friction, acting in the direction of motion, allows a car to accelerate and you to walk. In the absence of friction, say on a very icy road, it is much more difficult to accelerate (or brake), or walk. Friction, acting in the direction opposite to motion, causes sliding to be reduced – moving things slow down or stop.

Kinetic friction occurs when two surfaces slide relative to one another. This friction always opposes the direction of motion. For a surface, A, sliding relative to surface B, the kinetic friction force applied by B on A causes A to slow down. The more that surface A pushes into surface B, the greater the kinetic friction, *F*(by B on A). To a good approximation, kinetic friction is proportional to the normal force, since the normal force is a measure of the amount that A is pushing into B; that is, it is B's reaction to A's push.

When A and B are in contact, it takes some effort to start them sliding. This effort rises to a maximum. The external force required to overcome this **static friction** is usually greater than the kinetic friction that opposes the sliding, once it has begun. You notice this when you try to push a heavy box across the floor. It takes more effort to get it moving than to keep it moving (Figure 8.32).

As long as the static friction is acting, it can act as a forwards force to propel objects forwards, such as in walking or rolling of car tyres.



▲ Figure 8.32
Static friction increases up to a maximum, then sliding begins. The kinetic friction is less than the maximum static friction.

For two surfaces in contact:

a the normal force acts perpendicular to the surfaces to prevent penetration of one surface into the other.

b the friction force acts parallel to the surfaces.
  - **i** Kinetic friction always opposes sliding.
  - **ii** Static friction, by opposing sliding, enables forwards forces to assist walking and rolling, and backwards forces to assist slowing and stopping.

Fluids (liquids and gases) also exert frictional forces on objects that are moving relative to the fluid. We call these **drag forces**. The mechanism of these friction forces is different to that between solid surfaces.

Components of forces

We saw above that it is convenient to separate the contact force into two components. The component parallel to the surface is the friction force. The component perpendicular to the surface is the normal force. Both of these components must add up geometrically to the real contact force that is actually applied to the object.

We can separate or resolve any force into any two components. One very useful way is to resolve them at right angles. Each component then has no effect on the other – we can apply the components independently of each other. It is also useful because we can frequently identify right-angled triangles in our geometric analysis. Consequently, trigonometric ratios and Pythagoras' theorem can be deployed to solve problems analytically.

In Figure 8.33(b) three forces act on the rock. The person pulls on the rock along the line of the rope. The gravitational force by Earth on the rock (weight force) acts towards the centre of Earth. The contact force by the ground on the rock acts in a generally upwards direction. It can be resolved into a vertical component (normal force) and a horizontal component (friction) (Figure 8.33(d)).

Does the rock leave the ground? The force by the person on the rock can be resolved vertically and horizontally. In this way, the vertical component can be compared with the vertically directed normal force. The horizontal component can be compared with the horizontally directed friction force (Figure 8.33(d)).

Figure 8.33(d) is the force diagram shown in Figure 8.33(b) redrawn to show the forces all resolved into vertical and horizontal components. It is now easier to analyse this situation.

We can add the vertical components of the individual forces to find the vertical component of the net force. And we can add the horizontal components of the individual forces to find the horizontal component of the net force.

In the vertical direction, using the subscript ‘V’ to denote vertical components:

$$F_{net}(\text{on rock})_V = F(\text{by person on rock})_V - F(\text{by Earth on rock})$$

We can use trigonometry to find $F(\text{by person on rock})_V$:

$$\frac{F(\text{by person on rock})_V}{F(\text{by person on rock})} = \sin 60°$$

$$\Rightarrow F(\text{by person on rock})_V = F(\text{by person on rock}) \times \sin 60°$$

$$\Rightarrow F_{net}(\text{on rock})_V = F(\text{by person on rock}) \times \sin 60° - mg$$

Thus, the rock will be lifted from the ground when the net force is zero:

$$\Rightarrow F(\text{by person on rock}) \times \sin 60° - mg = 0$$

$$\Rightarrow F(\text{by person on rock}) \times \sin 60° = mg$$

In the horizontal direction, using the subscript ‘H’ to denote vertical components:

$$F_{net}(\text{on rock})_H = F(\text{by person on rock})_H - F(\text{by friction on rock})$$

We can use trigonometry to find $F(\text{by person on rock})_H$:

$$\frac{F(\text{by person on rock})_H}{F(\text{by person on rock})} = \cos 60°$$

$$\Rightarrow F(\text{by person on rock})_H = F(\text{by person on rock}) \times \cos 60°$$

In the horizontal direction,

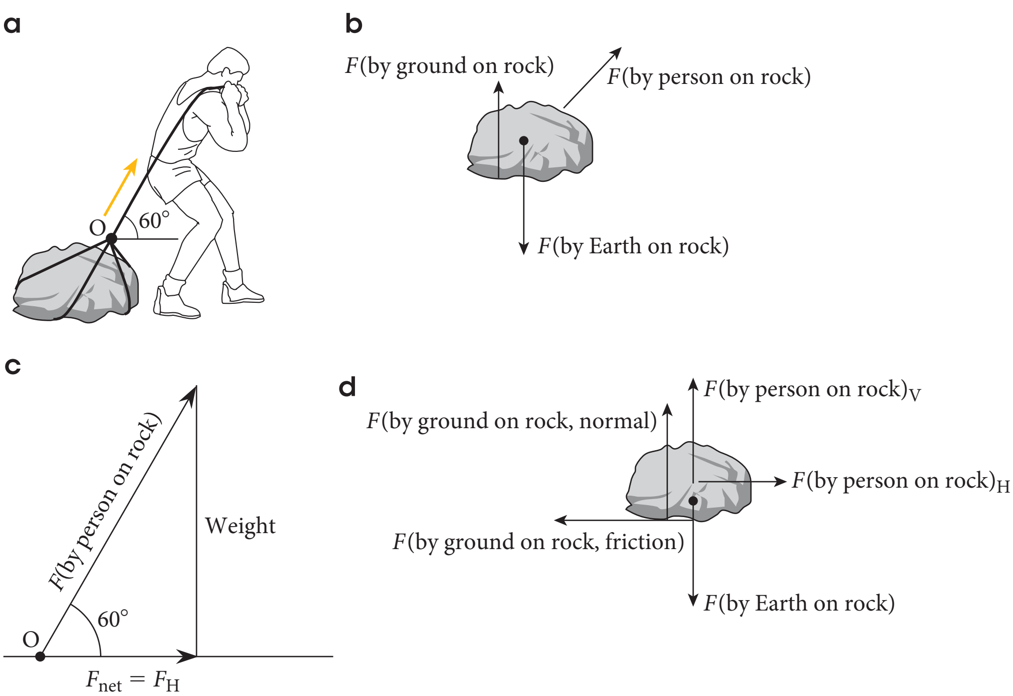
$$F_{net}(\text{on rock})_H = F(\text{by person on rock}) \times \cos 60° - F(\text{by friction on rock})$$

The rock will begin to slide across the ground when the horizontal component of the net force is zero:

$$\Rightarrow F(\text{by person on rock}) \times \cos 60° - F(\text{by friction on rock}) = 0$$

$$\Rightarrow F(\text{by person on rock}) \times \cos 60° = F(\text{by friction on rock})$$

Figure 8.33 ▶ Moving a rock by pulling. The angle of the rope affects the amount of force that can be applied along the direction of motion.



In this analysis we have separated our forces into horizontal and vertical components. This was the sensible choice for analysing the forces acting on an object on a horizontal surface. In other situations, as we shall see, we may want to choose different (but still mutually perpendicular) directions for the vector components.

WORKED EXAMPLE 8.7

A boy is pulling on the rigid handle of a heavy case. He applies a force of 23 N at an angle of 60° to the ground as shown in Figure 8.34. He tries to drag the case along the ground, but finds that the frictional force between the case and the ground is too great. We assume the same friction force applies between the edge of the case and the ground.

Find the components of the force applied by the handle to the case.

a Horizontal component (1 mark)

b Vertical component (1 mark)

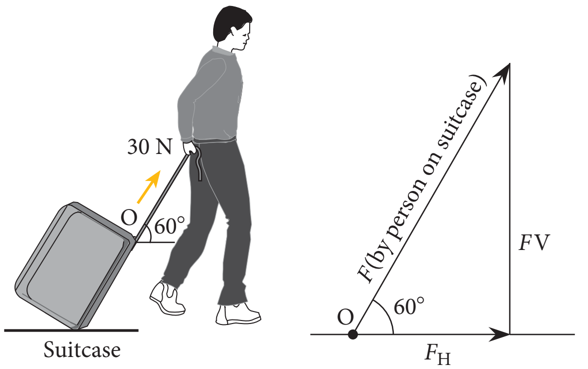


Figure 8.34
Moving a case with a rigid handle. The angle that the handle makes with the horizontal affects the magnitude of the horizontal component of the force. A smaller angle means more of the force can be applied horizontally.

Answers

a Horizontal component of the force:

$\vec{F}$(by handle on case)$_H$ = 23 cos 60° N = 11.5 N

b Vertical component of the force:

F(by handle on case)$_V$ = 23 sin 60° = 20 N

Note that by using a longer handle, the angle θ can be reduced. This increases the horizontal component of the force applied to the case, which may allow the case to be dragged along the ground.

Logic

Calculate the component correctly. 1 mark

Calculate the component correctly. 1 mark

Try these yourself

1 For the example above, find the horizontal and vertical components of the force applied to the case if the angle is reduced to 20°. (2 marks)

2 A runner pushes simultaneously backwards and downwards on the ground with a force, $\vec{R}$, at θ to the horizontal. What are the horizontal and vertical components of this force? (2 marks)

EXPERIMENT 8.2

STATIC FRICTION

The 'static' friction force is the maximum force applied by a surface to a stationary object to keep it from moving.

Aim

To find the static friction force applied by a table to a block.

Materials

- block of wood or brick (>5.0 kg)
- string
- spring balance (50 N)
- large protractor

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
Pulling a heavy object such as brick until it falls off the table or desk could cause it to land on a foot or toes.	Take care not to allow the block to fall off the edge of the table.

Procedure

1. Adjust the table so that it is horizontal.
2. Attach a 1 m length of string to the chosen block or brick with tape or by tying it.
3. Tie the other end of the string to a spring balance capable of reading up to 50 N.
4. Apply a constant force to the block by pulling on the spring balance at an angle of 90° to the horizontal.
5. Maintain the same force on the spring balance while reducing the angle the string makes with the horizontal. Move the spring balance slowly around and down (see Figure 8.35).
6. As soon as the block or brick begins to move, record the force applied by the spring balance to the string, and the angle, θ, made by the string with the horizontal table top.
7. If the block does not move, increase the force applied by the string on the block and repeat steps 5 and 6.

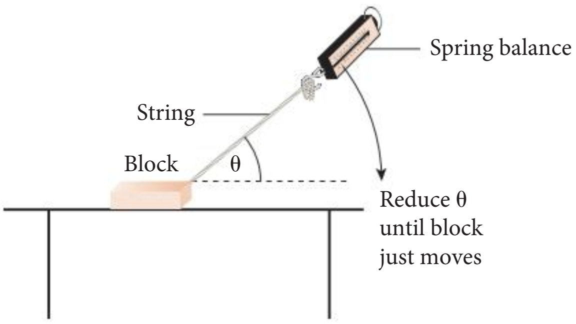


Figure 8.35 ◀
Experimental set-up

Results

Force (by spring balance on string): (_____ ± ___) N

Angle made by string with table top: (___ ± __) degrees

Analysis of results

1. Calculate the horizontal component of the force applied to the block, $F(\text{by string on block})_H$ using:

   $\vec{F}(\text{by string on block})_H = \vec{F}(\text{by string on block}) \sin\theta$
2. Use a numerical approach to find the maximum uncertainty in this value.

Discussion

1. Use Newton's third law to explain why the horizontal component of the force by the spring balance on the string, $\vec{F}(\text{by spring balance on string})_H$, has the same magnitude but opposite direction to the 'static friction'. Include a diagram with appropriately named force arrows.
2. Provide a value for static friction, including uncertainties.
3. An alternative procedure involves applying a horizontal force until the object begins to move. Would this provide a better value for static friction? Explain, including reference to data you collected.

Acceleration down a slope – the inclined plane

An object on a slope is affected by three forces: the weight force, w, acts vertically down; the normal force, N, acts perpendicular to the surface, and the friction force, f, acts parallel to the slope. The weight force is a non-contact force (gravity). The normal force and the friction force are contact forces (see Figure 8.36(a)).

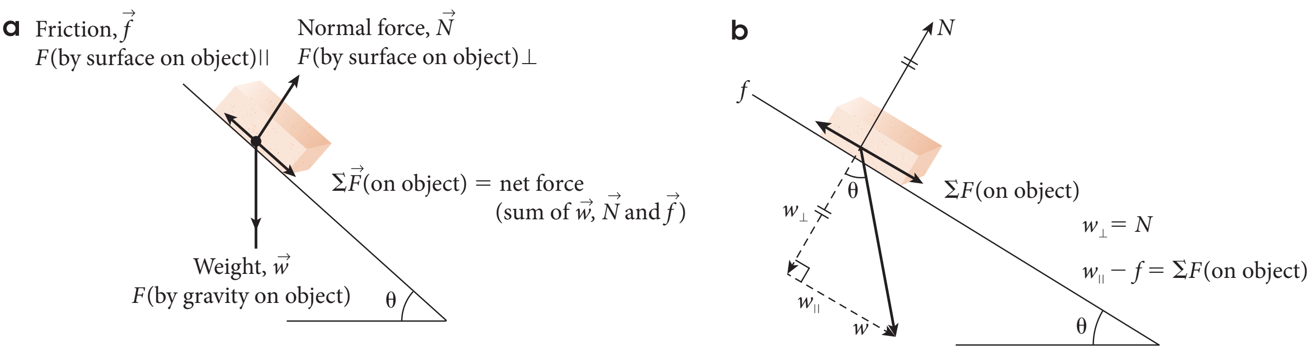
In Figure 8.36(b), N is the force by the surface on the object, acting perpendicular to the surface. It is equal to the component of the weight force that acts perpendicular to the surface, $w_{\perp}$, but which acts to draw the object into the surface. The normal force prevents the object from sinking into the surface. If the object remains on the surface and does not break through the surface, $w_{\perp} - N = 0$, i.e. $w_{\perp} = N$.

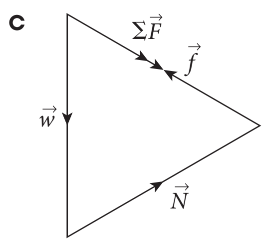
The friction force is the force applied by the surface on the object, and acting parallel to the surface. The component of the weight force that acts parallel to and down the surface, $w_{\parallel}$, is in the opposite direction to the friction force. If there is no friction force, this weight component causes the object to accelerate down the slope. The acceleration down the slope is reduced by the friction. The object will not move down the slope if $w_{\parallel} - f = 0$ (i.e. when $w_{\parallel} = f$.)

The geometry in Figure 8.36(b) shows how $w_{\parallel}$ and $w_{\perp}$ are related to $w = mg$:

$$mg\sin\theta - f_{\text{static}} = 0$$
$$\Rightarrow f_{\text{static}} = mg\sin\theta$$

Figure 8.36(c) shows the weight force, the normal force and the friction force added together as a geometric vector sum to find the net force.





▲ **Figure 8.36**
Forces acting on a mass on an inclined plane. a) Vector forces. b) Components of weight act perpendicular and parallel to the plane. c) Vector sum is zero for a non-accelerating object.

Measuring static and kinetic friction

We can use the object on a slope to measure static friction and kinetic friction. When the object is stationary, the friction is the static friction, f_{static}. Using Newton's second law:

$$mg\sin\theta - f_{\text{static}} = 0$$
$$\Rightarrow f_{\text{static}} = mg\sin\theta$$

When the object is accelerating down the slope, the friction is the kinetic friction, f_{kinetic}. Using Newton's second law:

$$mg\sin\theta - f_{\text{kinetic}} = ma$$
$$\Rightarrow f_{\text{kinetic}} = mg\sin\theta - ma$$

Friction, whether static or kinetic, varies strongly with the angle of the slope.

WORKED EXAMPLE 8.8

A 5.0 kg mass slides down a frictionless plane inclined at 30° to the horizontal (Figure 8.37).

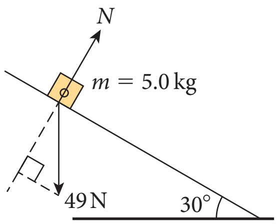


Figure 8.37 ▲

Calculate the magnitude of:

a the force applied by the surface to the mass. (2 marks)

b the force that causes the mass to slide down the slope. (2 marks)

c the acceleration down the plane. (2 marks)

Answers	Logic	Mark
a $F_{\perp} = F(\text{by gravity on mass}) \times \cos\theta$	Use the correct component.	1 mark
$\Rightarrow F_{\perp} = 5.0\,\text{kg} \times 9.8\,\text{N kg}^{-1} \times \cos 30°$		
$F_{\perp} = 42.4\,\text{N}$	Calculate the answer.	1 mark
b $F_{\parallel} = F(\text{by gravity on mass}) \times \sin\theta$	Use the correct component.	1 mark
$\Rightarrow F_{\parallel} = 5.0\,\text{kg} \times 9.8\,\text{N kg}^{-1} \times \sin 30°$		
$\Rightarrow F_{\parallel} = 24.5\,\text{N}$	Calculate the answer.	1 mark
c $a = \frac{F}{m}$		
$\Rightarrow a = \frac{24.5\,\text{N}}{5.0\,\text{kg}}$	Substitute the correct values.	1 mark
$\Rightarrow a = 4.9\,\text{m s}^{-2}$	Calculate the answer.	1 mark

Try these yourself

A 1.5 kg mass slides down a frictionless plane inclined at 20° to the horizontal (Figure 8.38).

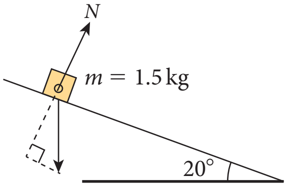


Figure 8.38 ▲

Calculate the magnitude of:

a the component of the weight force perpendicular to the plane. (2 marks)

b the component of the weight force parallel to the plane. (2 marks)

c the acceleration down the plane. (2 marks)

QUESTION SET 8.4

Remembering

1 a Describe two ways to find the net force on an object that is subjected to two forces applied along different lines in a plane.
  b For each method, identify one advantage and one disadvantage.
  c When is it possible to use Pythagoras' theorem to assist in finding the net force?

2 a Explain how the effect of several forces acting on the same object can be simplified as one net force.
  b Draw and label a diagram to show horizontal and vertical components of a force that is applied to an object at θ to the horizontal.

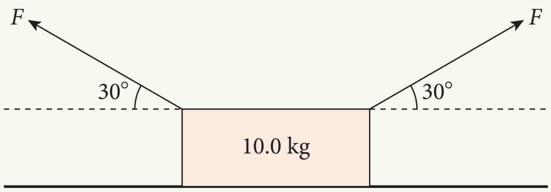
Understanding

3 A force of 30 N acts in a direction N30°E on a mass of 100 kg at rest on a smooth horizontal surface.
  a Calculate the northerly component of the force.
  b Calculate the easterly component of the force.
  c What is the magnitude and the direction of the acceleration of the mass?

4 Forces of 15.0 N and 12.0 N act at right angles on a block of ice of mass 5.0 kg on an ice rink, where friction is negligible.
  a Draw a vector diagram to show the sum of these two forces.
  b Find the acceleration of the block of ice.
  c Find the acceleration of the block of ice when a friction force of 7.0 N opposes its motion.

Applying

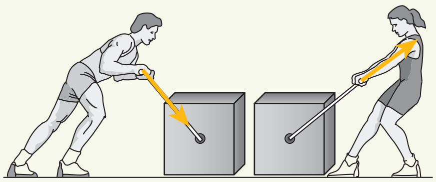
5 A 10.0 kg block is suspended from two ropes. Each rope applies the same force to the block, at an angle of 30° above the horizontal (Figure 8.39). What is the net force applied by each rope on the block?



▲ Figure 8.39

6 Estelle is a dancer. She pushes down on the floor in order to execute a vertical jump called a plié. Estelle says that this means that Newton's third law does not apply, because when she pushes down, the floor pushes up with a greater force. Explain how Estelle is wrong.

7 Is it easier to push or pull a box (see Figure 8.40)? Explain your answer in terms of the components of the forces involved.



◀ Figure 8.40

8 A 450 g mass slides down a frictionless plane inclined at 50° to the horizontal.
Calculate the magnitude of:
  a the normal force.
  b the component of the weight force parallel to the plane.
  c the acceleration down the plane.

Analysing

9 Figure 8.41 shows the velocity–time graph for a toy car of mass 0.5 kg travelling in a straight line to the right of the origin.

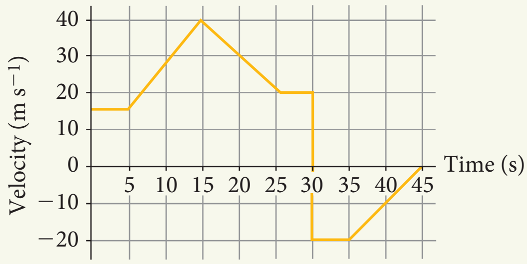


Figure 8.41 ▶

a Copy the following data table and use the graph to complete the table.

Time interval (s)	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45
Acceleration (m s^{-2})									
Force (N)									

b At a time of 10.0 s, the toy car experiences combined frictional forces of 0.2 N. What must be the forwards force to the right on the car when it is travelling at this time?

10 Two masses, A and B, are accelerated together along a frictionless horizontal surface by a force of 30 N, as shown in Figure 8.42.

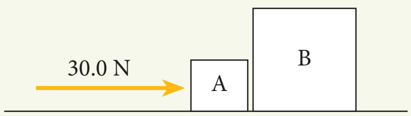


Figure 8.42 ▶

The mass of A is 1.5 kg and the mass of B is 3.0 kg.

a What is the acceleration of the two masses?

b What is the magnitude of the force exerted by B on A?

c The two masses are now accelerated together along the same smooth surface, in the opposite direction by a force of 30 N. What is the force exerted by B on A?

Reflecting

11 Describe the procedures you are able to use confidently to analyse and explain motion on a plane. Are there some techniques you need to develop further?

CHAPTER SUMMARY

- Forces are external actions applied by an agent on an object (receiver).
- A force is not a property of an object – objects do not possess a force.
- If object A applies a force to object B we denote this as $\vec{F}$(by A on B).
- The net force is the vector sum of all the forces acting on one body: $\Sigma\vec{F}$(on A)
- Forces can be classified into two main groups:
  - contact forces
  - non-contact forces (action-at-a-distance) forces.
- Mass is the amount of matter in an object.
- The mass of an object determines the rate at which it will accelerate when a non-zero net force acts on it.
- Masses attract each other due to the action of gravitational forces.
- Weight is the gravitational force applied on a mass by another mass. Near Earth, weight is the gravitational force applied by the mass of Earth on a much smaller mass: weight = 9.8 × mass (N).
- Aristotelians believe that:
  a things moved to their natural state (therefore no need for gravity).
  b a constant force caused constant motion.
  c there were two types of motion
    - natural motion (the striving of an object to return to its natural place)
    - violent motion (resulting from a push or a pull).
- Galileo showed that falling objects accelerated with the same acceleration regardless of their mass.
- Newton:
  - articulated the new understanding of motion.
  - demonstrated the importance of gravitational forces throughout the universe.
  - described three laws of motion: Newton's laws of motion.
- Newton's first law – the law of inertia:
  'If there are no non-zero net forces acting on an object, then its velocity will remain constant.'
- Newton's second law – a quantitative expression of his first law:
  $\vec{a} = \frac{\vec{F}_{net}}{m}$, also commonly written as $\vec{F}_{net} = m\vec{a}$
- Newton's third law – equal and opposites:
  'Forces act in pairs. The forces are:
  - equal in magnitude
  - opposite in direction
  and
  - act on different objects.
- A vector quantity is one that has direction as well as magnitude.
- A scalar quantity has no direction, therefore any force is a vector quantity.
- Free body diagrams are useful tools that show the forces acting on a body represented by arrows with the lengths of the arrows proportional to the magnitude of the forces acting.
- The vector sum can be found:
  - geometrically, using a scale diagram.
  - by using Pythagoras' theorem (if the forces are acting at right angles).
  - by trigonometry.

- The vector difference can be found by the addition of the opposite of the vector being subtracted.
- Force vectors can be resolved into two components at right angles to one another.
- For many situations it is useful to resolve force vectors into:
  - horizontal (H) and vertical (V) components, or
  - parallel and perpendicular components (e.g. inclined plane).
- The contact force applied by one surface on another has two components:
  - The parallel component is called friction.
  - The perpendicular component is called the normal force.

CHAPTER GLOSSARY

Aristotelian following in the tradition of Aristotle

balanced forces forces applied to an object that result in a vector sum of zero on that object

component of force the portion of a force acting in a given direction

contact force a force resulting from two objects being in contact

drag force frictional force acting on an object that is moving relative to a fluid (liquid or gas)

efficient cause the external force that was required to make things move in a horizontal direction, according to Aristotle

free body diagram a diagram that shows the magnitude and direction of all the forces acting on a body using arrows

friction a contact force applied by one surface on another; the parallel component of the contact force

friction force force by one surface on another that affects the relative motion of the surfaces

inertia the tendency of an object to resist a change in its motion

inertial mass the mass of an object determined by measuring the acceleration when a known force is applied

kinetic friction the force that occurs when two surfaces slide relative to one another; it always opposes the direction of motion

magnitude size

mass the amount of matter in an object

natural cause the source of natural motion

natural motion the striving of an object to return to its natural place

net force the sum of all the forces acting on a single object: $\Sigma\vec{F}$(on A)

non-contact force force that acts over a distance, including through a vacuum

non-zero net force forces acting on an object that, when added, result in a net force greater than zero on the object

normal force a force applied by a surface and which prevents surfaces sinking into each other; the perpendicular component of the contact force

reaction force the force that is applied to object B by object A when object A applies a force to object B

resultant force sum of forces

scalar quantities quantities with magnitude only

static friction the maximum force applied by a surface to a stationary object to keep it from moving

vector quantities quantities with magnitude and direction

vector subtraction this is the geometric addition of the negative; $\vec{A} - \vec{B} = \vec{A} + (-\vec{B})$

violent force an Aristotelian force that changes the natural motion of an object

weight force applied to a mass by another mass

CHAPTER REVIEW QUESTIONS

Remembering

1 How did Aristotle explain the constant motion of an object in terms of the forces acting? How was this idea changed by Galileo and Newton?

2 Define 'inertia'. How is inertial mass measured?

3 Draw a diagram to show the:

a vertical and horizontal components of a force vector directed at an angle θ to the horizontal.

b parallel and perpendicular components of the weight force on a mass sliding on a frictionless surface inclined at an angle θ to the horizontal.

4 Write out each of Newton's three laws in your own words.

Understanding

5 Galileo believed that objects with different masses fall with the same acceleration. However, a piece of paper falls more slowly than a rock. Explain this apparent discrepancy.

6 The magnitude of the acceleration of a freefalling object near Earth is modelled as being constant. Why, then, does this gravitational acceleration vary at different places on Earth's surface? Give three reasons.

7 Newton's second law is a quantitative consequence of Newton's first law. Explain.

Applying

8 Using a scale diagram, find the net force acting on a sailing boat when the force by the wind on the sails is 500 N to the west and the forwards combined force on the boat is 400 N to the north.

9 Use a graphical approach to find the average net force on an 80 kg athlete who accelerates from rest to $12\,\text{m}\,\text{s}^{-1}$ in 1.5 s.

10 A 1000 kg car and a 2.00×10^5 kg aeroplane accelerate at the same rate of $3.5\,\text{m}\,\text{s}^{-2}$. What is the ratio $\frac{\text{net force on aeroplane}}{\text{net force on car}}$?

11 A physics textbook is resting on the rear parcel shelf of a car moving at $50\,\text{km}\,\text{h}^{-1}$. The book becomes a dangerous missile when the car comes to a sudden stop. Explain.

12 A 40 kg block is on a frictionless plane inclined at 42°.

a Calculate the component of the weight force that is:

i parallel to the surface.

ii perpendicular to the surface.

b i What is the resultant force on the block?

ii What is the normal force on the block?

13 A toboggan and child of combined mass 67 kg slide down a frictionless snowfield inclined at 15° to the horizontal.

Calculate the magnitude of:

a the component of the weight force perpendicular to the snowfield.

b the force down the slope of the snowfield on the toboggan and child.

c the acceleration.

Analysing

14 A crowded school bus enters a left-hand corner while maintaining a constant speed. The younger pupils standing say that they were pushed to the right of the bus. Explain why they are incorrect and give a proper explanation for their apparent motion from within the bus.

15 It is often observed that a force is required to maintain the motion of an object, such as a car along a straight, level road. Does this mean that Aristotle was correct after all? Give reasons for your opinion.

16 A rocket launch involves motors expelling large quantities of gas downwards with high speed. Explain how this results in the rocket moving upwards.

17 The anchor chain of a large ship exerts a force on the ship of 6.0×10^3 N in a direction N30°E at an angle of 60° to the horizontal.

a sketch a diagram to show this situation.

b For this force, calculate the magnitude of the component in each of the following directions.

i Horizontally

ii Vertically up

iii North

iv East

v South

18 A child stands on a set of scales inside a lift. Describe how the reading on the scales will vary as the lift:

a begins to move upwards.

b travels upwards at a constant speed.

c slows to a stop.

d falls freely.

19 The friction, f applied by snow on a slope inclined at 38° to the horizontal is related to the normal force, N, by the relation:

$f = 0.080 \times N$

For a snowboard and rider of combined mass of 95 kg calculate the magnitude of:

a the net friction force applied.

b the force down the slope on the snowboard and rider.

c the acceleration.

20 Three masses, A, B and C, are accelerated together along a frictionless horizontal surface by a force of 168 N, as shown in Figure 8.43.

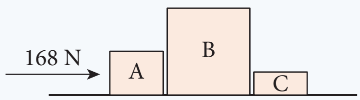


Figure 8.43 ▶

The mass of A is 6.0 kg, the mass of B is 12.0 kg and the mass of C is 3.0 kg.

a What is the acceleration of the three masses?

b Calculate the magnitude of force exerted by:

i A on B.

ii B on C.

iii B on A.

iv C on B.

Reflecting

21 Summarise three major differences between the views of Newton and Aristotle on the cause of motion of objects.

22 What is the common misconception that has led to the naming convention for forces, $\vec{F}$(by A on B)? Do you think you are able consistently to avoid this misconception in your thinking?

CHAPTER 9 WORK-ENERGY AND IMPULSE-MOMENTUM

By the end of this chapter you will have covered the following material.

Science Understanding

- Momentum is a property of moving objects; it is conserved in a closed system and may be transferred from one object to another when a force acts over a time interval (ACSPH064)
- Energy is conserved in isolated systems and is transferred from one object to another when a force is applied over a distance; this causes work to be done and changes to kinetic and/or potential energy of objects (ACSPH065)
- Collisions may be elastic and inelastic; kinetic energy is conserved in elastic collisions (ACSPH066)

Shutterstock.com/Fer Gregory

Introduction

When a net force acts on an object over a time interval, a change in the state of motion of that object occurs. The change to the motion will depend on the mass of the object. In sports such as sailing and car racing, millions of dollars are spent in producing materials that are light yet strong enough to withstand the forces that act on them. A racing car with less mass will accelerate and brake at a greater rate.

Previously, we have studied how a net force on an object causes acceleration of the object, and that once an object is in motion, no force is required to maintain that motion. Newton's three laws of motion predict and explain the motion of objects in terms of the forces applied to them.

Energy is transferred by the object applying the force, the agent. The receiver's energy therefore changes too. In this chapter, the energy transfers between objects when forces are applied to objects will be analysed.

Momentum is transferred when objects collide. Even in simple ball games, such as billiards, hockey and tennis, momentum transfers occur. The physics of car crashes also involves the transfer of momentum. The knowledge gained by high-speed observations has saved the lives of thousands of Australians by making cars safer, even though they are faster and more powerful, and there are many more cars than at any time in the past.

Newton used the idea of momentum – he called it 'the quantity of motion' – to derive the third law. Momentum is conserved in an isolated system. It is transferred directly from one object to another.

Energy

Energy is one of the fundamental ideas in physics. Energy is everywhere, yet it has no clear description and comes in many forms, as you have seen in previous studies. Energy can also be converted from one form into another using purpose-built devices. A light globe transforms electrical energy into light and heat (internal) energy.

Figure 9.1 ▲ Food, such as your breakfast, is a source of chemical potential energy.

There are two important forms of energy associated with movement. They are kinetic energy, the energy possessed by moving objects due to their motion, and potential energy due to the position of the object in a system. Potential energy is stored energy, ready to do work. A compressed spring has stored energy. If released, the spring is able to do work by applying a force through a distance. This is called elastic potential energy.

Other forms of energy include heat (or internal) energy, chemical potential energy, electrical energy, radiant energy and nuclear energy. Mass energy, also referred to as relativistic energy, which Einstein quantified so neatly in his $E = mc^2$ equation, has given rise to the view that mass and energy are interrelated. Mass can be converted to energy. Other equations are also used to model quantitatively how we can observe and measure the other forms of energy.

Energy conservation

Within the universe as a system, energy is neither created nor destroyed. The total energy in all its forms remains unchanged. This is the law of conservation of energy. Energy can be changed from one form to another, but it is never 'lost'. We rarely refer to the entire universe as our system. Rather, we limit our system to our surroundings. Sitting in a nice warm bath, we notice heat energy being 'lost' from our system. This lost energy is simply being removed from our open system and causing the air and the room to become warmer. But the universe has not lost any energy.

In an isolated system the total amount of energy remains constant. Energy can neither be created nor destroyed: it can only be transformed from one form to another. This is the law of conservation of energy.

When we consider a system in which to analyse energy, it is often best to select one that does not allow energy to leak away from or enter the system. If the law of conservation of energy does not seem to hold true in your chosen system, it means that your system is not closed or isolated.

The relation between work and energy

When a force acts through a distance, work is done on the object. Work, W, is defined as force multiplied by the distance moved in the direction of the force:

$$W = Fs$$

When work is done, energy is transferred:

$$\Delta W = \Delta E$$

where ΔW is the work done on the object and ΔE is the change in the energy of the object.

From $W = Fs$, the unit of work is $\mathrm{N\,m} = \mathrm{kg\,m\,s^{-2}\,m} = \mathrm{kg\,m^2\,s^{-2}} = \mathrm{J}$.

Both work and energy are measured in joules (J).

One joule of work done on an object is one newton of force acting on an object along its motion for one metre.

Energy may be transferred as kinetic energy or potential energy.

Kinetic energy

When a force acts through a distance on an object with mass m, energy may be given to the mass in the form of kinetic energy, E_k. Consider the object's change in speed over the duration of the action of the force, with initial speed u and final speed v. The equation can be written:

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

This equation can be compared to the equation for work:

$$W(\text{on mass}) = F(\text{on mass}) \times s$$

$$W(\text{on mass}) = mas \qquad \text{(by Newton's 2nd law)}$$

so that:

$$as = \frac{W(\text{on mass})}{m}$$

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

$$v^2 = u^2 + 2\frac{W(\text{on mass})}{m}$$

Thus:

$$mv^2 = mu^2 + 2W(\text{on mass})$$

$$W(\text{on mass}) = \frac{1}{2}mv^2 - \frac{1}{2}mu^2$$

This shows that the work done on a mass by an applied force moving through a distance changes the kinetic energy of the mass, where $E_k = \frac{1}{2}mv^2$. If the applied force, F, is in the direction of motion of the mass, then an increase in the kinetic energy of the mass will occur. However, if the applied force is in the opposite direction to the motion of the mass, a decrease in the kinetic energy of the mass will occur; that is, the mass will decelerate (speed decreases).

$$W(\text{on mass}) = \Delta E_k$$

Kinetic energy $= \frac{1}{2}mv^2$, so the unit of kinetic energy $= \mathrm{kg\,(m\,s^{-1})^2} = \mathrm{kg\,m^2\,s^{-2}} = \mathrm{J}$.

This confirms that the units of kinetic energy and work are the same.

Potential energy

A force acting over a distance may be used to store energy. This happens for springs when they are stretched or squashed. It also occurs when objects are lifted up above Earth. **Gravitational potential energy** is equal to the weight of the object multiplied by the distance, Δh, the weight is moved above Earth:

$$E_p = W(\text{on mass}) = Fs = mg \times \Delta h$$

WORKED EXAMPLE 9.1

A force of 900 N acts on a car with mass 1 tonne over 0.5 km.

a What work is done on the car by the applied force? (2 marks)

b If the car was initially at rest, what is its final speed? Ignore losses due to friction. (2 marks)

Answers	Logic	
a $W(\text{on car}) = F(\text{on car}) \times s$ $= 9.0 \times 10^2\,\text{N} \times 5.0 \times 10^2\,\text{m}$ $= 4.5 \times 10^5\,\text{J}$	First, restate the question using SI units: A force of 9.00×10^2 N acts through a distance of 5.00×10^2 m on a car of mass 1.00×10^3 kg.	2 marks
b $W(\text{on car}) = \Delta E_k$	Link work with energy.	1 mark
$W(\text{on car}) = \frac{1}{2}(mv_f)^2 - \frac{1}{2}(mv_i)^2$ $4.5 \times 10^5\,\text{J} = \frac{1}{2} \times 1000\,\text{kg} \times v^2 - \frac{1}{2} \times 1000\,\text{kg} \times (0\,\text{m s}^{-1})^2$ $v^2 = \frac{4.5 \times 10^5\,\text{J}}{500\,\text{m}}$ $v = 30\,\text{m s}^{-1}$	Substitute the correct values and calculate the answer.	1 mark

Try these yourself

1 Through what distance must a 40 N force act so that the work done on a mass is 1.6 kJ? (2 marks)

2 A cyclist and bike of combined mass 90 kg start from rest. A force of 150 N acts to accelerate the bike. For what distance must this force act so that the bike's speed becomes 12 m s^{-1}? (2 marks)

QUESTION SET 9.1

Remembering

1 To a non-physics student, the law of conservation of energy may not seem to apply to the heat energy in the liquid of a hot drink. Explain their observation.

2 Give three examples of how energy can be transformed from one form to another.

3 How is kinetic energy different from potential energy?

Understanding

4 Brakes apply a force to a vehicle to decrease its speed. Why do the brakes get hot?

5 Describe the relationship between 'work done' and 'energy'.

Applying

6 Compare the kinetic energy of a 1000 kg car moving at 60 km h^{-1} with the potential energy of the same car when it is stopped at the top of a 20 m high hill.

7 Which has the greater kinetic energy, E_k: a 20 tonne truck travelling at 20 km h^{-1} or a 1000 kg car travelling at 110 km h^{-1}? Show your reasoning.

8 **a** How much work must be done on a 1500 kg car to increase its speed from 60 km h^{-1} to 80 km h^{-1}?

b The same car travelling at 80 km h^{-1}, is now stopped quickly by the brakes. How much work must the brakes do on the car?

Analysing

9 Using graphs, explain the relationships that exist between:

a gravitational potential energy and height.

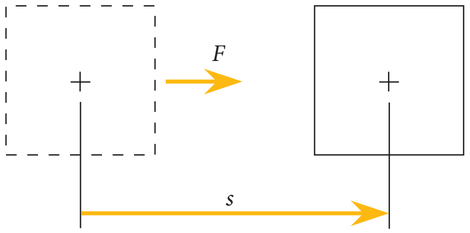
b kinetic energy and an object's speed.

Work: force acting while moving over a distance

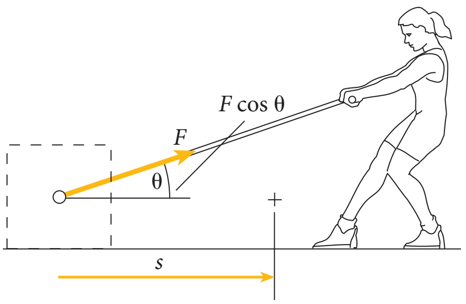
In physics we say that work is done when a force causes a particle or an object to move through some distance. As the force acts *externally* on the object, the work is done *on* the object.

It is possible to calculate how much work is done when the component of the force in the direction of the motion is known. If this is the case, then the work done on an object is the product of the force, F, and the distance, s, through which the object has moved with the force acting on it (Figure 9.2).

In cases where the applied force and the direction of motion are not in the same direction (Figure 9.3), we only consider the component of the force that is in the same direction as the motion.



▲ Figure 9.2
The force is applied in the direction of motion: $W = Fs$.



▲ Figure 9.3
The force is applied at an angle to the direction of the displacement.

Definition

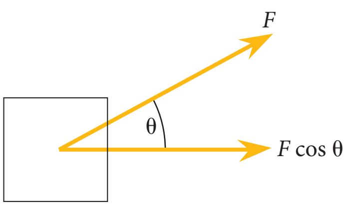
$$W(\text{on B}) = \vec{F}(\text{by A on B}) \bullet \vec{s} \quad \text{Unit: N m, J}$$

The '•' is a way of reminding us that we must always consider the component of force parallel to the distance, $\vec{s}$, moved by the point of application of the force (vector 'dot' or 'scalar' product).

The unit of work is the unit of force multiplied by the unit of distance (newton × metre), N m. This unit, called the newton-metre, is equivalent to the unit of energy, the joule, J.

Figure 9.4 shows that when we need to consider the component of force parallel to the distance moved by the point of application of the force, simple trigonometry is used:

$$W(\text{on B}) = \vec{F}(\text{by A on B}) \bullet \vec{s} \;\; = Fs\cos\theta$$



▲ Figure 9.4
The component of the force parallel to the displacement is $F\cos\theta$.

Work done by a force

A force applied to an object may be constant or it can vary over time in a number of ways. In every case, the value of the area under a F–s graph will give the work done by that force. The following examples show how this can be done.

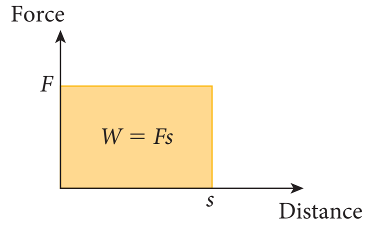
A force acting perpendicular to the object's motion does no work on the object.

Work done by a constant force

Figure 9.5 shows a constant force, F(by A on B), acting to move object B over a distance interval, s. The work done on B is:

$$W(\text{on B}) = F(\text{by A on B}) \times s = \text{area under } F\text{–}s \text{ graph}$$

Figure 9.5 ▶ Work done by a constant force



> **WOW**
>
> ### Humankind's most forceful machine
>
> The first stage of the Saturn V rocket that sent the Apollo missions to the Moon in the 1960s and 1970s was the most force-generating machine for such motion ever built. This part of the Saturn V stood as tall as a 14-storey building and weighed three million kilograms. The kerosene fuel and liquid oxygen, when combusted, could produce a combined force of 35 million newton. Each of the five rocket motors had a diameter of 3.8 m and fired for 161 seconds, using 13 000 kg of fuel per second. In this short time, the huge rocket and its payload would be lifted to an altitude of 61 km. The entire stage, including the rocket motors, was then jettisoned and fell back into the Atlantic Ocean. The smaller second and third stages would continue to lift the crew and lunar modules into Earth orbit.

Work done by a stepwise force

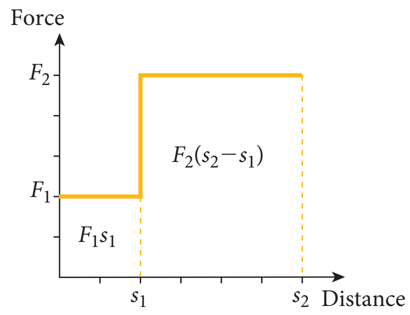
If the force by A on B changes from one constant to another constant value, then the total work done is the sum of the areas under each section, as shown in Figure 9.6:

$$W(\text{on B}) = \text{area under } F\text{–}s \text{ graph}$$

$$W(\text{on B}) = \Sigma\{F(\text{by A on B}) \times s\}$$

$$W(\text{on B}) = \{F_1(\text{by A on B}) \times s_1\} + \{F_2(\text{by A on B}) \times s_2\}$$

Figure 9.6 ▶ Work done on B is the sum of all the areas.



Work done by a steadily increasing force

For a force that commences from zero and increases steadily, the average force is $\frac{1}{2}F$, where F is the final force.

If the force increases in direct proportion to the distance moved, the work done is found from the area under the graph, as shown in Figure 9.7. In such a case we have:

$$W = ks$$

where k is the constant of proportionality between s and W. Using $y = mx + b$, k is the gradient.

The work done for a steadily increasing force is the area of the triangle: $W = \frac{1}{2}Fs$.

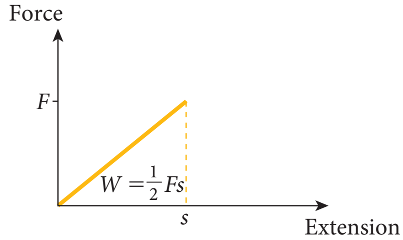
Note that this is the product of the average force and the distance through which the object is moved:

$$W = F_{ave} \times s$$

$$W = \frac{1}{2}Fs$$

$$= \frac{1}{2}(ks) \times s$$

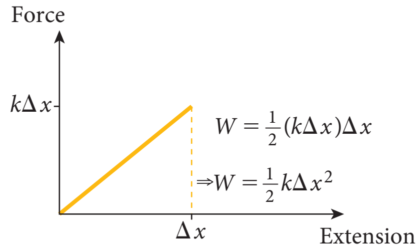
$$W = \frac{1}{2}ks^2$$



▲ **Figure 9.7**
The work done for a steadily increasing force is the area: $W = \frac{1}{2}Fs$.

Force that varies with displacement: springs

Springs are found in many applications, such as vehicle suspension systems, bicycles, bed mattresses and lounges. A spring can be modelled as an object that applies a force that is directly proportional to the change to its natural length.



▲ **Figure 9.8**
The magnitude of the force applied by the spring is a function of the magnitude of the extension.

Hooke's law

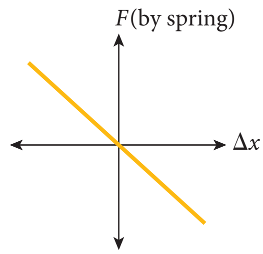
When a spring is stretched it extends by a distance Δx. The force applied to the spring to stretch it is the same magnitude as the force exerted by the spring on whatever is stretching it. This is an example of Newton's third law. The force applied to stretch the spring (*action*) is equal to the force applied by the spring, but in the opposite direction (*reaction*).

Hooke's law states that the force applied by the spring is proportional to the extension of the spring, Δx. This force is also opposite in direction to the extension of the spring. A spring can either be stretched (lengthened) or compressed (shortened). Hooke's law applies in both situations.

$$F(\text{by spring on object}) = k(-\Delta x)$$

The negative sign in Hooke's law indicates that the force applied by the spring is in the opposite direction to its extension or compression. k is the constant of proportionality between the extension or compression, Δx, and the force, F. The value of k is a measure of the **stiffness** of the spring (known as the spring constant). It is the gradient of the $F - \Delta x$ graph. An example of such a graph is shown in Figure 9.8. Only the magnitudes of the force and the extension or compression are used for these graphs.

The force, F, exerted by the spring on the object is in the opposite direction to the extension of the spring, Δx. The graph should have a negative gradient. This would appear as shown in Figure 9.9. However, graphs are usually, and sensibly, drawn to represent the magnitude of the force and the magnitude of the extension.



▲ **Figure 9.9**
The force applied by a spring, F, is in the opposite direction to the spring's extension, Δx. This results in the graph above having a negative gradient.

EXPERIMENT 9.1

THE POTENTIAL ENERGY STORED IN SPRINGS

Aim

To find, for three springs:

1 the stiffness.

2 the potential energy stored.

Materials

- 3 open springs that can be compressed and extended
- force measurer: a data logger or mechanical spring balance
- ruler
- clamp

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
The spring may flick back or flick an object into a person's eye.	Wear safety glasses when working with springs.

In your write-up, add any more risks you can think of, as well as ways to manage them.

Analysis

For each spring:

1 Construct a data table that has the following data included.

Spring	Extension of spring, Δx	Uncertainty in extension	Force applied by spring	Uncertainty in *F*(by spring)

2 Plot a graph of *F*(by spring) versus Δx, including uncertainty bars. Do not assume the line includes (0, 0).

3 Find the value of the stiffness, *k*, for each spring (gradient).

4 Find the amount of potential energy stored in the spring, E_p (area).

Discussion

1 Why can you not assume the graph line includes (0, 0)? Discuss this with reference to your data.

2 For each spring, was the spring constant the same for extension and compression?

3 By what factor did the springs differ from each other with respect to:

- **a** stiffness?
- **b** potential energy stored?

4 Explain why the comparisons of stiffness and stored energy rely on the springs being the same length *and* the same diameter.

Conclusion

1 Report your main findings in a data table.

2 Describe limitations to the validity and reliability of your findings.

3 Indicate what might be done better to make fair comparisons between springs, particularly if your springs were of different lengths and diameters.

Extension

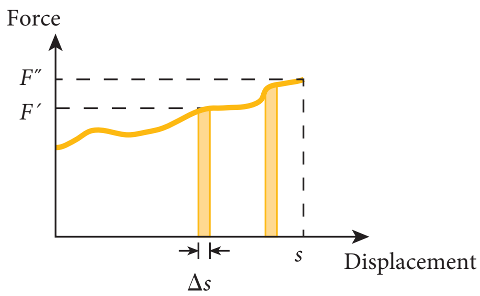
1 Find out about stress and strain of materials under compression and tension. How could these ideas be used to improve fair comparisons between springs in your experiment?

2 Create a visual report on the use of springs in industry and transport.

Any force that varies with displacement

If the magnitude of the force changes continuously as the displacement changes, without the direction changing, then the total work done must be found graphically, as shown in Figure 9.10.

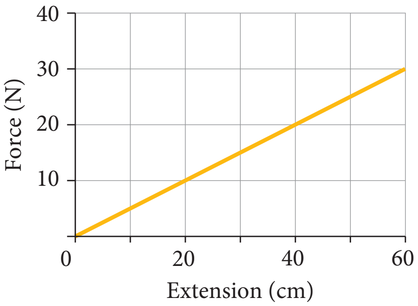
The work done in moving a small distance Δs is equal to $F'\Delta s$. To find the total amount of work done over the total distance moved, s, we would need to find the total area under the graph.



▲ **Figure 9.10**
The work done is equal to the area, which is the sum of all the small areas like the one shown: $W = \sum(F'\Delta s)$.

WORKED EXAMPLE 9.2

Figure 9.11 shows the magnitude of a force applied by a spring as a function of the magnitude of the extension of the spring.



◀ **Figure 9.11**

1 What is the magnitude of the force applied by the spring when it is extended by 40 cm? (1 mark)
2 The force applied by the spring when it is extended 60 cm is 30 N. Explain why this is the same magnitude as the force being applied to the spring to extend it this distance. (1 mark)
3 What is the spring constant (the value of k) of this spring? (2 marks)
4 How much work is done by the force applied to the spring to extend the spring 40 cm? (2 marks)

Answers	Logic	
1 From the graph, F(on spring) = 20 N	Read the value from the graph.	1 mark
2 For all extensions of the spring, the force applied by the spring is equal in magnitude to the force applied to the spring. But the forces act on different objects, hence they act in different directions. This is a re-statement of Newton's third law.	Give the correct explanation, citing the correct law.	1 mark
3 $F = k\Delta x$	Use the correct equation.	1 mark
$k = \frac{F}{\Delta x}$		
$= \frac{20\text{ N}}{0.40\text{ m}}$	Substitute the correct values.	1 mark
$= 50\text{ N m}^{-1}$		

▶

4 From the graph, at 40 cm: | Read the correct value from the graph, given the limit of reading. | 1 mark

$F = 20\,\text{N}$

Area under graph:

$$W = \frac{1}{2}F(\text{by spring}) \times \Delta x$$
$$= \frac{1}{2}20\ \text{N}(0.40\ \text{m})$$
$$= 4.0\,\text{J}$$

Calculate the area under the graph (remember to convert from cm to m). | 1 mark

Try these yourself

1 The spring in Figure 9.11 is extended 60 cm by an applied force. When released, the spring returns to its normal resting length. How much work does the spring do when returning to its normal length? (3 marks)

2 How would the graph shown in Figure 9.11 differ if the spring had double the stiffness? (2 marks)

Ideal springs

Hooke's law describes the relationship between force and displacement in an **ideal spring**. In an ideal spring, all the work done on the spring by the applied force is stored in the spring, ready to do work when the spring is released. In the real world, a small amount of work done on the spring by the applied force is transformed into heat energy. Also, when the spring is released, a small amount of the potential energy in the spring is lost from the spring system as heat energy. In addition, the range of extensions and compressions to which $F = kx$ applies is limited. Springs typically have a limited linear region in which this applies.

QUESTION SET 9.2

Remembering

1 Define work in terms of applied force, displacement and the angle between force and displacement.

2 How does the gradient of a F versus Δx graph for a very stiff spring compare with that for a spring that is not as stiff?

3 How does Hooke's law relate to the force applied to a spring?

Understanding

4 Explain why, when you are holding a book above your head, you are not doing any work on the book.

5 Is work done on an object if the force applied to the object is perpendicular to its motion? Give reasons for your answer.

Applying

6 How much work is done by a constant horizontal force of 75 N that moves a javelin through a horizontal distance of 2.0 m?

7 A force–extension graph for an ideal spring is shown in Figure 9.12.

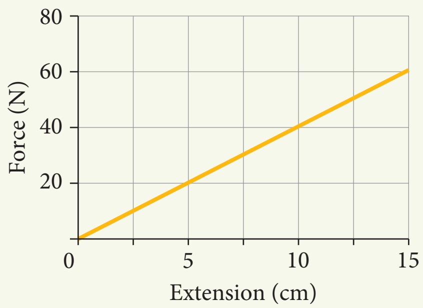


Figure 9.12 ▶

a What is the spring constant, k, for this spring? Express your answer with units of $N\,m^{-1}$.
b What force is required to extend this spring 10 cm?
c How much energy is stored in this spring when it is extended 5.0 cm?
d How much extra energy will be stored in the spring if it is extended from 0.05 m to 0.10 m?

8 A kite surfer's kite strings apply a force of 200 N at an angle of 30° above the horizontal. The kite surfer moves a distance of 300 m along the water. Sketch a diagram of this and calculate how much work is done on the kite surfer by the kite strings.

Analysing

9 How would the area under a force versus extension graph for a real spring differ from the potential energy stored in this spring?

Reflecting

10 Describe how your ideas about what 'work' is have been changed during your study of this chapter.

Work and energy transfers near Earth

Any object near Earth's surface experiences a gravitational force due to the large mass of Earth. This force has magnitude $F = mg$ where m is the mass of the object and g is the acceleration due to gravity close to Earth's surface, $9.8\,m\,s^{-2}$, which acts directly downwards.

Imagine holding an object at some height and dropping it. If we ignore air resistance, the only force acting on the object is the gravitational force, $F = mg$. The object falls directly downwards, which is in the same direction as the force, through some height Δh. Remember that work is $W = Fs$, hence the work done by the gravitational force in this case is:

$$W(\text{by gravity on object}) = F_g\Delta h = mg\Delta h$$

The object's kinetic energy must change by this amount. So, if it started at rest, it now has kinetic energy:

$$E_k = \frac{1}{2}mv^2 = mg\Delta h$$

We know that energy is conserved, so where did the energy that has now appeared as kinetic energy of the object come from? It must have come from some stored potential energy. In fact, it comes from energy stored due to the position of the object relative to Earth, because of the gravitational force due to Earth. By conservation of energy, the change in gravitational potential energy is:

$$\Delta E_p = -mg\Delta h$$

Note the minus sign: it is telling us that if kinetic energy has increased, the potential energy has decreased.

If we now define the zero of potential energy as being at the surface of Earth, we can say that any object at some height h above Earth's surface has gravitational potential energy:

$$E_p = mgh$$

Note that this choice for our zero of potential energy is arbitrary. It is a convenient choice when analysing the behaviour of objects close to Earth's surface, but other choices are better when we consider, for example, the motion of objects far from Earth.

This choice may mean that sometimes an object has a negative potential energy. For example, if we have defined the zero of gravitational potential energy as at the surface of Earth, then any object dropped into a hole will have a negative potential energy. This doesn't matter, as in practice we can only measure changes in potential and kinetic energy. These changes can be positive or negative. In fact, in any isolated system, any positive change in total kinetic energy must be accompanied by a negative change in total potential energy and vice versa.

Imagine lifting an object from rest at constant speed directly upwards. The net force on the object must be zero because the object has zero acceleration. (We will ignore for the time being the extra force required just for an instant to accelerate it to this constant speed.) The applied force must have equal magnitude, but opposite direction, to the gravitational force. Hence, when we lift an object from ground level to some height h we must do an amount of work equal to the change in potential energy, ΔE_p, which is again:

$$\Delta W = mg\Delta h = \Delta E_p$$

Earlier, it was stated that when a force acts on an object in the direction of its motion, the work done is converted into kinetic energy. Yet this object we have lifted at constant speed has not changed its kinetic energy. How can this be the case? Remember that at the same time as we are lifting the object by applying a force to it, the Earth is also applying a gravitational force in the opposite direction. The gravitational force does an equal amount of work on the object, but in the opposite direction. Hence the total change in kinetic energy is zero – the object begins and ends at rest because, whatever work we have done, the Earth has done an equal but negative amount. This is because the gravitational force is in the opposite direction to the displacement. This now appears as if we have a violation of the conservation of energy – the potential energy of our system (the object and Earth) has increased, but there has been no change in kinetic energy. The answer is that the extra energy has come from outside the system – from us, lifting the object. If we define our system as just the object and Earth, this is not an isolated system. We can transfer energy into the system by doing work on it. This illustrates how important it is to clearly define your system when you wish to apply conservation of energy!

When an object is infinitely far away from Earth it has no kinetic energy (it cannot overshoot infinity). If we assign the potential energy at infinity to be zero also, then the total energy is zero. As the object falls towards Earth it gains kinetic energy and potential energy must be lost. This is covered in Unit 3 in more detail.

Where is the energy?

There is one final important point to make about potential energy. Although we often say that 'an object has potential energy', this is not strictly correct. An object above Earth's surface only has potential energy because of its position relative to Earth. It is more correct to say that the energy is stored in the *system* – the Earth plus the object. An isolated object with no forces acting on it cannot be said to have any potential energy. It is only the fact that a force is acting, due to some other object, that means there is potential energy stored. Hence *the energy belongs to the system*, not the object. The same is true of kinetic energy. Kinetic energy is given by $E_k = \frac{1}{2}mv^2$ We need some reference frame or object against which we measure v. Usually this is the surface of Earth. But the surface of Earth is itself moving because Earth is rotating, and orbiting about the Sun. The Sun is itself in orbit about the galactic centre. So, in fact, kinetic energy is only meaningful when we can say what we are measuring velocity relative to. In an otherwise empty universe, it would be meaningless to say that an object has either kinetic or potential energy.

WOW

Rocks falling from space

If you have ever seen a meteor, or 'shooting star', you have seen how fast they travel. They appear to be much lower than they really are due to their very high speed. Most meteors are small pieces of rock that crash into Earth's upper atmosphere with such a high speed that they heat up and vaporise before hitting the surface. It is not only Earth's gravitational field that gives meteors their high speed. As Earth orbits the Sun, it moves through space at about 30 km s^{-1}. A meteor crashing to Earth is rather like a small bug squashing on the windscreen of a fast-moving jet aeroplane!

Shutterstock.com/jdwfoto

Figure 9.13 ▲ A meteor in the night sky

Scientific literacy: Getting to Mars faster

There's a growing chorus of calls to send astronauts to Mars rather than the Moon, but critics point out that such trips would be long and gruelling, taking about six months to reach the Red Planet. But now, researchers are testing a powerful new ion engine that could one day shorten the journey to just 39 days.

Traditional rockets burn chemical fuel to produce thrust. Most of that fuel is used up in the initial push off Earth's surface, so the rockets tend to coast most of the time they're in space.

Ion engines, on the other hand, accelerate electrically charged atoms, or ions, through an electric field, thereby pushing the spacecraft in the opposite direction. They provide much less thrust at a given moment than do chemical rockets, which means the rockets can't break free of Earth's gravity on their own.

Several space missions have already used ion engines, including NASA's Dawn spacecraft, and Japan's spacecraft Hayabusa, which rendezvoused with the asteroid Itokawa in 2005.

But a new engine, called VASIMR (Variable Specific Impulse Magnetoplasma Rocket), will have much more 'oomph' than previous ones. That's because it uses a radio frequency generator, similar to transmitters used to broadcast radio shows, to heat the charged particles, or plasma.

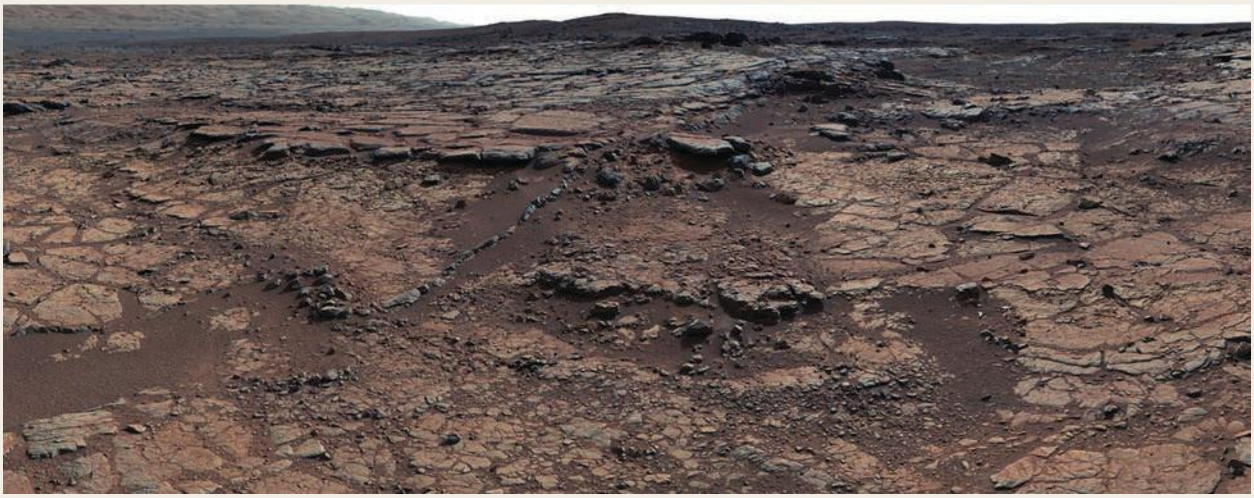
The engine is being developed by the Ad Astra Rocket Company, which was founded in 2005 by plasma physicist and former space shuttle astronaut Franklin Chang-Diaz.

VASIMR works something like a steam engine, with the first stage performing a duty analogous to boiling water to create steam. The radio frequency generator heats a gas of argon atoms until electrons 'boil' off, creating plasma.

The plasma could produce thrust on its own if it were shot out of the rocket, but not very efficiently. To optimise efficiency, the rocket's second stage then heats the ions to about a million degrees, a temperature comparable to that at the centre of the Sun.

It does this by taking advantage of the fact that in a strong magnetic field – like those produced by superconducting magnets in the engine – ions spin at a fixed frequency. The radio frequency generator is then tuned to that same frequency, injecting extra energy into the ions.

Grossman, L. (2009) 'Ion engine could one day power 39-day trips to Mars'. *New Scientist*, 24 July.



NASA/JPL-Caltech/MSSS

▲ **Figure 9.14**
An image of the surface of Mars taken from the Rover Curiosity's mast camera, showing the Yellowknife Bay formation, December 2012

Questions

1 Are ion drive engines being used to launch rockets from Earth's surface?
2 When have ion drive motors been used previously?
3 How are the ions produced for the VASIMR motor?
4 Explain how Newton's third law is applicable to an ion drive motor.
5 Why would ion drive motors be excellent motors for very long voyages in space?

Projectiles

When an object is thrown upwards with an initial speed u, its speed decreases as its height increases. The work being done on the object causes a loss of kinetic energy, E_k. This energy is converted into energy that is able to do work later, gravitational potential energy, E_p. At the top of the flight, where it is momentarily stopped, the object has lost all of its kinetic energy and gained potential energy:

$$\Delta W = mg\Delta h = +\Delta E_p = -\Delta E_k = -\frac{1}{2}mu^2$$

We know that this object has potential energy because it immediately begins to fall back down, gaining kinetic energy and losing its gravitational potential energy.

Potential energy is always calculated as an increase or a decrease. The point where the value of gravitational potential energy is assigned a zero value depends on the situation. For vertical motion, this is often taken to be at Earth's surface. Other situations assign the zero value for gravitational potential energy to be at an infinite distance away from Earth.

In cases where all kinetic energy is converted into gravitational potential energy or vice versa:

$$\frac{1}{2}mv^2 = mg\Delta h$$

$$\Rightarrow \quad v = \sqrt{2g\Delta h}$$

The result for the vertical speed of the object is not dependent on its mass. This would go against the beliefs of an Aristotelian physicist; however, Galileo and Newton showed this is indeed the case. Falling objects fall at the same rate regardless of their masses. As this equation shows, the speed depends only on a change of height and the strength of gravity.

WORKED EXAMPLE 9.3

1 How much work is done by a crane lifting a 200 kg load through a height of 25 m? (2 marks)

2 A waterslide starts from a height of 15 m above the ground. A 40 kg person sits at the top of the waterslide.
   a What is the maximum kinetic energy the person will gain when sliding back down? (2 marks)
   b How fast will the person be going at the bottom of the slide if there is no friction? (2 marks)
   c Will a 60 kg person go down the slide faster than the 40 kg person? Give reasons for your answer. (1 mark)

Answers	Logic	
1 $W(\text{on load}) = F_g\Delta h = mg\Delta h$	Use the correct equation.	1 mark
$= 200\text{ kg} \times 9.8\text{ m s}^{-2} \times 25\text{ m}$ $= 4.9 \times 10^4\text{ J}$	Substitute the correct values and calculate the answer.	1 mark
2 a $W(\text{on person}) = F_g\Delta h = mg\Delta h$	Use the correct equation.	1 mark
$= 40\text{ kg} \times 9.8\text{ m s}^{-2} \times 15\text{ m}$ $= 5.9 \times 10^3\text{ J}$	Substitute the correct values and calculate the answer.	1 mark
b $\Delta W = -\Delta E_p = +\Delta E_k$	Use the correct equation.	1 mark
$\Delta E_k = 5.9 \times 10^3\text{ J}$ $\frac{1}{2}mv^2 = 5.9 \times 10^3\text{ J}$ $v = \sqrt{\frac{2 \times 5.9 \times 10^3\text{ J}}{40\text{ kg}}}$ $= 17\text{ m s}^{-1}$	Calculate the answer.	1 mark
c No, the mass of the falling object has no bearing on the rate at which it falls, so their higher mass will have no effect on the speed of the 60 kg person.	Use correct reasoning.	1 mark

Try these yourself

The Moon has no atmosphere and a gravitational acceleration of about 1.6 m s^{-2}.

a How much gravitational potential energy does a lunar lander of mass 400 kg lose when it descends from 2.0 km above the surface? (2 marks)

b If the rocket motors failed completely at 2.0 km above the surface, what would be the speed of the lunar lander when it hits the surface of the Moon, assuming it began its descent with no speed? (2 marks)

Power

A crane that can lift a 10 tonne block of concrete to the top of a building in 1 minute is more powerful than a crane that can do the same task but takes 2 minutes. Both cranes do the same amount of work on the load but they take different times to do so. The rate at which work is done is called power, P. It is the rate at which energy is being transferred:

$$P = \frac{\Delta W}{\Delta t} = \frac{\Delta E}{\Delta t}$$

The unit of power is the watt (W). One watt is 1 joule per second or 1 newton-metre per second. A 100 W light globe is transforming 100 J of electrical energy into light and heat energy every second. A 1 kW crane motor increases the gravitational potential energy of a load by 1 kJ every second (assuming 100% efficiency and ignoring friction).

Power and energy are two quantities in physics that are often confused and misused. Power is the time rate of change of energy, so the amount of energy used is the power multiplied by the time interval.

$$\Delta E = P\Delta t$$

A 100 W light globe uses as much electrical energy in one hour as a 1000 W kettle uses in 6 minutes. This is one reason why energy-efficient devices – those that use less power to do the same task – have helped to reduce the overall use of electricity in Australia in recent years.

As power is used to measure the rate of energy usage in society, it is common to see large units being used:

$$1\ \text{kW} = 1\ \text{kilowatt} = 10^3\ \text{W}$$

$$1\ \text{MW} = 1\ \text{megawatt} = 10^6\ \text{W}$$

$$1\ \text{GW} = 1\ \text{gigawatt} = 10^9\ \text{W}$$

It is usual for the highly populated states of New South Wales and Victoria to use 10 GW of electrical power during the day, with peak demand well in excess of this. Domestic electricity tariffs measure electricity quantity in kilowatt hours (kW h). This is the amount of electricity used at the rate of one kilowatt for one hour.

One of the factors that determines the top speed of a car is the power of its engine. Frictional forces on the car oppose its motion. Air resistance on the car increases significantly as the speed of the car increases. The work done on the car by this air resistance decreases the energy of the car. In order to keep the car travelling at constant speed, the engine must provide this amount of energy by converting the energy stored in the fuel, so fuel consumption increases with speed.

From the definition of power:

$$P = \frac{\Delta W}{\Delta t} = \frac{Fs}{\Delta t} = F\frac{s}{\Delta t}$$

$$\frac{s}{\Delta t} = v \quad \text{so} \quad P = Fv$$

This means that the power, or rate at which energy is being transferred, is equal to the force applied multiplied by the speed at which the object is moving.

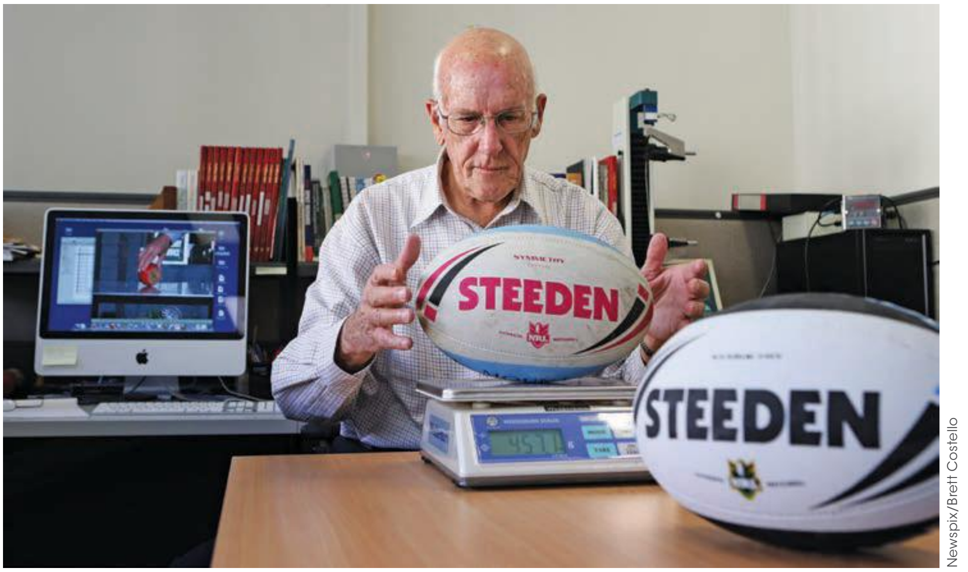
Case study

Associate Professor Rod Cross – using physics to solve crime, and the mystery of sports

Associate Professor Rod Cross has been working in the field of physics for more than 45 years. Recently he has been analysing the physics of many sports and has been involved as an expert witness in murder cases. Forensic physics is a little-known area of expertise. When police alleged that a body had been thrown off a cliff and the defence lawyers said the person had jumped, Professor Cross became involved. Projectile motion and the laws of physics showed that the person had been thrown, and the prosecution won the case.

Whether to use a light or heavy cricket bat in order to hit the ball further is an example of Professor Cross's later work at the University of Sydney. Fascination with the underlying principles of tennis, cricket, golf, softball and many other sports has resulted in published books and other publications on how these games actually work.

A curved ball in softball or baseball is a very hard ball to hit, but it can be understood by Professor Cross's analysis of projectile motion and spinning balls. Although high school physics largely ignores air friction, this force and the way it interacts with a moving ball in any ball sport is crucial to understanding and playing the game.



▲ Figure 9.15
Associate Professor Rod Cross weighing rugby balls as he tests the differences between a conventional NRL ball and a pink Women In League NRL ball at Sydney University, New South Wales, 2012

Professor Cross worked for many years on plasma physics. Plasma is sometimes referred to as the fourth state of matter in which the atoms and molecules of a gas have been ionised at very high temperatures. This allows the plasma to be controlled using magnetic and electric fields. A kind of plasma exists in the fluorescent tube that just might be above your head as you read this.

Questions

1. Members of the jury in a criminal trial may not have any background in physics. Should they take the opinion of a physics professor as evidence in such trials? Give reasons for your answer.
2. What variables would need to be considered when estimating how fast a car must have been moving if it was driven off a cliff?
3. In what ways might a furry tennis ball behave differently from a smooth cricket ball when thrown?
4. Why is it not possible to analyse adequately the motion of spinning balls using the projectile motion equations?
5. What is plasma?

QUESTION SET 9.3

Remembering

1 What energy transfer is taking place when an object falls vertically downwards near Earth's surface?

2 How are power, work and time related?

3 A continuous, constant force is applied to an object to overcome gravity and push it upwards. What are the two energy transformations that are most likely to occur?

Understanding

4 In a real situation, would a falling mass near Earth expect to have all of its gravitational potential energy converted to kinetic energy? Explain your answer.

5 A 500 kg boulder and a 5 kg rock fall off a cliff edge at the same time and fall to the valley floor 200 m beneath.

a Compare the loss of gravitational potential energy of the boulder and the rock.

b Which would you expect to land first? Explain your reasoning.

6 Explain why a child takes longer to go down a slide that is 3.5 m high than to fall from this height.

Applying

7 A skydiver wearing a new frictionless suit jumps from a height of 5.0 km. By considering the change in her kinetic and potential energies, determine the vertical distance she must fall to achieve a speed of 150 $m\,s^{-1}$.

8 A crane is lifting a 900 kg elephant upwards.

a What work is done by the crane if the elephant is lifted through a vertical displacement of 16 m?

b The crane is lifting the elephant at a speed of 1.5 $m\,s^{-1}$. At what power is the crane operating?

Analysing

9 The air resistance on a small object of mass 1.2 kg travelling horizontally is given by the equation:

$$F_{air} = 9.0 \times 10^{-4} v^2$$

The object has a constant forwards force of 20.0 N applied to it. When it is travelling at 30 $m\,s^{-1}$, find:

a the air resistance on the object.

b the net force on the object.

c the acceleration of the object.

Reflecting

10 All modern cars can easily perform at and above the maximum road speed limits. Why, then, is there such an emphasis on power?

Impulse: force acting over time

When a force is applied to an object over a time interval, we say an **impulse** has been applied to the object by the force. Impulse is also a vector quantity in that it has the same direction as the direction of the force being applied. The magnitude of the impulse, *I*, is found by the product of the force being applied, *F*, and the time interval, Δt, over which the force is applied.

Definition of impulse, *I*

For a force being applied to object B by object A:

$$\vec{I}(\text{on B}) = \vec{F}(\text{by A on B}) \times \Delta t$$

The unit of impulse is the unit of force × time, newton-seconds, *Ns*.

Impulse exerted by a constant force

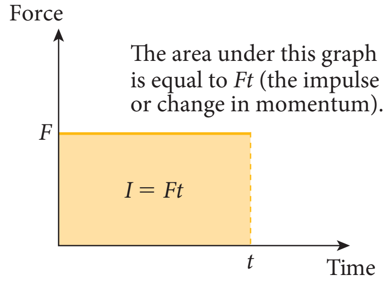


Figure 9.16 ▲ Impulse is the area under the *F*–*t* graph.

When a constant force acts on an object for a time interval, the graph of force versus time appears as shown in Figure 9.16. The impulse I is the area under the F versus t graph, which in this case is a rectangle. Put simply, $I = Ft$.

Impulse exerted by a variable force acting over time

We can find the impulse as the area under any force versus time graph. This is similar to the way in which we previously found the area under force versus displacement graphs to find the work done. Figure 9.17 shows the area for a short time interval Δt. Figure 9.18 shows the area for the length of time t, which began at the origin.

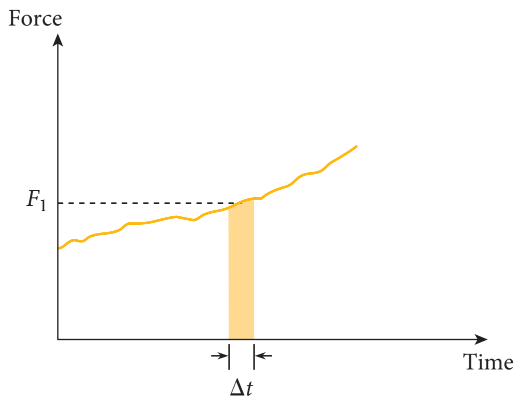


Figure 9.17 ▲ Impulse is approximately the area of the shaded area $F_1\Delta t$.

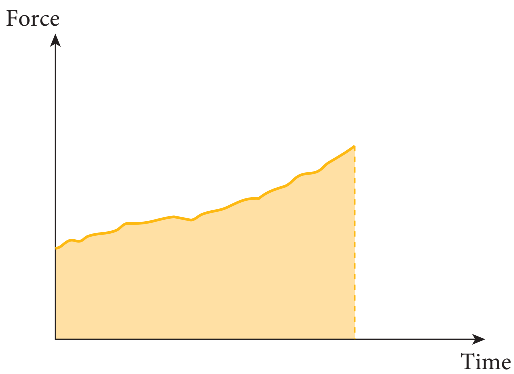


Figure 9.18 ▲ Impulse is the sum of all the small areas, which is the total area under the *F*–*t* graph.

Impulse and momentum

Recall that t and Δt both refer to a time interval.

We know that when a force F acts on a mass m the force will cause the mass to accelerate, changing its speed from its initial value u to a final value v over a time interval t. This can be written as:

$$v = u + at$$

Using the equation for impulse, $I = Ft$, we can find an expression for the at part of this equation:

$$I(\text{on mass}) = F(\text{on mass}) \times t$$
$$I(\text{on mass}) = mat$$
$$at = \frac{I(\text{on mass})}{m}$$

Now, when this is combined with the previous equation:

$$v = u + at$$
$$v = u + \frac{I(\text{on mass})}{m}$$

Multiply by m:
$$mv = mu + I(\text{on mass})$$
$$I(\text{on mass}) = mv - mu$$

This expression shows that the impulse on the mass is the difference between two quantities derived from the product of the mass and its velocity. This quantity is called **momentum**, $\vec{p}$. For any mass m travelling at velocity $\vec{v}$, its momentum is found by:

$$\vec{p} = m\vec{v}$$

The impulse of the force acting on the mass is therefore the *change* in the momentum of the mass, $\Delta\vec{p}$:

$$\vec{I}(\text{on mass}) = \Delta\vec{p}$$

The arrows over the symbols are a reminder that these are vector quantities and that their direction matters.

Units for quantities are sometimes written in different but equivalent forms. Using Newton's second law we can show that the unit for impulse, N s, is indeed equivalent to the unit for momentum, kg m s^{-1}:

$$\begin{aligned} \text{Impulse} &= \text{force} \times \text{time} && (\text{N s}) \\ &= (\text{mass} \times \text{acceleration}) \times \text{time} && (\text{kg m s}^{-2}\,\text{s}) \\ &= \text{momentum} && (\text{kg m s}^{-1}) \end{aligned}$$

Momentum is a conserved quantity

A conserved quantity is one for which the sum of the initial amounts is equal to the sum of the final amounts. Unlike energy, for which transformations occur between kinetic energy and potential energy, there are no transformations of momentum into other forms.

Isaac Newton used the idea of momentum – he called it 'quantity of motion' – when he developed his three laws of motion. Newton noticed that, when two things interacted, one lost this 'quantity of motion', or momentum, and the other gained the same amount. The total amount of momentum was therefore the same before and after the interaction.

Further, Newton realised that, as the interaction occurred continuously, there was no moment when the two quantities were not being exchanged and that, at all moments in an interaction, momentum was conserved. Newton's third law was derived using these considerations. We can show this through a further analysis of the impulse–momentum equivalence for the simplest of interactions, a collision between two objects.

Collisions between two objects

Two objects, A and B, of momenta $(\vec{p}_A)_i$ and $(\vec{p}_B)_i$ respectively, collide. The impulse of their interaction transfers momentum. Their final momenta are $(\vec{p}_A)_f$ and $(\vec{p}_B)_f$ respectively. The sum of final momenta and the sum of the initial momenta are equal:

$$\begin{aligned} &(\vec{p}_A)_i + (\vec{p}_B)_i = (\vec{p}_A)_f + (\vec{p}_B)_f \\ &\Rightarrow (\vec{p}_A)_f - (\vec{p}_A)_i = (\vec{p}_B)_i - (\vec{p}_B)_f \\ &\Rightarrow (\vec{p}_A)_f - (\vec{p}_A)_i = -[(\vec{p}_B)_f + (\vec{p}_B)_i] \quad \dots (1) \end{aligned}$$

The impulse of the force by A on B transfers momentum to B. Simultaneously, and over the same time interval, the impulse of the force by B on A transfers momentum to A:

$$\vec{F}(\text{by A on B}) \times \Delta t = (p_B)_f - (p_B)_i$$

and

$$\vec{F}(\text{by B on A}) \times \Delta t = (p_A)_f - (p_A)_i$$

Taken together with equation (1), this means that:

$$\vec{F}(\text{by A on B}) \times \Delta t = -\vec{F}(\text{by B on A}) \times \Delta t \quad \text{... (2)}$$

$$\Rightarrow \vec{F}(\text{by A on B}) = -\vec{F}(\text{by B on A}) \quad \text{... (3)}$$

This is Newton's third law. The forces are equal in size, opposite in direction and act on different objects. Because they act on different objects, they cannot both be used as part of a net force equation (Newton's second law) on one object, either A or B. Thus, $\vec{F}(\text{by A on B})$ can be used with other forces acting on B to form a net force *on* B. Similarly, $\vec{F}(\text{by B on A})$ can only be used with other forces acting on A to form a net force. As $\vec{F}(\text{by A on B})$ does not act on A, it cannot be used.

The equation for Newton's third law (equation 3) can be transposed to:

$$\vec{F}(\text{by A on B}) + \vec{F}(\text{by B on A}) = 0$$

This way of writing Newton's third law is quite common. You might feel justified to conclude, incorrectly, that these forces can be added to make a net force and that force is conserved. But this way of writing an equation for Newton's third law hides the fact that it is really the impulse–momentum equation, and the time interval has been divided out.

WORKED EXAMPLE 9.4

1 A 1.5×10^3 kg car is moving to the right at 20 m s^{-1} while a 5.0×10^3 kg truck is moving to the left at 10 m s^{-1}. The car and truck collide and move off as one mass, stuck together.

- a Draw a diagram showing this situation. (1 mark)
- b Describe how the momentum of the truck will change. (1 mark)
- c What is the velocity of the wreckage immediately after the collision? (1 mark)

2 Two billiard balls of equal mass, one blue the other red, slide towards each other at 2.0 m s^{-1} in opposite directions towards each other. They collide head-on and rebound, both at 1.8 m s^{-1} in directions opposite to their original velocities, as shown in Figure 9.19.

Show how momentum is conserved in this collision. (3 marks)

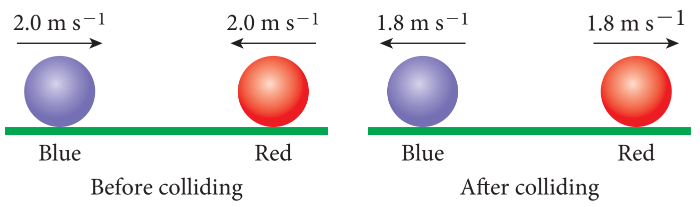


Figure 9.19 The two billiard balls before and after colliding

Answers

1 a

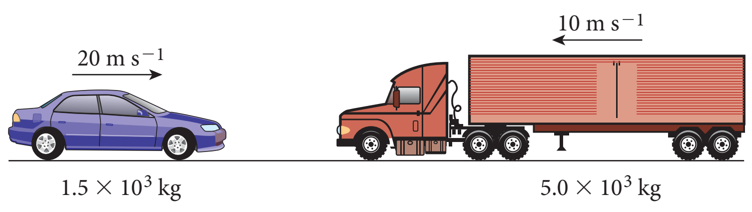


Figure 9.20 The car and truck before they collide

b Conservation of momentum of the system: The sum of the momentums of the car and truck after the collision is the same as before the collision.

Logic

1 a Draw a correct diagram showing mass and velocity. — 1 mark

b Use the conservation law. — 1 mark

c Taking to the right as the positive direction:

$$(m_{car} + m_{truck})\vec{v}_{car+truck} = m_{car}\vec{u}_{car} + m_{truck}\vec{u}_{truck}$$

$$\vec{v}_{car+truck} = \frac{m_{car}\vec{u}_{car} + m_{truck}\vec{u}_{truck}}{m_{car} + m_{truck}}$$

Rearrange to make *v* the subject. — 1 mark

$$= \frac{(1.5 \times 10^3\ \text{kg})(20\ \text{m s}^{-1}) + (5.0 \times 10^3\ \text{kg})(-10\ \text{m s}^{-1})}{(1.5 \times 10^3 + 5.0 \times 10^3)\ \text{kg}}$$

$$= -3.1\ \text{m s}^{-1}$$

The wreckage is moving at 3.1 m s^{-1} to the left just after the collision occurs.

Calculate the correct speed and direction. — 1 mark

2 Conservation of momentum of the system:

The sum of the momentums of the two balls after the collision is the same as before the collision.

Taking to the right as the positive direction:

Use conservation of momentum. — 1 mark

$$\vec{p}(\text{system})_{after} = \vec{p}(\text{system})_{before}$$

$$m_B\vec{v}_B + m_R\vec{v}_R = m_B\vec{u}_B + m_R\vec{u}_R$$

$$m_B \times 2.0\ \text{m s}^{-1} + m_R \times -2.0\ \text{m s}^{-1} = m_B \times -1.8\ \text{m s}^{-1} + m_R \times 1.8\ \text{m s}^{-1}$$

Substitute the correct values. — 1 mark

But $m_R = m_B$: $\Rightarrow 0 = 0$

Momentum has been conserved.

Give reasoning and answer. — 1 mark

Try these yourself

1 Two asteroids of equal mass, alpha and beta, are moving in opposite directions at 200 m s^{-1} to the right and 300 m s^{-1} to the left respectively, towards each other. They collide and form one mass. Draw a diagram. Predict and then calculate the velocity of the combined mass after the collision. (4 marks)

2 Find the change in the momentum of a 50 kg occupant of the car in the previous worked example. Find the average force applied to the occupant. Assume the change in the occupant's momentum was effected by the application of a force lasting 0.80 s. (5 marks)

ACTIVITY 9.1

NEWTON'S CRADLE

Aim

To observe nearly-elastic collisions and the conservation of kinetic energy as well as the conservation of momentum in action

You will need

Access to a Newton's cradle toy and/or search for an online video using the search words 'Newton's cradle video'

What to do

Observe the behaviour of a Newton's cradle when one, two and three balls are pulled back and released.

What did you discover?

Relate your observations to the conservation laws mentioned above.

Can you make the Newton's cradle behave differently?

▲ **Figure 9.21** An example of a Newton's cradle

WOW

Small rocks with large energy

Comets and asteroids periodically slam into Earth. Even a relatively small comet approximately 100 m in diameter and moving at 30 km s^{-1} has sufficient kinetic energy to do significant damage when it hits Earth, even in unpopulated areas. In 1908, such an object is thought to have exploded just above the ground over Siberia, flattening trees for a radius of 50 km. The explosion released about 1000 times more energy than the atomic bomb that destroyed the Japanese city of Hiroshima in 1945. Such events have prompted scientists to monitor the orbits of these objects and to have a plan to deflect them well before they hit Earth.

Elastic collisions

There are occasions when objects can 'collide' without producing sound or heat. If this occurs, the total kinetic energy of the objects in the system is conserved. Recall from Chapter 1 that such collisions are said to be elastic collisions.

Perfectly elastic collisions can occur when a non-contact force between the colliding objects is applied by each object on the other. Magnetic air pucks, electrostatically charged balls and objects in space can 'collide' without touching. The non-contact force – magnetic, electrostatic or gravitational – causes a change in the velocities of the objects.

Nearly perfect elastic collisions occur when objects collide briefly and rebound. Metal ball-bearings, such as those found in Newton's cradles collide without losing much of their kinetic energy.

Inelastic collisions

In Worked example 9.4, some of the kinetic energy of the objects in the systems was lost during the collisions. This lost kinetic energy must be accounted for in some other energy transformation that occurs during the collision. This must be done so that the law of conservation of energy, which covers all energy forms, is obeyed. These collisions would involve the production of sound and heat, which is lost to the surroundings beyond the system and equals the lost kinetic energy of the objects within the system. Such collisions are called **inelastic collisions**.

Inelastic collisions involve a loss of total kinetic energy of the objects within the system. If sound or heat (caused by the deformation of the colliding objects) is produced, that energy must have come from somewhere. It has come from the decrease in the objects' kinetic energy. The law of conservation of energy holds.

Elastic collision

Momentum is conserved throughout the collision.

Total kinetic energy before the collision is the same as total kinetic energy after the collision. Some energy is transformed to potential energy.

Inelastic collision

Momentum is conserved throughout the collision.

Total kinetic energy before the collision is different from total kinetic energy after the collision. Some energy is transformed to potential energy, and some energy is stored in permanent deformations of structures, and transformed to sound, light, heat etc., which are effectively unavailable for return to kinetic energy.

WORKED EXAMPLE 9.5

How much kinetic energy is lost as heat and sound when the car and the truck shown in Figure 9.19 collide? (2 marks)

Answer	Logic	
E_k of system before collision	Calculate E_k before and after the collision.	1 mark
$= E_k(\text{car}) + E_k(\text{truck})$		
$= \frac{1}{2}m_{(\text{car})} \times v_{(\text{car})}^2 + \frac{1}{2}m_{(\text{truck})} \times v_{(\text{truck})}^2$		

$= \frac{1}{2} \times 1.5 \times 10^3\,\text{kg} \times (20\,\text{m s}^{-1})^2$

$+ \frac{1}{2} \times 5.0 \times 10^3\,\text{kg} \times (10\,\text{m s}^{-1})^2$

$= 5.5 \times 10^5\,\text{J}$

E_k of system after collision

$= E_k$(car and truck combined)

$= \frac{1}{2} m_{(\text{car + truck combined})} \times v_{(\text{car + truck combined})}$

$= \frac{1}{2} \times 6.5 \times 10^3\,\text{kg} \times (3.1\,\text{m s}^{-1})^2$

$= 3.1 \times 10^4\,\text{J}$

E_k lost during collision	Calculate loss of E_k.	1 mark

$= 5.5 \times 10^5\,\text{J} - 3.1 \times 10^4\,\text{J}$

$= 5.2 \times 10^5\,\text{J}$ which is most of the original kinetic energy of the system.

Try this yourself

Given that each of the billiard balls shown in Figure 9.18 has a mass of 100 g, how much E_k was lost when they collided, as shown? (2 marks)

EXPERIMENT 9.2

MOMENTUM DURING A COLLISION

Aim

To analyse the momentum during a collision

Materials

- two dynamics trolley carts
- motion sensor and data logger (or ticker tape timing apparatus)
- springs

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
The spring may flick back or flick an object into a person's eye.	Wear safety glasses when working with springs.

In your write-up, add any more risks you can think of, as well as ways to manage them.

Procedure

1. Determine the mass of the two dynamics trolleys carts.
2. Tie the two trolleys together using string with a compressed spring between them, as shown in Figure 9.22.
3. Release the trolley cars by cutting or burning the string holding them together.
4. Measure and record the velocities of each trolley car after the 'collision' (the explosion, or cutting of the string). Repeat this three times using the same spring and compression, Δx, each time, to obtain an average result.

5 Find the momentum of the system of the two trolley cars before and after each collision. Find the change in the momentum of the system.

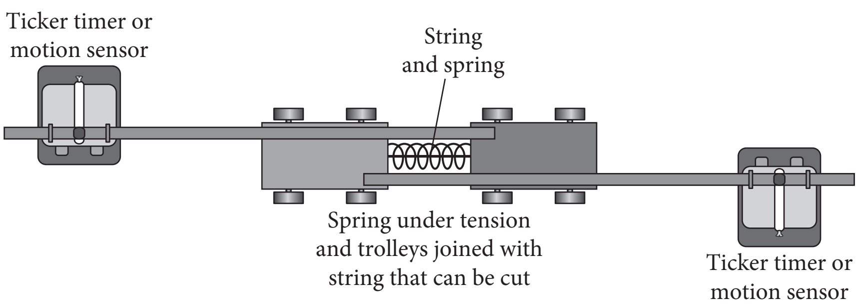


Figure 9.22 ◀
The set-up of the apparatus for the experiment

4 We usually think of a collision as objects hitting each other. Explain how the situation of the trolleys flying apart can also be considered an explosion.

Results and analysis of results

Enter your results into a table like the one below.

	Mass of car A (kg)	*v* of car A after string is cut	*p* of car A ($kg\,m\,s^{-1}$)	Mass of car B (kg)	*v* of car B after string is cut	*p* of car B ($kg\,m\,s^{-1}$)	Δp of the system ($kg\,m^{-1}$)
Trial 1							
Trial 2							
Trial 3							
Average							

Discussion

1 Is the total momentum conserved in the collisions? If not, why not?

2 Is the total kinetic energy conserved in the collisions? If not, identify the source of the kinetic energy of the trolleys.

3 Suggest how the potential energy stored in the spring could be found.

INVESTIGATION 9.1

WEIGHING A SKATEBOARD RIDER WITHOUT A BALANCE

A moving skateboard will slow down when a person jumps sideways onto it. A stationary skateboard rider will start to move, and a moving skateboard rider will slow down when they collide and move off together.

What is your aim?

Decide which one or both scenarios you will use to collect data. Then write an aim in the form 'To … by measuring …'.

What will you need?

Decide what you will need to do the experiment. A skateboard would seem to be essential, but you might choose another, safer carriage.

What measuring system will you use? How will you set it up to collect good data?

What are the risks?

Consider all risks to a person's safety and ways to reduce or eliminate them. Skateboard riders have a number of essential items of safety equipment – knee pads, wrist guards etc. Make sure they are listed against the relevant risk.

Construct a table similar to the one below. Identify specific risks involved in the investigation and ways that you will manage the risks to avoid injuries or damage to equipment. Consult a skateboard safety manual.

What are the risks in doing this investigation?	How can you manage these risks to stay safe?

How will you carry out your investigation?

Write a list of steps you intend to undertake in order to collect sufficient, good data to achieve your aim.

Safety: Do not attempt this investigation until you have approval from your teacher.

What results will you collect?

What data will you collect? How?

How will you record your data?

What will you do to reduce errors and estimate uncertainties?

How will you analyse your results?

Think about how you are going to display your results. What techniques will you use to analyse your results? What theory will you use to analyse the data so you can achieve your aim?

What have you found?

Display your results.

What do you conclude?

Write a conclusion consistent with the aim of your investigation. Use your results as evidence to support your conclusion.

Ideas for improvement or further investigation

Consider whether your plan was satisfactory. What could you have done better?

What did you find out about the process of doing an investigation?

How confident are you that your results are defensible?

The physics of car safety

In 2013, 1193 people were killed on Australian roads compared with 3798 in 1970, even though there are many more cars on the roads now than 40 years ago.

Physics has, in part, led to the dramatic fall in the road toll in Australia over the past 40 years. Many would say that 1193 is still 1193 too many; however, the application of physics has made cars safer than ever before.

When a car collides with a stationary object a rapid deceleration occurs. The occupants of the car must have a force applied to them so that they decelerate with the car. The alternative is to be thrown out of the car, a far more dangerous proposition. For a car travelling at $90\,\text{km}\,\text{h}^{-1}$ ($25\,\text{m}\,\text{s}^{-1}$), a sudden stop involves a change in the momentum of the car and its occupants. Assuming an occupant has a mass of 60 kg, this change in momentum is:

$$\Delta p = 60\,\text{kg} \times 25\,\text{m}\,\text{s}^{-1} = 1500\,\text{kg}\,\text{m}\,\text{s}^{-1}$$

Crumple zones

This change in momentum must be transferred from the car to the occupant.

It was once believed that a car should be made as strong and as rigid as possible to protect its occupants. Modern cars are now designed with a crumple zone. This is a specially designed area surrounding the passenger compartment that is actually supposed to crumple in a collision.

CRASHING A SMART CAR AT 110 KM H^{-1}

Even modern small cars are designed with a crumple zone as this video clip shows.

The reason for a crumple zone lies in the way in which the impulse is given to the occupant. From the equation $I = F \times \Delta t$, we can see that the longer the time interval, Δt, the smaller the force, F, that is applied to the occupant. Allowing the front of the car to crumple increases the time taken for the car to stop. This decreases the forces applied to the occupants.

Alamy/CTK

▲ **Figure 9.23**
The crumple zones in these crash test cars are clearly visible.

WORKED EXAMPLE 9.6

A car travelling at $90\,\text{km}\,\text{h}^{-1}$ ($25\,\text{m}\,\text{s}^{-1}$) collides with a stationary immovable object and comes to a stop in 0.08 s. Assuming the car decelerates at a constant rate, find the force applied to a 60 kg occupant who remains in the vehicle. (2 marks)

Answer	Logic	
$I(\text{given to occupant}) = \Delta p$ $F(\text{on occupant}) \times t = mv - mu$	Use the correct equations.	1 mark
$F(\text{on occupant}) = \dfrac{(mv - mu)}{t}$ $= \dfrac{60\text{ kg} \times 0\text{ m s}^{-1} - 60\text{ kg} \times 25\text{ m s}^{-1}}{0.08\text{ s}}$ $= -1.9 \times 10^4\text{ N}$	Calculate the answer.	1 mark

The negative answer indicates that the direction of the applied force on the occupant is opposite to the direction of motion.

This force is the equivalent of having an almost 2 tonne elephant sitting on your chest and hips for this time, a most unpleasant experience.

Try this yourself

The immovable object with which the car collides in the previous example now has a collapsible guard rail around it that increases the time taken for the car to stop to 0.20 s. Find the average force applied to the occupant in this case. (2 marks)

Combining energy and momentum

In many cases in which the analysis of motion and forces is performed, work–energy and impulse–momentum ideas can be used interchangeably.

Seatbelts

The compulsory wearing of seatbelts means that the force is applied to the occupant across the strongest parts of the human body – the chest and hips. If worn incorrectly seatbelts can cause injuries to the abdomen and other soft parts of the body. Seatbelts are designed to stretch a small amount, which also assists in decreasing the maximum force applied to the wearer while the impulse is being given during the collision.

In the 1960s, before seatbelt laws were introduced, many cars did not have seatbelts, and many people did not wear them, even if they were fitted. The introduction in 1972 of laws making the wearing of seatbelts compulsory throughout Australia has been credited with a large decline in the road toll. Without a seatbelt, the occupant's inertia results in the person continuing to move forwards until coming into contact with some part of the car. Striking the steering wheel, dashboard or windscreen at such speeds is often fatal.

WOW

Saving thousands of Australians' lives

Seatbelt wearing and anti-drink-driving laws together have had the most impact in lowering the road toll in Australia over the past 40 years. Although authorities say that you can never eliminate the 'idiot factor', raising the awareness of the general population has had a positive influence on the behaviour of most people. The next big issue is that of driver distraction by phones, global positioning systems and other in-car devices.

WORKED EXAMPLE 9.7

A 2000 kg car travelling at 14 m s^{-1} crashes into a rock wall and comes to a stop in 0.5 m. Use work–energy and impulse–momentum ideas respectively to determine:

a the average force applied to the car by the rock wall. (3 marks)

b the time taken for the car to come to a stop. (3 marks)

Answers	Logic	
a Work–energy	Correctly use work–energy linkage.	1 mark
$W(\text{done on car}) = \Delta E_k = \frac{1}{2}mv^2$		
$F(\text{on car}) = \frac{1}{2}\frac{mv^2}{s}$		
$= \frac{1}{2}\frac{(2000\text{ kg})(14\text{ m s}^{-1})^2}{0.5\text{ m}}$	Substitute the correct values.	1 mark
$F(\text{on car}) = 3.9 \times 10^5\text{ N}$	Calculate the answer.	1 mark
b Impulse–momentum	Correctly use impulse–momentum linkage.	1 mark
$F(\text{on car}) \times t = \Delta p = mv$		
$t = \frac{mv^2}{F(\text{on car})}$	Substitute the correct values.	1 mark
$t = \frac{2000\text{ kg} \times 14\text{ m s}^{-1}}{3.9 \times 10^5\text{ N}}$		
$= 7.2 \times 10^{-2}\text{ s}$	Calculate the answer.	1 mark

Try these yourself

A trail bike and rider come to a stop from 12 m s^{-1} in 2.5 s. The combined mass of trail bike and rider is 250 kg.

a Find the average force applied. (3 marks)

b Calculate the distance covered. (3 marks)

Airbags

Airbags, or supplemental restraint systems, were introduced into cars in the 1980s. Front airbags, as shown in Figure 9.24, are designed to increase the time taken for an occupant to decelerate. They also spread more evenly the load over which the force is applied. Later developments, including side and curtain airbags, prevent an occupant's head making impact with hard parts of the car during collisions other than front-on collisions. In every case, the time taken to effect a change in momentum is increased so that the maximum force applied to any part of an occupant is decreased.

Figure 9.24 ▶ Air bags help protect occupants in a collision.

Alamy/Photo Network/Tom Tracy

QUESTION SET 9.4

Remembering

1 In what circumstances is momentum conserved in a collision in an isolated system?

2 Define 'impulse'.

3 For any given impulse, how can the average applied force be decreased?

Understanding

4 Why is it desirable to have a crumple zone in a vehicle?

5 Two metal balls collide and move off. By observing the collision, how could you tell if it was elastic or inelastic?

▶

Applying

6 A force of 500 N is applied for 4.0 s to a 100 kg satellite in space.

  a What impulse is given to the satellite?

  b What is the change in the satellite's momentum?

7 A 50 kg ice-skater exerts an average force of 25 N on the ice in her direction of motion for 3.0 s.

  a If the skater is initially moving at 4.5 $m\,s^{-1}$, what is her final speed?

  b If the force the skater applied to the ice had been 75 N, for how long would she have needed to apply it to reach the same speed?

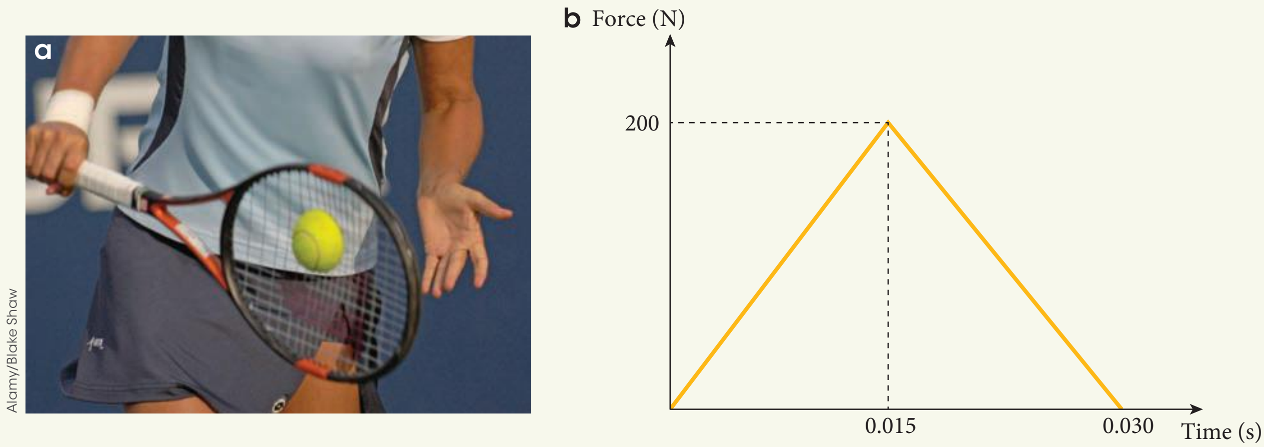
8 A 58 g tennis ball travelling at 30 $m\,s^{-1}$ to the right is struck by a tennis racquet and returned along the same direction (Figure 9.25(a)).

The force applied by the tennis racquet to the ball is modelled by the graph in Figure 9.25(b).

  a What is the impulse applied to the ball?

  b What is the change of momentum of the ball?

  c With what speed does the ball begin to return?



▲ Figure 9.25

a) A tennis ball is deformed when struck by a racquet b) The *F*–*t* graph for a tennis ball being struck.

Analysing

9 The speed of an object of known mass and speed is forced to change. What determines whether you will start with a work–energy or an impulse–momentum analysis?

Reflecting

10 What is similar and what is different about the concepts of work–energy and impulse–momentum?.

CHAPTER SUMMARY

Work-energy

- Work is done by a force acting through a distance.
- Work leads to a change in energy.
- When a force acts through a distance, work is done and energy is transferred.
- $W(\text{by A on B}) = \vec{F}(\text{by A on B}) \bullet \vec{s}$
  $W(\text{by A on B}) = F(\text{by A on B}) \times \cos\theta \times s$
- $W(\text{by A on B})$ = area under F–s graph
- The work done by a steadily increasing force is the area under the F–s graph:
  $$W = \tfrac{1}{2}Fs$$
- Hooke's law: the force applied by a spring is proportional to the extension of the spring:
  $$\vec{F}(\text{by spring}) = k(-\Delta\vec{x})$$
- k is a measure of the stiffness of the spring
- $W = \tfrac{1}{2}k(\Delta x)^2$
- Law of conservation of energy: In an isolated system, energy is neither created nor destroyed.
- $\Delta E_k + \Delta E_p = 0$
- $\Delta W = \Delta E$
- $W(\text{on mass}) = \Delta E_k$
- $E_k = \tfrac{1}{2}mv^2$
- Elastic collision: Both kinetic energy and momentum are conserved in the system.
- Inelastic collision: Only momentum is conserved; some kinetic energy of the system is lost as other forms of energy (sound, light, heat, deformation, potential energy).
- Near Earth:
  - $E_{total} = E_k + E_p = \tfrac{1}{2}mv^2 + mg\Delta h$
  - $\Delta E_{total} = 0$
    $\Rightarrow +\Delta E_p = -\Delta E_k$
    or
    $-\Delta E_p = +\Delta E_k$
- Power:
  $$P = \frac{\Delta W}{\Delta t} = \frac{\Delta E}{\Delta t} = Fv$$

Impulse-momentum

- Force acting over a time leads to impulse.
- Impulse causes to a change in momentum.
- When a force acts over a time, impulse is applied and momentum is transferred.
- $I(\text{by A on B}) = \vec{F}(\text{by A on B}) \times t$
- $I(\text{by A on B})$ = area under F–t graph
- The impulse for a steadily increasing force is the area under the F–t graph:
  $$I = \tfrac{1}{2}Ft$$
- Law of conservation of momentum: In every collision, total momentum is always constant.
- $\vec{I}(\text{on mass}) = \Delta\vec{p}$
- $\vec{p} = m\vec{v}$
- In all collisions, momentum is conserved.
- $\vec{p}_{total} = \Sigma\vec{p}_{i = \text{constant}}$
  or
  $\Delta\vec{p} = 0$

CHAPTER GLOSSARY

gravitational potential energy the energy associated with the gravitational force acting on an object

Hooke's law the force applied by a spring is proportional to the extension of the spring

ideal spring a spring that obeys Hooke's law

impulse the action of a force over a time interval; the change in momentum

inelastic collision a collision between two or more objects in which momentum is conserved but total kinetic energy is not

momentum a quantity related to the action of a force over a time interval; it is the product of the mass and velocity of an object

stiffness a measure of how much force is required to extend or compress a spring (the spring constant)

CHAPTER REVIEW QUESTIONS

Remembering

1 What is the purpose of the arrow over a quantity, for example, $\vec{F}$?

2 What is the SI unit of work?

3 What is the quantity calculated by finding the area under a F–s graph?

4 What other factor must be found to calculate the change in the momentum of an object when the force applied to the object is known?

5 What observations could be made to determine that a collision was not elastic?

6 What is a crumple zone in a vehicle?

7 Why have seatbelts been so important in the decrease in the road toll?

8 What does 'conservation' mean in the context 'conservation of momentum'?

Understanding

9 How could you tell if work has been done on an object?

10 Why is the force required to extend a spring the same magnitude as the force exerted by the spring?

11 How is the stiffness of a spring related to the gradient of the F–Δx graph for the spring?

12 How does a falling object gain its kinetic energy, given that energy cannot be created or destroyed?

13 A student carries a heavy backpack (it must be the maths textbook) to school along a level street. The pupil says 'that was hard work'. Evaluate the pupil's statement.

14 A meteor enters the atmosphere and slows down quickly while burning up and vaporising. Describe how the law of conservation of energy applies to this event.

15 The impulse given to the occupant of a car with a crumple zone is exactly the same as the impulse given to the occupant of a rigid vehicle. Why are crumple zones considered desirable?

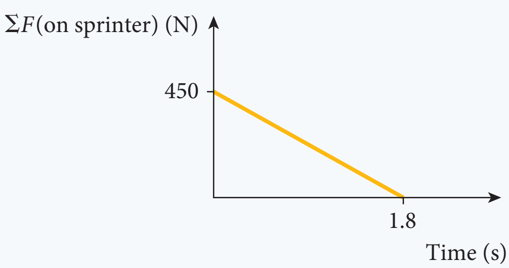
Applying

16 The kinetic energy of a 200 kg satellite is raised by 4.0×10^7 J in a 10-minute rocket burn. What was the average power of the rocket motors?

17 Figure 9.26 shows the net force acting on a 60 kg sprinter at the start of a race.

a What is the sprinter's speed 1.8 s after the start?

b Assuming the net force on the sprinter remains zero after 1.8 s, explain why the sprinter must still use considerable energy to complete the race.



◀ Figure 9.26

18 A force of 25 N is applied to a coil spring and compresses the spring by 5.0 cm.

- **a** What is the spring constant, *k*, of this spring?
- **b** How much work must be done on the spring to compress it by 5.0 cm?
- **c** Where does this work energy go when the spring is compressed?

19 Using the work–energy idea, calculate the speed with which a 10 kg rock released from a height of 30 m hits the ground.

20 With what constant vertical speed could a 2.5 kW winch motor lift an injured hiker into the helicopter if the hiker and stretcher have a combined mass of 100 kg?

Analysing

21 Would it be possible to have a collision in which neither kinetic energy nor momentum are conserved?

22 Show that, if conservation of momentum is applied, the result of a Newton's cradle collision cannot be anything other than what is observed to happen.

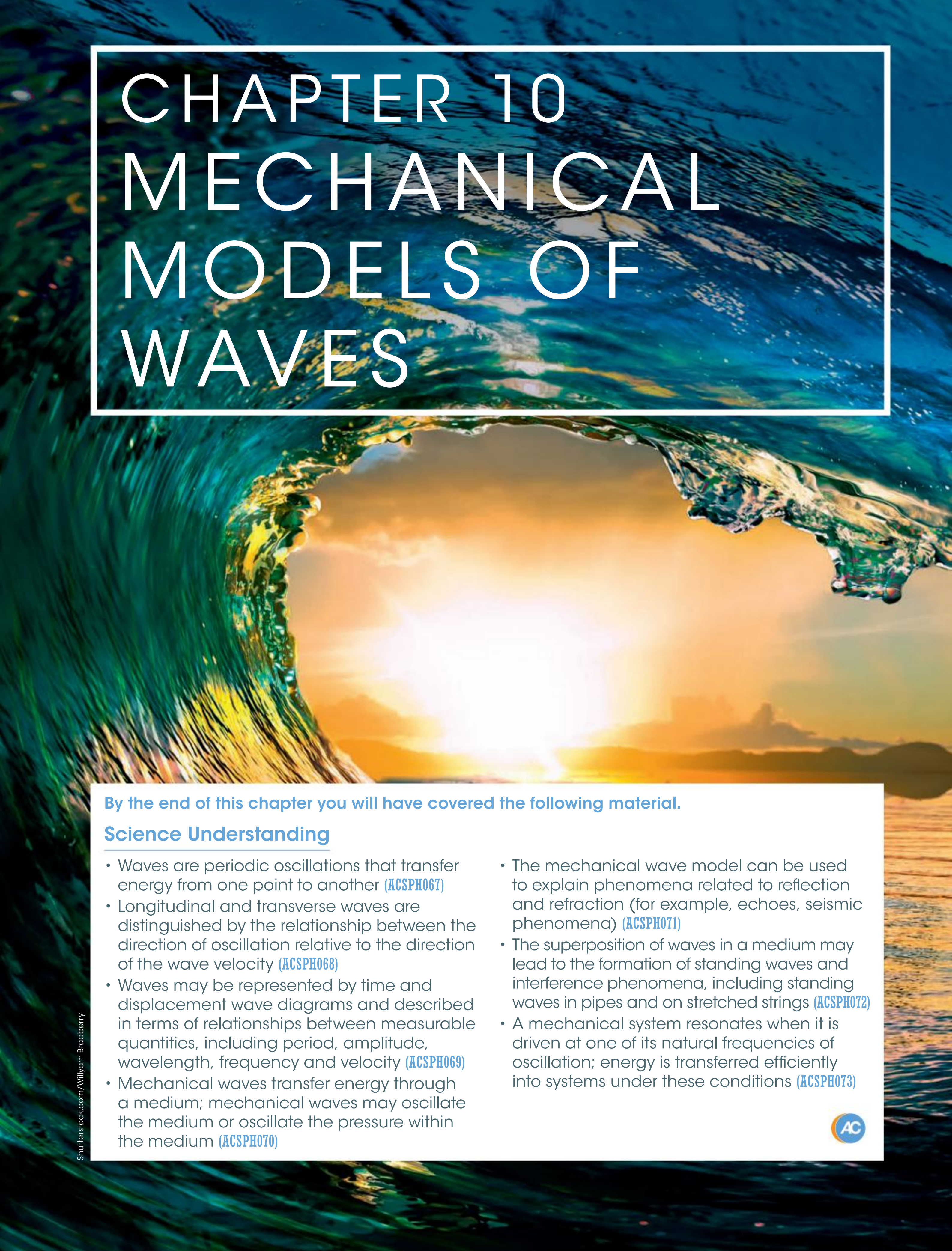
23 When a car brakes suddenly from 20 m s^{-1}, the head of a 90 kg occupant comes to a stop on the windscreen. The skull is pushed inwards towards the brain by 3.5 cm. The mass of the head is 8% of the total body mass.

- **a** What is the mass of the occupant's head?
- **b** What is the kinetic energy of the head when it hits the windscreen?
- **c** With what force does the windscreen strike the head?
- **d** How long did it take for the head to be crushed by the windscreen?

Reflecting

24 Why is the study of physics so important to car safety?

25 An object is said to transfer energy, but never to transfer force. Explain the difference between the two ideas.



Shutterstock.com/Willyam Bradberry

CHAPTER 10 MECHANICAL MODELS OF WAVES

By the end of this chapter you will have covered the following material.

Science Understanding

- Waves are periodic oscillations that transfer energy from one point to another (ACSPH067)
- Longitudinal and transverse waves are distinguished by the relationship between the direction of oscillation relative to the direction of the wave velocity (ACSPH068)
- Waves may be represented by time and displacement wave diagrams and described in terms of relationships between measurable quantities, including period, amplitude, wavelength, frequency and velocity (ACSPH069)
- Mechanical waves transfer energy through a medium; mechanical waves may oscillate the medium or oscillate the pressure within the medium (ACSPH070)
- The mechanical wave model can be used to explain phenomena related to reflection and refraction (for example, echoes, seismic phenomena) (ACSPH071)
- The superposition of waves in a medium may lead to the formation of standing waves and interference phenomena, including standing waves in pipes and on stretched strings (ACSPH072)
- A mechanical system resonates when it is driven at one of its natural frequencies of oscillation; energy is transferred efficiently into systems under these conditions (ACSPH073)

Shutterstock.com/Kostiantyn Fastov

Figure 10.1 ▲ Concentric circular waves formed by drops of water

Introduction

Water waves, sound and seismic waves are examples of mechanical waves. A single ripple or pulse on a pond spreads out, losing amplitude. A leaf floating nearby moves up and down around its original position but does not travel with the pulse.

We are surrounded by sounds: the chirping of birds in the bush or the roar of the traffic on a busy road. Sound energy can be carried to our ears through any medium by **sound waves**. The air does not move in a rush from the source to our ears but, rather, the energy rippling through the medium causes the local vibrations of particles. A gossamer-light spider's web in the path of the sound will move backwards and forwards as the wave passes.

Sound waves vibrate the eardrum backwards and forwards. The energy is converted into electrical impulses in the ear and sent to the brain where it is interpreted. The ear is a delicate structure that needs to be protected from sounds that are too intense. Science has developed strategies and devices to prevent acoustic damage as well as methods of repairing or enhancing impaired hearing.

Seismic waves travel through Earth from explosions and earthquakes. They carry information from within Earth.

WOW

Large, destructive water waves

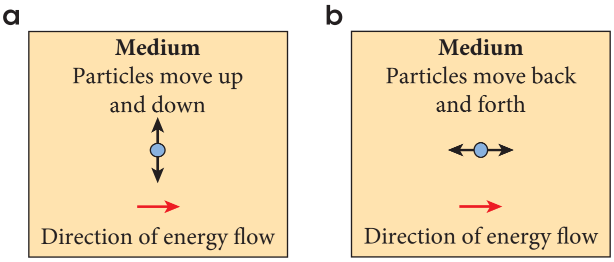
Larger forms of water waves, such as ocean swells and tsunamis, can carry huge amounts of energy. The massive tsunami that occurred off the coast of Sumatra in 2004 had enormous destructive power; 240 000 people were killed and millions of homes were destroyed. The whole of the Indian Ocean basin was affected. A huge earthquake thrust 1600 km of sea floor upwards, shifting massive amounts of water. The resultant tsunamis travelled at about 600 km h^{-1}. When they reached shallower water near land they slowed down, the wave heights grew to 30 metres and they penetrated well inland.

AAP Images/AP

▲ Figure 10.2
Banda Aceh, the capital of Aceh province, Indonesia, after the 2004 tsunami

The mechanical wave model

Sound and water waves are **mechanical waves**. Mechanical waves transfer energy through a medium without the whole medium being transported from place to place. The particles making up the medium are free to move a little from their usual positions. They do not move forwards together in bulk. To a first approximation, water waves are **transverse waves**. Particles oscillate at right angles to the direction of motion of the wave. Sound waves are **longitudinal waves**. Particles oscillate about mean positions in the same direction as the wave movement.

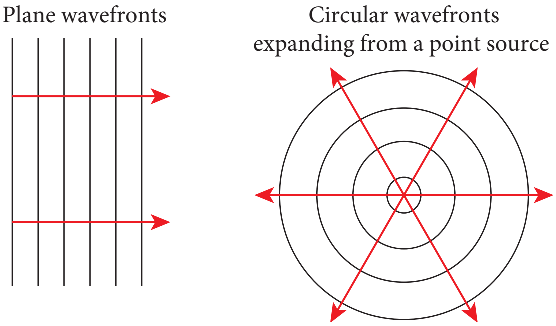


◀ **Figure 10.3**
Mechanical waves cause particles to move around their equilibrium positions:
a) up and down;
b) back-and-forth.

Mechanical transverse and longitudinal waves travel in a material medium made of inter-connected particles that are progressively disturbed. Energy, but not particles, is transferred through the medium.

We can describe a wave using the ideas of wavefronts and rays. A **wavefront** is a surface joining all points in space that are reached at the same instant by a wave propagating through a medium. A **ray** is a line drawn at right angles to the wavefront in the direction of propagation. This description is called the ray model.

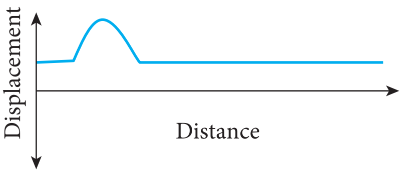
WAVES
This provides an excellent tutorial on the fundamental properties of mechanical waves.



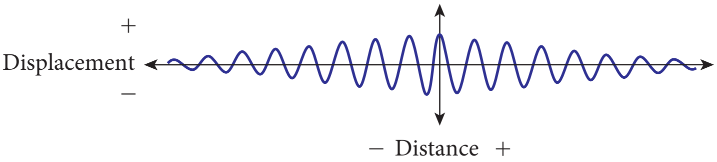
◀ **Figure 10.4**
Wavefront models of waves. The arrows represent the direction of propagation of the wavefronts.

In Figure 10.1 a finger dipped into the water creates surface disturbances that radiate from the point where the finger broke the water. The leading edge of the entire wave forms a circle that is the wavefront. If only a single wavefront passes through a medium it is called a **pulse**.

If the finger is dipped into the water in a regular pattern, wavefronts are produced continuously. A **continuous wave** moves outwards at a constant speed in all directions. As it spreads the energy covers a larger area. The amplitude of the wave diminishes with distance from the source.



▲ **Figure 10.5**
Representation of a wave pulse in a stretched string or spring



▲ **Figure 10.6**
Representation of a continuous wave arising from a central disturbance in a pond

It is important to understand the terminology used to describe waves. Graphical representation of waves allows visualisation of the terms. Plotting a graph of displacement versus time enables a visualisation of the period and the amplitude of a wave.

Period is the time it takes before a wave repeats itself and **amplitude** is the largest distance of the particle from the mean position before returning. The top of the wave is called a **crest** and the bottom a **trough**.

The **frequency** of a wave is the number of crests generated in a time interval.

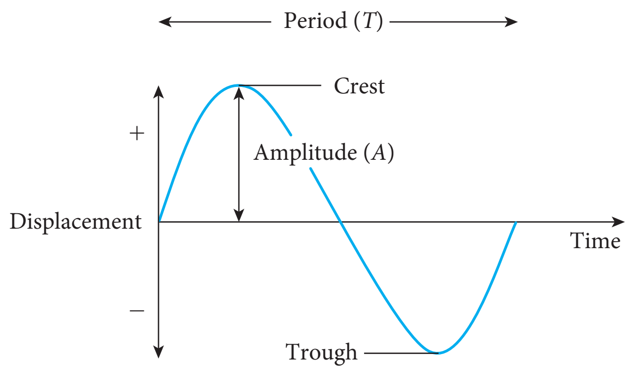


Figure 10.7 ▶ A displacement versus time graph of a mechanical wave represents the displacement of one particle of the medium experiencing a wave disturbance over time. It also shows the amplitude of the wave.

Plotting a displacement verses distance graph enables a visualisation of the wavelength of a wave.

Wavelength is the distance that one wave covers before it repeats itself.

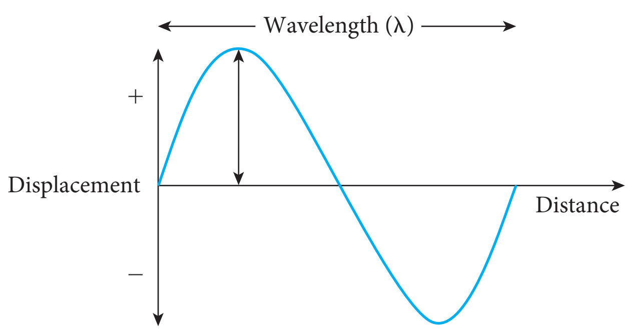


Figure 10.8 ▶ The displacement versus distance graph of a mechanical wave shows the displacement of all the particles of the medium experiencing a wave disturbance at an instant in time.

Table 10.1 Wave terminology

Term	Definition	Symbol	Unit
Amplitude	The largest distance away from the mean position that a particle moves before returning	A	metre (m)
Frequency	The number of crests generated in a time interval	f	hertz (Hz) or (s^{-1})
Wavelength	The distance between successive crests	λ	metre (m)
Period	The time it takes before a wave repeats itself	T	second (s)
Speed	The rate at which a wave covers distance	v	metres per second ($m\,s^{-1}$)

ACTIVITY 10.1

INVESTIGATING WAVE PROPERTIES

Aim

To investigate the properties of transverse and longitudinal mechanical waves

You will need

Two slinky springs, one with a large diameter and one with a significantly smaller diameter

What to do

Use two slinky springs to produce and describe:

- transverse pulses.
- longitudinal pulses.

Work with a partner to prepare a risk assessment.

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
The springs will be stretched and could whip into your eye if accidentally released.	Wear safety glasses and take care when handling springs.

1 Investigating a transverse wave pulse

Stretch out the thinner of the two slinky springs along a bench or on the floor so there is a small amount of tension in it (turns about 5 mm apart). Create a transverse wave pulse in the spring. To do this have your partner hold one end stationary while you give the spring one vigorous shake sideways, to send a single pulse down the spring.

- Draw a diagram showing the direction of the displacement of the coils and the direction of travel of the pulse.
- What part of the spring represents the particles in a medium?
- Draw a diagram showing two things you notice that changed in the pulse after it was reflected.
- Summarise all your findings.

2 Increasing the energy of the wave pulse

Create a more energetic pulse by applying a larger and quicker movement of the hand.

- Did the pulse travel faster?
- What happened to the size (amplitude) of the pulse?
- What difference did your partner feel when the pulse reflected?
- Summarise all your findings.

3 Investigating a longitudinal wave

- Take the larger-diameter spring and stretch it to a low tension (turns about 10 mm apart). Create a longitudinal wave pulse in the spring. To do this have your partner hold one end stationary while you give the spring one vigorous push, to send a single pulse down the spring in the direction of your partner.
- Draw a diagram showing the direction of the displacement of the spring's coils and the direction of travel of the pulse.

4 Comparing springs

Join the ends of the two springs together to make one long spring and stretch the combination with a 'low' tension. Measure the length of each spring under this tension. Stretch each spring to that length before carrying out the task.

Put a transverse wave pulse into both the springs with about the same amplitude.

- What did you notice about the speed of the wave in the smaller-diameter spring compared with the larger-diameter spring?
- Explain your findings.

5 Investigating frequency and wavelength

Create a continuous transverse wave by shaking the spring from side to side at a constant rate (frequency). Then try a number of different constant rates (frequencies).

- Draw two diagrams that show how the rate of shaking (frequency) affects the number of complete waves travelling down the spring.
- Draw two diagrams that show how the distance between the top of each wave (wavelength) varies with the rate of pulse input (frequency).
- Does the speed of the wave vary with the rate of pulse input (frequency)?
- Summarise all your findings.

Transverse wave model

Individual particles of a medium move up and down about their rest position. A series of wave crests and wave troughs moves through the medium. A graph of the position of particles along the medium at one particular time looks like a sine graph (Figure 10.9).

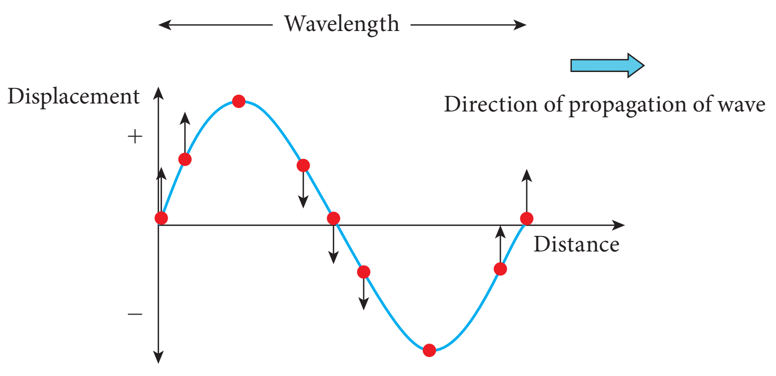


Figure 10.9 ▶ A displacement versus distance graph of a transverse mechanical wave showing the position of the particles at an instant in time.

At a crest or a trough, particles are stationary and about to move towards the mean position. Particles on either side of the crest or trough are moving away from or towards the mean position.

Reflection of transverse waves in strings or springs

When the string or spring is fixed at one end, a crest of a transverse wave is reflected as a trough. When free at one end, a crest of a transverse wave is reflected as a crest.

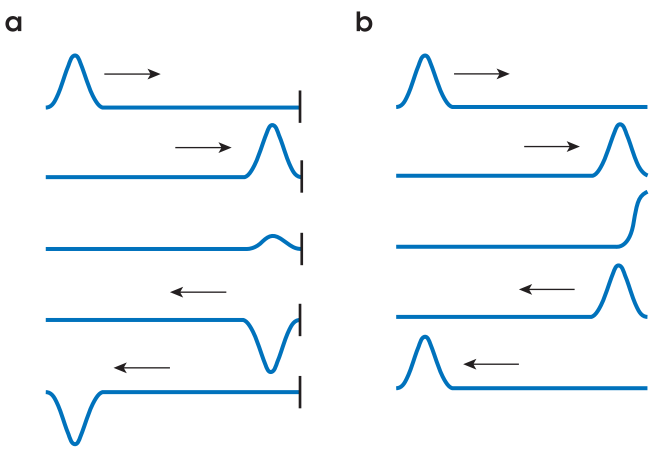


Figure 10.10 ▶ a) A stretched string fixed at one end reflects waves (fixed-end reflection) upside down. The wavelength and frequency are unchanged. b) A stretched string free at one end reflects waves (free-end reflection) the same way up (in phase). The wavelength and frequency are unchanged.

Longitudinal wave model

Particles move back and forth in the same line as the direction of the transfer of energy. The further away from the mean position a particle moves before returning, the greater the amplitude of the disturbance.

When the particles around a point are all moving towards the point, there is a local **compression**. If they are all moving away from the point, there is a local **rarefaction**. A particular point in the medium through which the wave disturbance is travelling experiences a series of compressions and rarefactions (changes to the undisturbed pressure) as the energy passes through it. Figure 10.11 shows a snapshot at an instant in time of where the particles along the top have been displaced to as the wave passes. Maximum pressure occurs where there is minimum displacement.

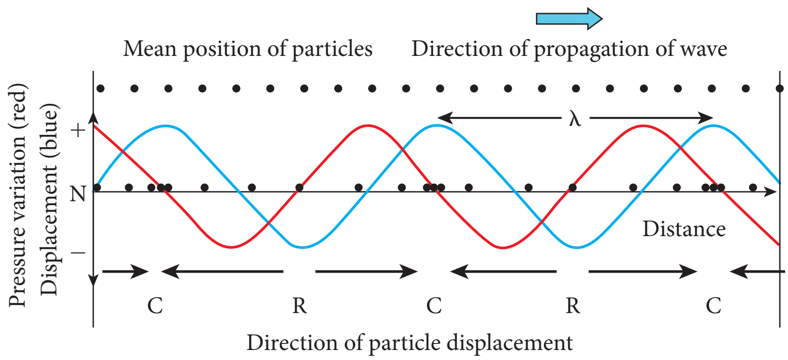


Figure 10.11 Longitudinal wave showing rarefactions (R) and compressions (C) in a medium at the same time

Maximum pressure occurs when the particles are most constrained (blue), hence displaced little. The pressure is lowest when the displacement of particles from their mean positions (red) is greatest (Figure 10.11).

Characteristics of continuous waves

The wave model explains the relationship between amplitude, wavelength, period, frequency and speed that can be used to predict the behaviour of continuous waves.

Consider a slinky spring that you are putting pulses into at a rate of one every 0.5 s. This is the period of the wave (T). The distance between the tops of two adjacent pulses, the wavelength (λ), is found to be 0.6 m. It will take one period for a crest to travel a distance of one wavelength. One period is the time between identical 'snapshots'. This means that the wave in the spring is travelling 0.6 m every 0.5 s.

As:

$$v = \frac{\text{distance}}{\text{time}}$$

$$v = \frac{0.6\,\text{m}}{0.5\,\text{s}}$$

$$v = 1.2\ \text{m s}^{-1}$$

Distance is λ and the time is T, so the formula can be generalised as:

$$v = \frac{\lambda}{T}$$

If the single wave took 0.5 s (T) to generate, then in 1.0 s two crests would have been generated. The number of crests generated per second is the frequency of the wave (f). This gives the relationship:

$$T = \frac{1}{f}$$

$$f = \frac{1}{T}$$

$$f = \frac{1}{0.5}$$

$$f = 2\ \text{Hz}$$

When you substitute $T = \frac{1}{f}$ into $v = \frac{\lambda}{T}$ you get:

$$v = f\lambda$$

Putting in the values:

$$v = 2\,\text{s}^{-1} \times 0.6\,\text{m}$$

$$v = 1.2\,\text{m}\,\text{s}^{-1}$$

WORKED EXAMPLE 10.1

A tsunami travels at 800 km h^{-1} (222 m s^{-1}) across the ocean. It has a wavelength of 150 km. What is the time interval between wave crests? (7 marks)

Answer	Logic	
$v = f\lambda$	Select the correct relationship to find the frequency.	1 mark
$f = \frac{v}{\lambda}$	Rearrange to make *f* the subject.	1 mark
$f = \frac{222\ \text{m s}^{-1}}{1.5 \times 10^5\ \text{m}}$ $f = 1.48 \times 10^{-3}\ \text{s}^{-1}$	Insert values with correct units and solve correctly.	2 marks
$T = \frac{1}{f}$	Select the correct relationship to find *T*.	1 mark
$T = \frac{1}{1.48 \times 10^{-3}\ \text{s}^{-1}}$	Substitute the correct values.	1 mark
$T = 6.76 \times 10^2$ s	Solve correctly.	1 mark

Try these yourself

You dip your finger into a pond at a regular rate of 12 dips per second and observe that the wavefronts created take 24 s to travel the 16 m to the other side of the pond.

a What is the wavelength of the wave? (1 mark)

b What is the period of the wave? (1 mark)

QUESTION SET 10.1

Remembering

1 What are mechanical waves?

2 What is the major difference between a longitudinal and a transverse wave?

3 What is the relationship between the frequency and the period of a wave?

Understanding

4 Where would you find the maximum displacement of the turns in a slinky spring that has a transverse wave passing though it?

- **a** At a crest only
- **b** At a trough only
- **c** At both a crest and a trough

5 How is the displacement of particles in a wave related to the amplitude of the wave?

Applying

6 A person standing 1.2 km from a firing range hears the report of a gunshot. How long did it take the sound to travel from the range to the person? The speed of sound in air is 340 m s^{-1}.

7 A concert pianist plays a single note of frequency 310 Hz. The wavelength is 1.07 m. What is the speed of sound in the air at that time?

Analysing

8 Classify the following examples of mechanical waves as transverse or longitudinal waves.

a A pulse transmitted along a string stretched at right angles to the direction of motion of the pulse

b The wave produced by dropping a stone into a calm pond

c Sound waves produced by a radio

d Waves produced in the air by vibrating vocal cords

Reflecting

9 Summarise your knowledge of waves. How has it changed?

Describing sound waves

In a sound wave, particle oscillations are always parallel to the direction of energy flow; therefore sound waves are longitudinal waves. This sound wave model is not perfect. If, for example, the particles of a gas are in constant random motion, how do they oscillate about a mean position? Models have their limitations and cannot exactly describe physical reality.

Sound transmission as a longitudinal pressure wave

When sound travels through a medium, the particles form a series of compressions and rarefactions as in Figure 10.12(a). At compressions, the pressure is higher than the normal pressure. At rarefactions, the pressure is lower than normal. The graph of pressure variation against the distance from source is shown in Figure 10.12(b).

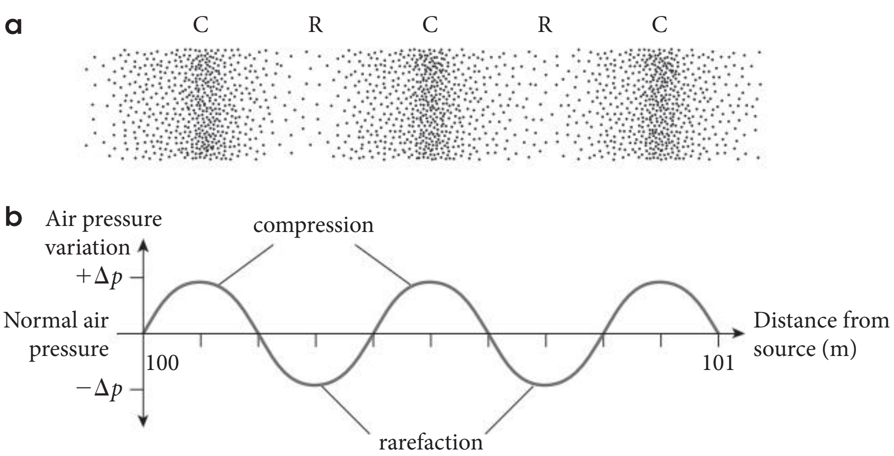


Figure 10.12 a) When sound travels through a medium, the particles form a series of compressions and rarefactions. b) Graph shows pressure variation with distance.

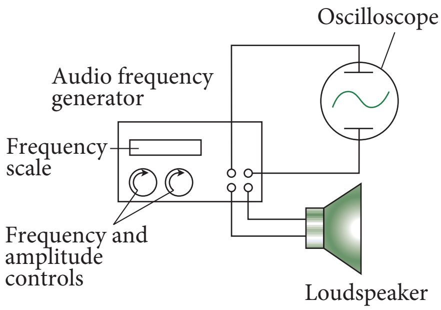


Figure 10.13 ▲ Apparatus to investigate the waveform of sound waves

Representing sound waves

An oscilloscope shows how electrical signals change with time. A microphone converts sound into electrical signals. These can be sent to an oscilloscope. The vertical axis on the oscilloscope represents the pressure variations in the sound waves. The horizontal axis represents the time over which these changes take place.

We can use an audio frequency generator, a loudspeaker and an oscilloscope to investigate the waveforms of sound waves at the same time as they are heard.

A sound of known frequency is produced by the audio generator. This can be heard through the loudspeaker and viewed on the oscilloscope. The audio generator vibrates the loudspeaker. Simultaneously, the oscilloscope displays an *apparently* transverse wave. The y axis represents particle movement at a point, and the x axis represents time. It displays a graph of amplitude versus time. It is not a snapshot of the wave at a particular time.

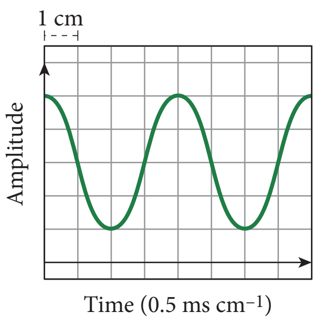


Figure 10.14 ▲ Sound wave on an oscilloscope

The **time-base scale** (x axis) enables the period of the wave to be determined (Figure 10.14). The **pitch** of the note increases with frequency.

For the waveform in Figure 10.14, one complete wave on the oscilloscope is completed in 4 cm. If the time-base scale of the oscilloscope is set on 0.5 ms cm^{-1}, then each centimetre on the scale on the screen is equivalent to half a millisecond. The period (T) of the wave is 4 cm $\times$ 0.5 $\times$ 10^{-3} s cm^{-1}, and therefore equal to 2.0 $\times$ 10^{-3} s. As $f = \frac{1}{T}$, frequency $f = \frac{1}{2.0 \times 10^{-3}}$ s. The frequency is 500 Hz.

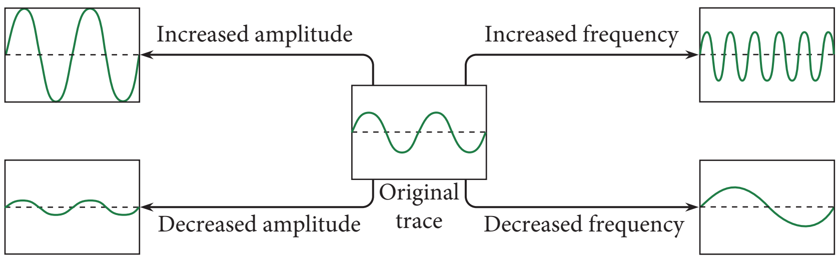


Figure 10.15 ▶ Oscilloscope traces show the qualitative relationship between sound wave amplitude and frequency.

A tuning fork is used as a standard of pitch to tune other musical instruments. It comprises a two-pronged fork formed from a U-shaped bar of elastic metal. When the fork is struck, the prongs resonate at a specific constant frequency.

EXPERIMENT 10.1

MEASURING THE FREQUENCIES OF TUNING FORKS

Microphones convert sound pressure changes to voltage changes displayed on the oscilloscope screen.

Aim

To measure the frequencies of tuning forks

What you will need

- an oscilloscope
- a good quality microphone
- 3 tuning forks with different frequencies (preferably attached to a sounding box)

Shutterstock.com/MilanB

Figure 10.16 ▲ Tuning fork mounted on sound box

▶

Prepare a risk assessment box like the one shown below and have it approved by your teacher.

What are the risks in doing this experiment?	How can you manage these risks to stay safe?

Procedure

Your teacher will demonstrate how to set up and use the oscilloscope.

See page 320 for details of what the screen displays.

1 Strike the tuning fork and hold it close to the microphone. If you are using a tuning fork with a sounding box, hold the microphone at the opening of the box.
2 Adjust the time-base switch until a good, well-spaced, fairly stable sine wave is obtained.
3 Practice using the features of the oscilloscope, such as freezing the signal to find the period, wavelength and speed of the wave.
4 Measure the distance on the screen from one crest to the next and convert this to the period.
5 Record your results in a data chart similar to the one below.
6 Repeat the procedure for the other tuning forks.
7 Calculate the frequency for each tuning fork.

Results

Tuning fork	Time base setting (ms cm^{-1})	Crest to crest distance (cm)	Period (ms)	Frequency (Hz)

Analysis of results

1 What do you notice about the frequencies of the different tuning forks?
2 Why do you think the amplitude of the signals displayed on the screen died away quickly?
3 Did you notice any relationship between the frequency and the amplitude on the screen? If so, what was it?

Discussion

1 What is the advantage of mounting the tuning forks on the sounding boxes?
2 Tuning forks can be constructed to have any fundamental frequency. Why is it that most of the tuning forks made resonate at a set of specific frequencies?

Conclusion

Write a conclusion to answer your aim.

QUESTION SET 10.2

Remembering

1 What is the difference between a compression and a rarefaction?

Understanding

2 Sound waves are classified as longitudinal waves. Why is this?
3 When a trumpet is played, the sound waves that are produced spread out in all directions from the source at a speed of 340 m s^{-1}. Which one of the following best describes the motion of the air particles at a distance of 12.0 m from the trumpet? For simplicity assume the air particles were stationary before the trumpet was sounded.
   A The air particles transferring the sound wave are vibrating parallel to the direction of motion of the sound waves.
   B The air particles are vibrating at right angles to the direction of motion of the sound waves.

C The air particles are moving away from the trumpet with a speed of 340 m s^{-1}.

D The air particles are moving away from the trumpet with a speed greater than 340 m s^{-1}.

E The air particles are moving away from the trumpet with a speed less than 340 m s^{-1}.

4 A sound wave of wavelength 34 cm travels at 340 m s^{-1} in air. What is the frequency of the sound?

5 A sound wave of frequency 210 Hz travels at 340 m s^{-1} in air. What is the wavelength of the sound?

Applying

6 Figure 10.17 is the displacement–distance graph for a periodic wave.

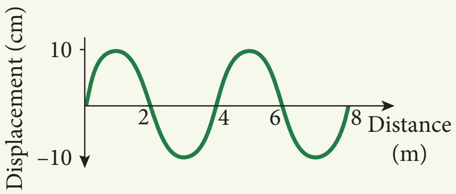


Figure 10.17 ◀

a What is the wavelength of the periodic wave?

b What is the amplitude of the wave?

7 Figure 10.18 shows the trace produced on an oscilloscope when a microphone picks up a pure note sung by an opera singer. Copy the trace and use it as the basis for drawing the oscilloscope trace that would result from:

a a louder sound of the same frequency.

b a note of the same amplitude but with double the frequency.

c a sound from an audio frequency oscillator having the same pitch as the singer.

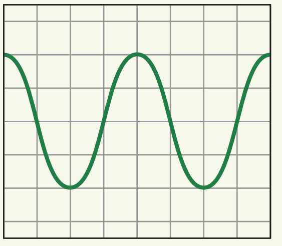


Figure 10.18 ▲

Analysing

8 A sound wave consists of regions of air pressure that alternate from slightly higher to slightly lower than normal. At one particular instant, the variation in the air pressure of a sound wave a long way from its source is as shown in Figure 10.19.
The wave is travelling at a speed of 340 m s^{-1}.

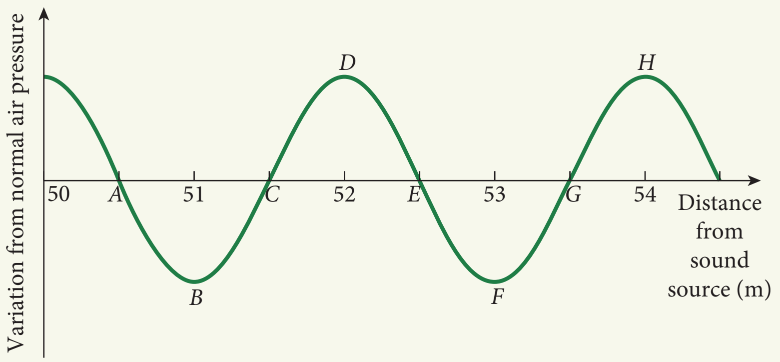


Figure 10.19 ◀

a What is the frequency of the sound wave shown in Figure 10.19?

b Which of the points A–H in Figure 10.19 correspond to point(s) where the sound wave is causing zero displacement of the air particles?

c Copy the graph in Figure 10.19 and show the pressure variations of the wave:

i a quarter of a period later.

ii half a period later.

Reflecting

9 In Question 3 it was stated, 'For simplicity assume the air particles were stationary before the trumpet was sounded'. What does this tell you about the limitations of the model used to describe sound propagation?

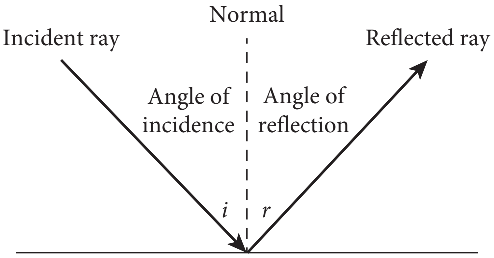
Wave behaviour

Sound waves can interact with surfaces, edges and interfaces between different materials. These interactions include reflection, refraction and diffraction.

Reflection of sound waves

Echoes are evidence of the reflection of sound. Careful measurements show that all waves obey the **law of reflection**.

If we draw a normal to the surface (that is, a perpendicular line) such that the incident and reflected wave are in the plane of the normal, then:



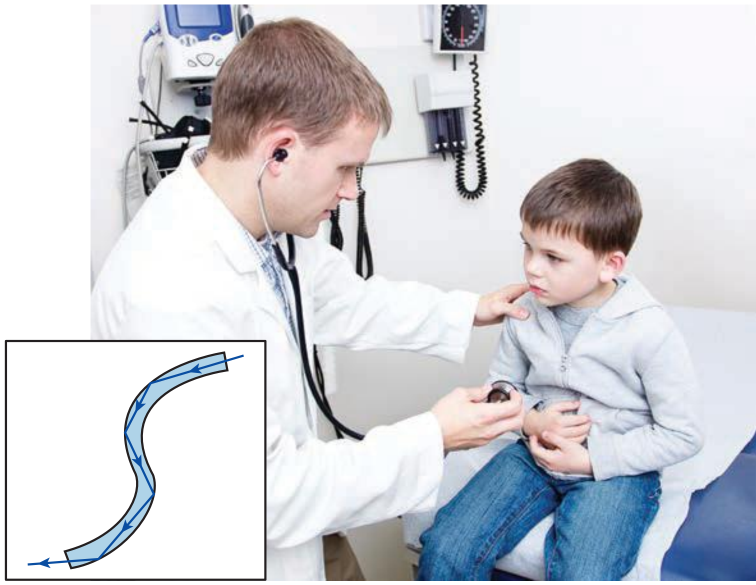
▲ Figure 10.20 Law of reflection for waves

> The angle of incidence, *i*, is equal to the angle of reflection, *r*.

Reflection of sound is used in stethoscopes and sonar depth sounders. In stethoscopes, sound waves reflect back and forth along the inner walls of a tube. This is known as **total internal reflection**.

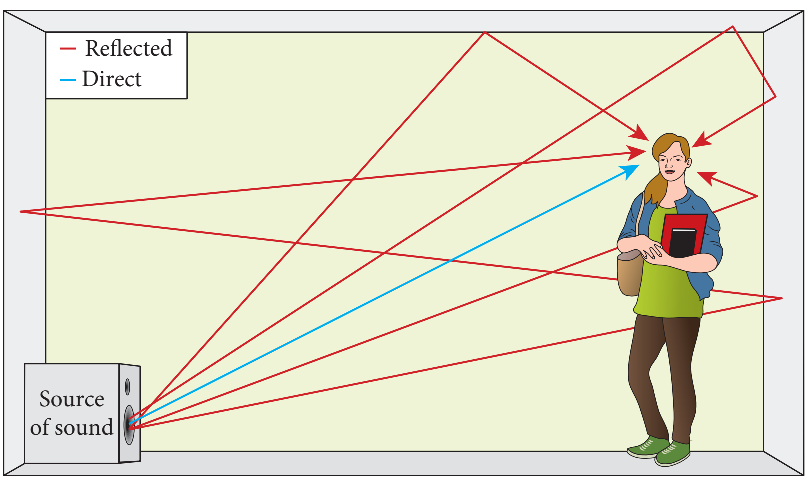
Reverberation

When sound is produced in an enclosed space, a large number of echoes or **reverberations** build up and then slowly **decay** as the sound is absorbed by the walls and air. When the sound source stops, the reflections continue, decreasing in amplitude, until they can no longer be heard. The longer the time of decay the greater the reverberation.



iStockphoto/Yobro10

▲ Figure 10.21 Stethoscopes allow sound to travel to the ear by total internal reflection, as shown in the insert.



◀ Figure 10.22 Multiple reflections and direct sound cause reverberation.

Echoes

If you clap your hands or make a loud sharp sound at a distance from a good sound-reflecting surface, you hear an echo. The human ear can distinguish sounds that are about one fifteenth of a second apart. In air, the speed of sound is about 340 m s^{-1}, so the minimum distance the sound must travel for you to hear the echo is:

$$v = \frac{\text{distance}}{\text{time}}$$

$$\text{distance} = v \times \text{time}$$

$$= 340 \text{ m s}^{-1} \times 6.67 \times 10^{-2} \text{ s}$$

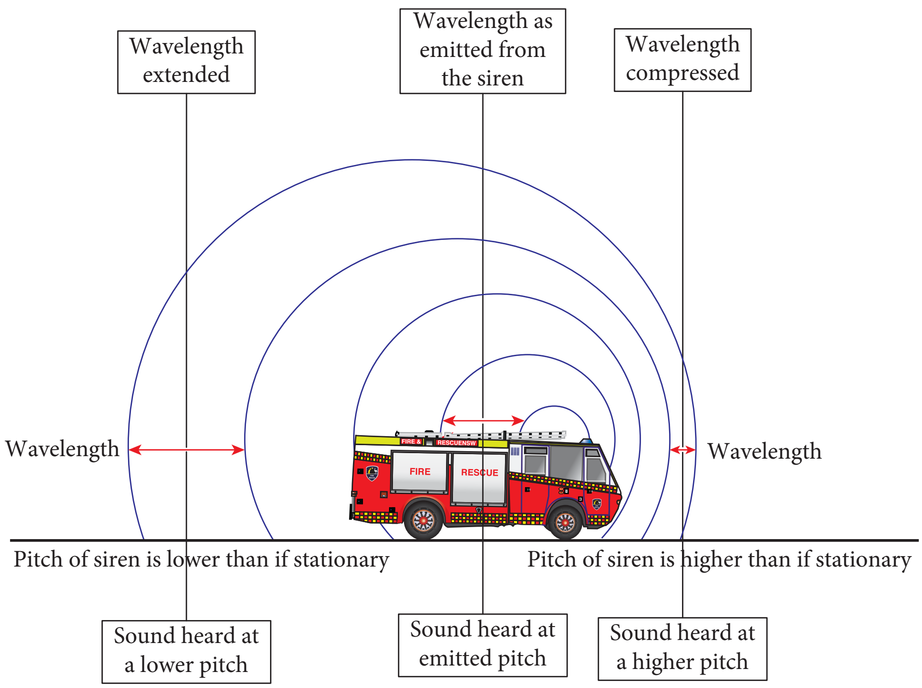
$$22.7 \text{ m}$$

The minimum distance between the clap (source) and the echo surface is $\frac{22.7 \text{ m}}{2} = 11.4$ m.

Doppler effect

When a source of waves, for example a siren, is approaching a receiver, the pitch heard is greater than that emitted. This is because each wave is emitted a little closer to the observer than the previous one. The reverse is heard when the source moves away from the receiver. This is known as the **Doppler effect**. It is the relative motion of source and receiver that changes the wavelength of the sound received relative to the sound emitted. The number of sound waves received per second, the frequency, depends on the relative speed of source and receiver.

Figure 10.23 ▶ The Doppler effect for a fire engine siren



Speed of sound

Sound, like all mechanical waves, must travel through a medium – it cannot travel through a vacuum. Typically there are two essential properties of a medium that affect the speed of a wave: its **density** and its elastic properties.

The phase (state) of matter has a very large impact upon the elastic properties of the medium. In general, solids have the strongest interactions between particles, followed by liquids and then gases. The stronger the interaction, the more rapidly the particles respond to a collision and the greater the speed of the wave. For a particular material, longitudinal sound waves travel in the solid state faster than they do in the liquid state, which is faster than in the gaseous phase. The mass density factor favours the gas state, but the elastic response is the more influential.

Refer back to the information on the kinetic particle model of matter on page 6.

Wave speed depends on the material and its state – solid, liquid or gas. Speed differences between materials in the same state are affected most by density. The greater the density of the material, the more sluggish will be the interactions between its neighbouring particles.

This results in the wave travelling more slowly. For example, a sound wave will travel nearly three times faster in helium than it will in air.

The speed of a sound wave in air depends on temperature and humidity. Temperature has the most influence because it affects density.

Warm dry air is less dense than cold air. Humidity is also important. Moist air is less dense than dry air because water vapour is less dense than both oxygen and nitrogen.

The speed of the wave depends upon the properties of the medium through which the wave is travelling, not the frequency or wavelength.

Table 10.2 The speed of sound in different media at 25°C and 1 atmosphere pressure

State	Substance	Speed $m\,s^{-1}$
Solids	Aluminium	6420
	Nickel	6040
	Steel	5960
	Iron	5950
	Brass	4700
	Glass (flint)	3980
Liquids	Water (sea)	1531
	Water (distilled)	1498
	Ethanol	1207
	Methanol	1103
Gases	Hydrogen	1284
	Helium	965
	Air	346
	Oxygen	316
	Sulfur dioxide	213

The speed of sound in a wire also depends on how much it is being stretched – the tension in the wire. The greater the tension, the greater the speed of the sound wave.

WOW

The sonic boom

When a plane flies through the air it creates a series of pressure waves in front of it and behind it. These waves travel at the speed of sound. As the speed of the plane increases, the waves are forced together, merging into a single shockwave at the speed of sound. At 'supersonic' speeds, the plane begins pushing the air like a plough. This high-pressure shockwave sounds like a boom, and is continuous along what is called the boom carpet for the entire supersonic flight. Low-pressure rarefactions cause rapid condensation, hence clouds, around the plane.

Alamy/US Navy Photo

◀ Figure 10.24
A jet fighter plane just breaking the sound barrier and forming a typical condensation cloud as the shockwave forms

SONIC BOOM

View a jet plane as it produces a sonic boom.

Scientific literacy: Bats hunt using echolocation

When hunting their prey, some animals such as dolphins and bats use echolocation.

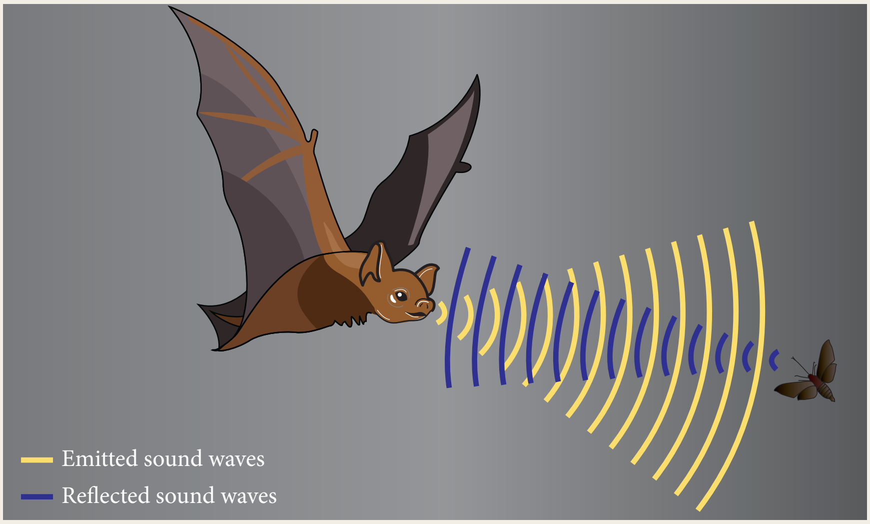


Figure 10.25 ▲
Bats use echolocation to hunt their prey.

Most bats use echolocation to navigate in the dark and to locate and catch food (Figure 10.25). Bats send out sound waves using their mouth or nose. When the sound hits an object, an echo comes back. The bats can identify the object by the sound of the echo. They can tell the size, shape and texture of a tiny insect from its echo.

Alamy/Arterra Picture Library

Figure 10.26 ▲
Bats such as the Brown long-eared bat, *Plecotus auritus*, have large ears that can capture the faintest sounds.

The echolocation calls emitted are usually ultrasonic, ranging in frequency from 20 kHz to 200 kHz. These are above the human hearing limit of 20 kHz or less. Even so, we can hear echolocation clicks produced by bat sonar. These noises resemble the sounds made by hitting two pebbles together. In general, echolocation calls are characterised by their frequency, their loudness in decibels (dB), and their duration in milliseconds (ms).

The bats produce a complicated sequence of echolocation calls that combine both constant frequencies and varying frequencies. Although low-frequency sound travels further than high-frequency sound, the higher frequency calls give the bats more-detailed information about the size, range, position, speed and direction of their prey in flight. Thus, high-frequency sounds are used more often.

Bats use the Doppler shift of the reflected signal to determine if the prey is flying towards or away from them. The Doppler shift occurs when a pulse of sound from a source, such as a bat, strikes a target that is moving relative to the source. If the target is moving away from the source, the frequency is shifted down to a lower frequency. If the target is moving towards the source, the frequency is shifted up to a higher frequency. By analysing the shift, the bat is able to determine the velocity of the target and the direction of its movement.

Bats emit calls as soft as 50 dB and as loud as 120 dB, which is louder than the alarm in a smoke detector 10 centimetres from your ear and intense enough to damage human hearing. To protect themselves, bats' ears have adaptations that reduce the intensity of the sounds received, thus protecting the ear from damage.

The ears and brain cells in bats are especially tuned to the frequencies of the sounds they emit. A concentration of receptor cells in their inner ear makes bats extremely sensitive to frequency changes. The large variations in the sizes and shapes, and the number of folds in their ears aid in the reception of the returned sound. They do this by funnelling echoes into their inner ear.

BATS HUNTING THEIR PREY

View this video of bats using echolocation to hunt their prey.

Questions

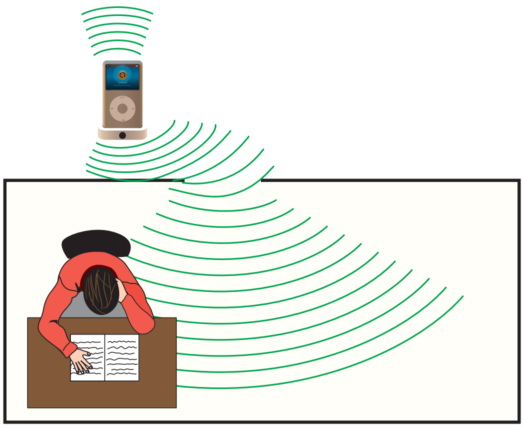
1 a Define 'echolocation'.
  b What do bats use echolocation for?
2 A bat that emits tracking sounds with a loudness of 50 dB needs larger ears than a bat that emits signals at 100 dB. Why do you think this is so?
3 Bats have large highly modified ears. Give two reasons why they need the modifications.
4 A bat sends out a 20 kHz signal while hunting. How long does it take for a signal it has sent to a moth 20 cm away to return to the bat?
5 Bats can tell the direction and the speed of their prey. Explain how they can do this.
6 What extra understanding of the process of echolocation did you gain by watching the 'Bats hunting their prey' video?

Diffraction of sound waves

For humans, the audible spectrum of sound ranges from about 20 Hz up to 20 kHz. This corresponds to a wavelength range from 17 m down to about 2 cm in air.

If you are outside and someone calls out to you from the other side of an oval or park, you turn in their direction. This is because you have binaural hearing. 'Binaural' literally means 'for two ears'.

If the sound arrives at an angle to the axis of your ears, then the sound will arrive at each ear at slightly different times. The sound waves bend around the head and reach the ear facing away from the sound source a fraction of a second after they reach the ear facing towards the sound source. The brain processes the information and gives you an approximate location of the source of the sound.



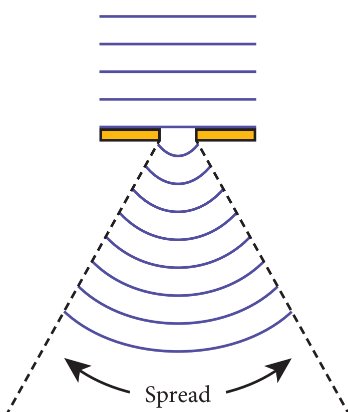
▲ **Figure 10.27** Simplified diagram of sound waves diffracted through an open door

If the sound comes from directly in front, behind or above you, the sound arrives at both your ears at the same time. This makes it difficult for you to get an accurate location of the source of the sound. This is why you automatically tilt or turn your head when trying to locate sound sources in those positions. The sound wave has travelled to you in a straight line. So when you turn, you can use the off-axis differences to locate the source of the sound.

If you are sitting in a room reading, you can still be aware of a radio or TV in another room or a noisy truck in the street outside. In these situations, the sound will have both reflected and bent around corners and/or obstacles to reach your ears. This phenomenon of waves bending around corners or obstacles is known as **diffraction**. The diffraction pattern spreads into the whole region. The central maximum, containing the bulk of the energy, spreads more for longer waves (Figure 10.28).

When you are listening to a loudspeaker, the higher pitched sounds (shorter wavelengths) will be best heard in front of the speaker. This is because higher pitched sounds have shorter wavelengths and are diffracted less than the longer wavelengths. The lower pitched sounds (longer wavelengths) will be heard in front and to the sides of the speaker.

The shorter wavelengths will be more directional; that is, they will pass through openings with less bending or spreading than the longer wavelengths. The amount of diffraction of sound depends on the ratio of λ to w, where w is the slit width. The smaller the opening for a given wavelength, the more diffraction there will be and the greater the angle of spread. Also the longer the wavelength for a given opening, the more diffraction there will be and the greater the angle of spread.



▲ **Figure 10.28** Waves spread out after passing through a gap. This is called diffraction.

Loud music on a car system always sounds much the same outside the car no matter what is being played because only the low frequency sound waves are diffracted effectively through a window. When sound waves are diffracted there is no change to the speed, wavelength or frequency.

Diffraction is the spreading of waves into a space beyond a gap or obstacle. The amount of spread is greatest when the wavelength is greater than the gap width or obstacle.

When sound waves are diffracted there is no change to the speed, wavelength or frequency of the wave.

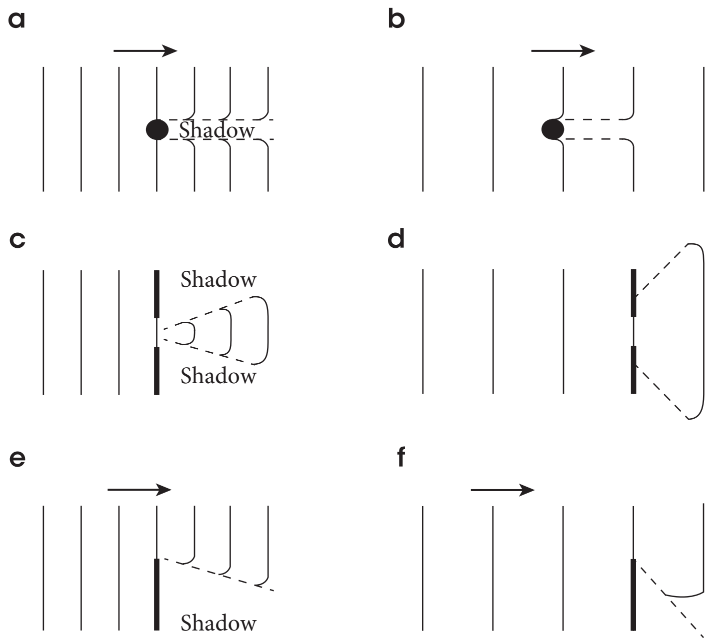


Figure 10.29 Diffraction of water waves: a) short wavelength around an object, b) long wavelength around an object, c) short wavelength through a gap, d) long wavelength through the same gap, e) short wavelength around the edge of a barrier and f) long wavelength around the edge of the same barrier.

You read earlier that bats use **ultrasound** to locate their prey. The frequencies used range from 20 kHz up to 200 kHz. Sound waves can only reflect from an object if its wavelength is similar to or less than the size of the object. If a sound wave has a long wavelength, it is completely diffracted around the object and there is no reflection. This puts a limit on the smallest size prey a bat can hunt.

WORKED EXAMPLE 10.2

If the echolocation frequency of a bat is 100 000 Hz, what is the smallest prey it can locate? (Assume the speed of sound in air is 340 m s^{-1}.) (4 marks)

Answer	Logic	
$v = f\lambda$	Choose the relationship.	1 mark
$\lambda = \frac{v}{f}$	Rearrange the formula.	1 mark
$\lambda = \frac{340 \text{ m s}^{-1}}{100000 \text{ s}^{-1}}$	Substitute the known values.	1 mark
$\lambda = 0.0034$ m	Calculate the answer.	1 mark
That is 3.4 mm, a mosquito-sized meal.		

Try this yourself

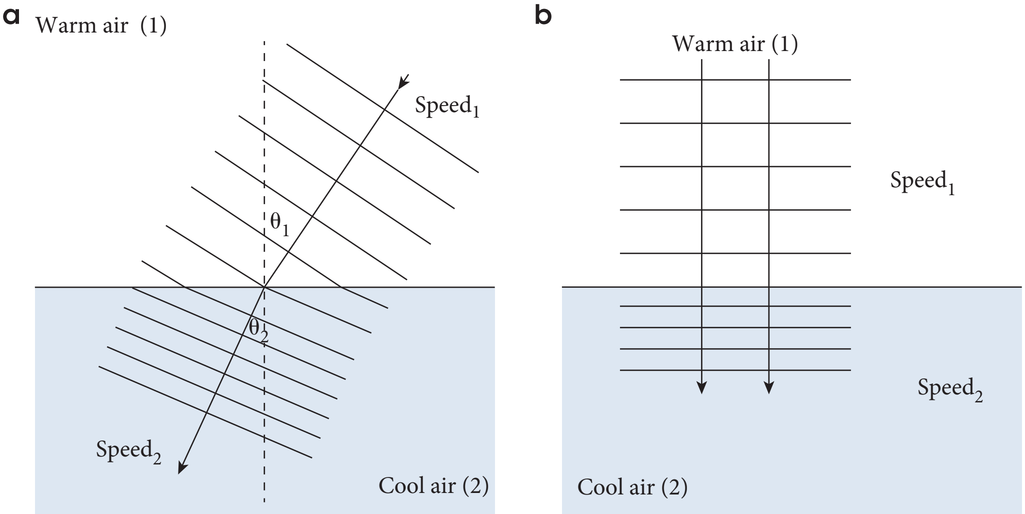
Dolphins also use echolocation to catch their prey, using frequencies of about 20 000 Hz. Given that the speed of sound in salt water is 1531 m s^{-1}, what is the smallest fish a dolphin could locate? (4 marks)

Refraction of sound waves

Refraction of waves involves a change in the direction of the waves as they pass from one medium to another. Refraction, or bending of the path of the waves, is accompanied by changes in the speed and wavelength of the waves. If the medium (or its properties) are changed, the speed of the wave is changed. The frequency does not change, therefore the wavelength must change if speed changes Thus, when waves pass from one medium to another, refraction can occur. This is due to the new medium having a different elastic property and/or mass density that affect the rate of transmission of the wave energy.

Temperature differences in the same medium will also cause refraction because the density of the air changes with temperature. Sound travels faster in warm air.

If a wave meets the interface at right angles it will not change direction, but its speed and wavelength will change. If it meets the interface at any other angle, its direction will change, as shown in Figure 10.30.

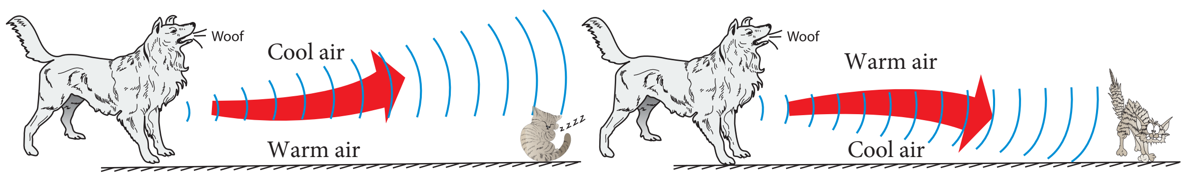


◀ Figure 10.30
a) A sound wave is refracted (changes direction) when it meets the boundary between two layers at an angle other than a right angle. b) At right angles there is no change in direction. Speed and wavelength both change.

Refraction: Waves change direction when they strike a surface at an angle other than 90°. The speed and wavelength of the wave also change. The frequency remains constant.

Refraction of sound waves is most evident in situations in which the sound wave passes through a medium with gradually varying properties. This commonly occurs in the atmosphere where there is a gradual change in temperature or where there is a temperature inversion. During the day the air close to the ground is warmed by the hot ground. The air higher up is cooler, creating a **temperature gradient** (Figure 10.31).

Sound travels faster in warmer air, so the edge of the wavefront close to the ground travels faster than the edge of the wavefront in the cool air. This causes a gentle refraction (bending) of the wavefronts away from the ground. The reverse is true at night, when the ground is cooler than the air above it. The implication is that on nights when the ground is cooler than the air above it you can clearly hear sounds coming from a distant source.



▲ Figure 10.31
Refraction of sound in reverse temperature gradients

QUESTION SET 10.3

Remembering

1 Define the following terms and explain how they differ from each other.
   a Reflection
   b Refraction
   c Diffraction

2 Reverberation and echoes are both reflection phenomena. How are they different?

Understanding

3 At room temperature the speed of sound in helium is 1007 m s^{-1} and 326 m s^{-1} in oxygen.
   a Helium and oxygen are both gases, so what property of helium allows sound to travel faster in helium than in oxygen?
   b What could you do to oxygen gas to make sound travel faster in it? Explain your reason.

4 Why must the frequency of a wave remain the same when it travels from one medium into an adjacent medium?

5 Why is the sound of an approaching siren heard at a higher pitch?

6 Waves diffract around the edge of a barrier or obstacle. What effect does decreasing the wavelength have on the amount of diffraction around that same object?

Applying

7 If you were hiding behind a large tree trunk in an open field and a high-pitched whistle was blown on the other side of the tree, would you hear it? Explain your answer.

Analysing

8 An echo returns to a scientist in Antarctica in 0.21 s. How far away is the echo surface? The speed of sound in cold dry air is 343 m s^{-1}.

9 A noisy speedway 2 km from your house holds race meetings every Friday night. On some nights the sound is very loud and on other nights you can barely hear the noise. Explain how refraction could be the phenomenon that causes these outcomes.

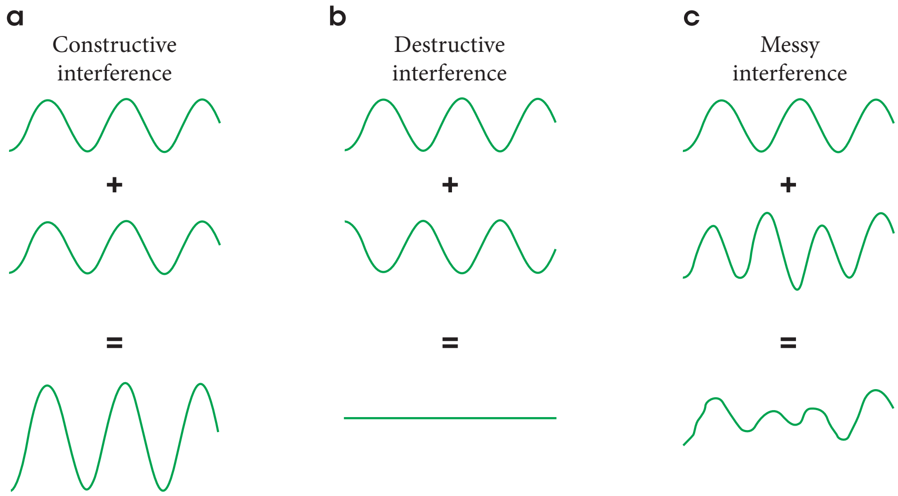
Reflecting

10 What are some situations in which you have experienced the phenomenon of wave diffraction?

The law of superposition

Waves of the same nature (either longitudinal or transverse) and in the same medium can pass through each other. They add algebraically as they interact, then resume their original shape. This is **superposition**. Usually superposition results in a messy combination. At any period, **constructive interference** (larger amplitude) or **destructive interference** (zero amplitude) may occur, depending on the two waves.

Superposition
When two or more waves of the same nature travel past a point in a medium, the medium will undergo a resultant displacement at that point. The resultant displacement of the medium at that point is the sum of the individual particle displacements due to the waves at that point.

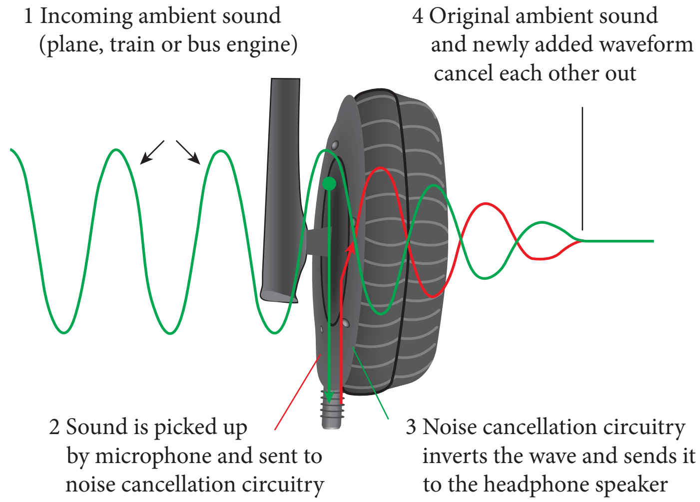


◀ Figure 10.32
a) Constructive, b) destructive and c) messy wave interactions are created by the superposition of the two waves.

WOW

Noise-cancelling earphones

A clever use of sound interference is to cancel noise. Headphones designed to cancel noise with destructive interference create a sound wave exactly opposite to the incoming sound. They use a microphone to receive the sound waves and with real-time, fast electronics produce a signal into sound waves that is the exact reverse of the incoming signal. This new signal interferes with the original signal, as shown in Figure 10.33. 30 dB reductions are common.



▲ Figure 10.33
How noise cancelling headphones work

THE LAW OF SUPERPOSITION

This animation demonstrates the law of superposition.

Stationary waves

A **standing wave** or **stationary wave**, is a wave that does not appear to be moving. If you shake waves onto a string that is fixed at the other end, the forwards and reflected waves will interfere. Usually, the effect is messy. But if you get it just right, a fixed pattern of maximum and minimum displacement will appear. This is an example of **resonance**.

See page 333, Resonance

Standing waves can be created in stretched strings, springs and in air columns in pipes. This is used extensively in musical instruments to make resonant sounds.

A standing wave such as that shown in Figure 10.34 is created when two waves of the same frequency and amplitude but travelling in opposite directions exist in a medium.

The **nodes** are points where destructive interference always occurs. At these points, at any moment in time, the amplitudes of the two waves are always the same magnitude but in

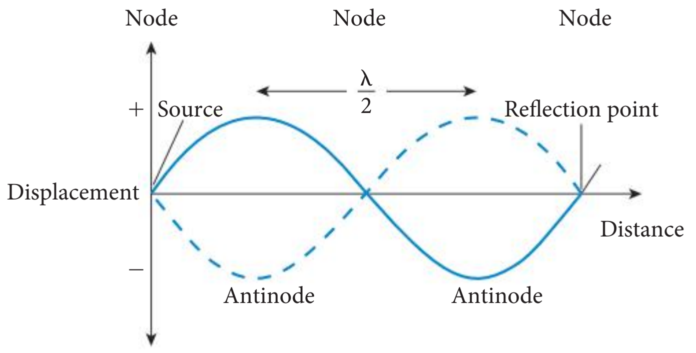


Figure 10.34 ▲ A stationary wave in a stretched string fixed at both ends showing the nodes and antinodes. The solid line represents the string's displacement at an instant in time and the dotted line the string's displacement half a period $\left(\frac{T}{2}\right)$ later.

opposite directions. Hence when a crest of one wave passes through this point, a trough of the same size is also passing through. When the displacement due to one wave is half the amplitude, the displacement due to the other wave is also half the amplitude, but in the opposite direction, and so on. Whatever the displacement due to one wave, the displacement due to the other is the same size but the opposite direction. Superposition means that these two displacements add to give zero total displacement at any time at a node, hence a particle at a node does not move.

At the **antinodes**, the particles move up and down constantly and reach the maximum displacement possible. The maximum displacement occurs when two crests (or two troughs) meet at this point to give a displacement twice that of the amplitude of the individual waves. So at an antinode the particles oscillate up and down between displacements of $-2A$ and $2A$. The frequency with which they move up and down is the same as the wave frequency.

In between nodes and antinodes the particles oscillate up and down with the same frequency, but with smaller amplitudes, to produce the pattern shown in Figure 10.34.

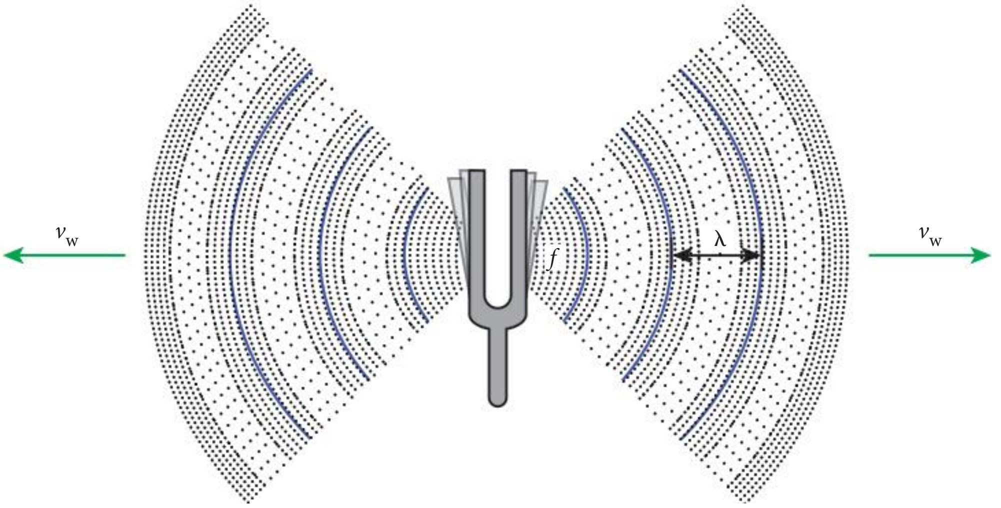
Standing wave patterns are always characterised by an alternating pattern of nodes and antinodes. The distance between nodes or between antinodes along the string or spring is half a wavelength, $\frac{\lambda}{2}$.

Free vibrations

Free (natural) vibrations occur when an object is displaced from its equilibrium position and then left to vibrate by itself.

When a tuning fork is struck, the prongs vibrate about their mean position. Elastic restoring forces strongly pull the prongs back and forth. The tuning fork vibrates at its natural frequency. This phenomenon can be observed in guitar strings, organ pipes, wind instruments, drums, pendulums and masses hanging on the end of springs (think bungee jumping!) – all have natural frequencies.

Figure 10.35 ▶ Sound waves produced from the freely vibrating tuning fork with a frequency f, speed v_w and wavelength λ



The frequency (and period) of vibration is determined by the properties of the vibrating object. For example, a plucked guitar string vibrates at different natural frequencies depending on its length, mass per unit length and the tension in the string.

The only energy driving a free vibration is the initial energy; thus, in time these vibrations die away because of friction – energy transfers to the surroundings.

Forced vibrations

A **forced vibration** occurs when one vibrating object makes another object vibrate. If a vibrating tuning fork is struck on a rubber stopper, it emits a low-intensity sound that can be heard only with difficulty. However, if the same vibrating tuning fork is held with its shaft on a wooden bench or tabletop, the sound is heard throughout a classroom. Why is this?

The sound is louder when the fork is in contact with the bench, because the fork causes the bench to vibrate with the same frequency. The benchtop has a larger vibrating area than the tuning fork. Consequently, these forced vibrations disturb a greater volume of air and produce a louder sound.

▲ **Figure 10.36**
Blowing across an open bottle

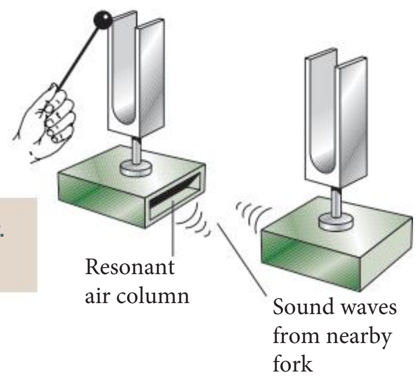
Resonance

If you blow across the mouth of a bottle as shown in Figure 10.36, the air in the bottle is made to vibrate and you will hear a note. The frequency of this note is determined by the dimensions of the bottle. The sound results from the free vibrations of the air in the bottle.

Purse your lips and blow through them. You will hear the sound of the rushing air. This sound is made up of waves of very many different frequencies. This sort of sound is called 'white noise' in analogy to 'white light', which is composed of many frequencies of light.

When you blow air across the top of a bottle like this, you are providing waves of many different frequencies to the air column inside the bottle. Most of these waves transfer energy very inefficiently to the air column. But waves of one particular frequency, the natural or resonance frequency, transfer energy very efficiently and set up a standing wave in the bottle. The frequency of this standing wave is the frequency of the note you hear.

- **Resonance will only occur when the driving frequency matches the natural frequency.**
- **The amplitude of vibration of the resonating object will increase dramatically.**



▲ **Figure 10.37**
Energy supplied to one tuning fork forces the other to resonate.

This is how clarinets and other woodwind instruments work. You make a particular mouth shape, called the *embouchure*, and blow air through your lips over a reed. The reed vibrates with the many frequencies of the waves produced by your blowing. The wave of just the right frequency then creates a standing wave in the pipe of the instrument. In effect, the body of the instrument selects and amplifies one particular frequency from the many you provide it with. By covering different holes, you determine which frequency (which note) is selected. Brass instruments, for example trumpets, work the same way, but your vibrating lips do the job of the reed. How the particular note (frequency) is selected is discussed in more detail later.

Now sound a tuning fork over the mouth of the bottle. If the frequency of the tuning fork (the driving frequency) differs from the natural frequency of the air column, only a feeble sound is heard. The frequency of this feeble sound is the same as the frequency of the fork and is due to the forced vibrations in the air column in the bottle.

When a tuning fork vibrating at the same natural frequency as the air column in the bottle is held over the mouth of the bottle, the sound intensity is increased considerably.

When the frequency of the forced vibration coincides with the natural frequency of the system, energy is transferred with maximum efficiency. For these examples, a standing wave is produced. This phenomenon is called resonance. The energy of the vibration across the top of the bottle is transferred very efficiently to the vibrating air column inside the bottle.

When an object is resonating, energy is being transferred very efficiently from the driving oscillator to the receiving oscillator.

Figure 10.37 shows forced vibrations when two tuning forks are mounted on sounding boxes. The length of the sounding box should be one quarter of the wavelength of the sound wave produced when the tuning fork vibrates.

THE DESTRUCTION OF THE TACOMA NARROWS BRIDGE

This spectacular footage shows the Tacoma Narrows Bridge collapsing due to resonance created by a steady wind that caused aero-elastic flutter. Flutter is a more complex phenomenon than pure resonance.

QUESTION SET 10.4

Remembering

1 What is the difference between a free and a forced vibration?

2 Give two examples of:
  a free vibrations. b forced vibrations.

3 a Explain how blowing across the top of a bottle can produce a loud, clear note.
  b What is this phenomenon called?

Understanding

4 What conditions are required for a standing wave to form?

5 The vibrating air column in the bottle in Figure 10.36 is a standing wave. Why is it called a standing wave?

Applying

6 How could you use one tuning fork to force another tuning fork to vibrate at its natural frequency?

Analysing

7 When a wave is reflected off a beach and meets an incoming wave, a larger wave is formed as they pass through each other. Sketch diagrams to show the two waves:
  a before they interact. b in the middle of interacting. c after interacting.

Modes of vibrations

Figure 10.38 shows possible vibration modes, or harmonics, for stationary waves on a string or wire fixed at both ends. There is a node at each fixed end for all modes of vibrations.

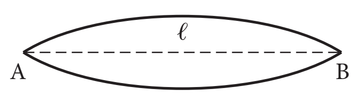
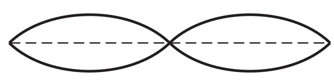
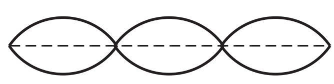
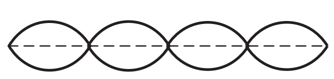
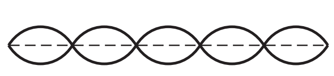
Vibration mode	Wave pattern	f & λ
Fundamental mode of vibration 1st harmonic		$\lambda_1 = 2\ell$ $f_1 = \frac{v}{2\ell}$
1st overtone 2nd harmonic		$\lambda_2 = \ell$ $f_2 = 2f_1$
2nd overtone 3rd harmonic		$\lambda_3 = \frac{2}{3}\ell$ $f_3 = 3f_1$
3rd overtone 4th harmonic		$\lambda_4 = \frac{1}{2}\ell$ $f_4 = 4f_1$
4th overtone 5th harmonic		$\lambda_5 = \frac{2}{5}\ell$ $f_5 = 5f_1$

Figure 10.38 ▲ The first five harmonics (four overtones) of a string fixed at both ends.

The fundamental mode of vibration is also referred to as the **first harmonic**. The other modes of vibration are called the second harmonic, third harmonic and so on. Wires and strings of musical instruments can be made to vibrate at frequencies other than their fundamental frequency.

Harmonics other than the first harmonic are also called overtones. Overtones are notes or tones of higher frequency than the fundamental or natural frequency (and are of smaller amplitude). Thus, the second harmonic is the first overtone; the third harmonic is the second overtone and so on.

The frequency of each harmonic is its harmonic number times the fundamental frequency. If the fundamental frequency of a stretched string is 40 Hz, the fourth harmonic has a frequency of $4 \times 40 = 160$ Hz.

The fundamental mode of vibration (the first harmonic) is generated when the stretched string is plucked in the middle. If you look at the fundamental vibration mode in Figure 10.38, the pattern represents half a wave, so wavelength λ_1 is twice the length of the string: $\ell = \frac{\lambda_1}{2}$.

The second harmonic (first overtone) is generated when the stretched string is plucked a quarter of the way along the string. The second harmonic mode in Figure 10.38 shows that the pattern represents a complete wave, so the wavelength is the length of the string: $\ell = \lambda$.

The third harmonic is generated when the stretched string is plucked a sixth of the way along the string. The third harmonic mode in Figure 10.38, the pattern represents one and a half complete waves so wavelength is two-thirds the length of the string: $\ell = \frac{3\lambda_1}{2}$.

For strings attached at both ends (and as you will see later for open pipes) to be resonating, the length of the string is related to the resonating wavelength by the relationship:

$$\ell = n\frac{\lambda_n}{2}$$

where ℓ is the length of the string and λ is the wavelength of the nth harmonic (n = 1, 2, 3 …).

Putting this together with $v = f\lambda$ the following relationships become apparent.

- The fundamental or first mode has frequency $f_1 = \frac{v}{\lambda_1} = \frac{v}{2\ell}$
- The second harmonic has frequency $f_2 = \frac{v}{\lambda_2} = \frac{2v}{2\ell} = 2f_1$
- To generalise, the nth harmonic has frequency $f_n = \frac{v}{\lambda_n} = \frac{nv}{2\ell} = nf_1$

WORKED EXAMPLE 10.3

The fundamental frequency of a string 2.4 m long, fixed at both ends, is 20 Hz.

a What are the frequencies of the first three overtones? (3 marks)

b Is it possible to produce stationary waves of frequency 50 Hz in this string? (2 marks)

c What is the speed of the waves in the string? (3 marks)

d What is the wavelength of the first harmonic? (1 mark)

Answers	Logic	
a The frequency of each harmonic is its harmonic number times the fundamental frequency.	The overtone number is one less than the harmonic number.	
i 2 × 20 Hz = 40 Hz		1 mark
ii 3 × 20 Hz = 60 Hz		1 mark
iii 4 × 20 Hz = 80 Hz		1 mark
b As 50 Hz is not the frequency of one of the harmonics of this string (a whole number multiple of the fundamental frequency), it is not possible to produce stationary waves of this frequency with the string under the same tension.		2 marks
c $f_1 = \frac{v}{2\ell}$	Use the fundamental frequency formula.	1 mark
$v = 2\ell f_1$	Rearrange for the speed of the wave.	1 mark
$v = 2 \times 2.4\text{ m} \times 20\text{ s}^{-1}$	Calculate the answer.	1 mark
$v = 96\text{ m s}^{-1}$		
d For the fundamental frequency, the wavelength is 2ℓ,so the wavelength is: 2 × 2.4 m = 4.8 m		1 mark

Try these yourself

A guitar string is 0.66 m in length. When plucked at its centre it produces a fundamental note of 80 Hz.

a What is the wavelength of the note? (1 mark)

b A guitarist presses the string against the fret, shortening the string to produce notes of different frequencies. How long should the string be to produce a note with a fundamental frequency of 160 Hz? (1 mark)

QUESTION SET 10.5

Remembering

1 What is the difference between the fundamental mode and the first harmonic mode of vibration in a string?

Understanding

2 A standing wave in a spring results from the interference between an incident wave and its reflection. The two waves cancel at the nodes.

Does this mean that energy is destroyed? Explain your answer.

3 When the two component waves producing a standing wave pattern each have a wavelength of λ,what is the distance between:

a adjacent nodes?

b adjacent antinodes?

c a node and the closest antinode?

4 Which pattern(s) in Figure 10.39 could represent a standing wave pattern on a string of length (ℓ) fixed at both ends?

Figure 10.39

Applying

5 The apparatus used to investigate the vibrations of a stretched string or vibrating wire is called a sonometer or monochord (see Figure 10.40).

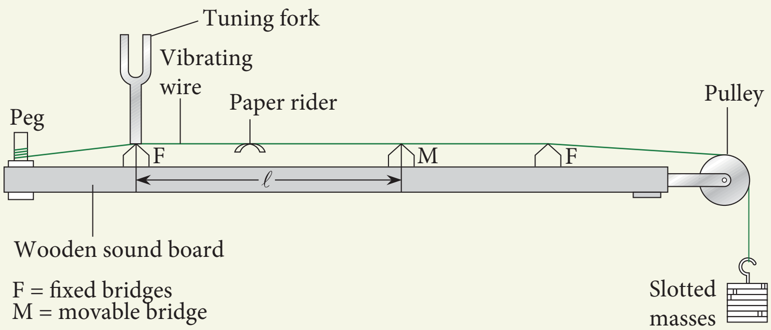


Figure 10.40 The sonometer

The stretched wire on a monochord is 0.80 m long.

a What is the wavelength of the fundamental mode of vibration?

b If the speed of the wave in the wire is 200 m s^{-1}, what is the fundamental frequency?

c If the vibrating length of the wire is shortened, does the fundamental frequency increase or decrease? Give a reason for your answer.

d If you added more slotted masses to the sonometer the frequency of the note it produces will increase. Why is this?

Analysing

6 What is the longest wavelength of a standing wave that can be created between fixed supports 12 cm apart?

7 Two successive overtones of a vibrating string are 300 Hz and 360 Hz. What is the fundamental frequency of the string?

Vibrations in air columns

Longitudinal stationary sound waves can be created in both open and closed pipes. Resonance occurs when sound waves match one of the harmonic wavelengths of the pipe. Open pipes are pipes that are open at both ends and closed pipes are open at only one end. Resonance in air columns is related to the length of the pipe and the speed of sound in air, which is temperature dependent.

Shutterstock.com/bikeriderlondon

▲ **Figure 10.41**
Trombones change the length of the resonator by sliding one tube through another.

Reflection of sound waves in pipes

Waves confined in pipes travel as plane waves, not spherical waves, as they do in the open air. As a compression travels along the pipe it continually reflects off the walls of the pipe. This maintains the compression. When the compression reaches the end of an open pipe it is no longer confined and rapidly expands into the air leaving behind a rarefaction that travels back down the pipe. The compression has been reflected as a rarefaction. This is similar to a fixed end reflection in strings and springs.

When a rarefaction travels along the pipe it continually reflects off the walls of the pipe. This maintains the rarefaction. When the rarefaction reaches the end of an open pipe, the higher pressure air outside the pipe rapidly expands into the rarefaction creating a compression that travels back down the pipe. The rarefaction has been reflected as a compression.

When a compression strikes the closed end of a pipe (the non-yielding part of a pipe) it is reflected as a compression. This is similar to a free-end reflection in strings and springs.

Rarefactions are also reflected as rarefactions from the closed end of the pipe. The standing wave formed in the tube has its maximum air displacement (an antinode) at the open end. This means there will be a pressure node at the open end of the pipe and a pressure antinode at the closed end.

Stationary waves in open pipes

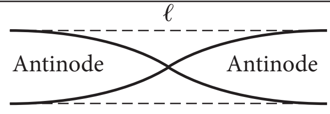
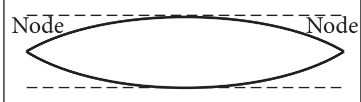
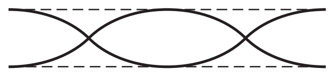
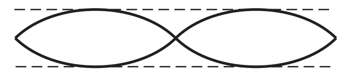
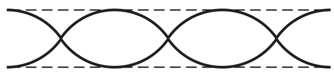
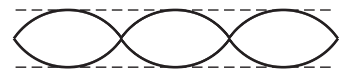
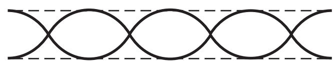
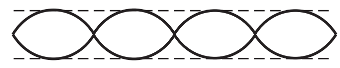
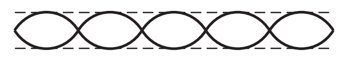
To indicate the stationary wave pattern in an air column, either the pressure variation or the particle displacement could be used. Figure 10.42 represents the particle displacement and pressure variations in an open pipe. In all open pipes, the maximum air displacements (displacement antinode or a pressure node) occur at both ends of the tube, so that its natural frequencies are different from those of a tube closed at one end.

For pipes open at both ends to resonate, the length of the pipe must be related to the resonating wavelength by the relationship:

$$\ell = n\frac{\lambda_n}{2}$$

where ℓ is the length of the pipe, λ is the wavelength of the resonance frequency and n is a whole number relating to that harmonic at which it is resonating (n = 1, 2, 3 ...). In open pipes all harmonics are possible.

Figure 10.42 ▶ The particle displacement and pressure variations of the resonant frequencies of a tube open at both ends

Vibration mode	Particle displacement	Pressure variation	
Fundamental mode of vibration 1st harmonic			$\lambda_1 = 2\ell$ $f_1 = \frac{v}{2\ell}$
1st overtone 2nd harmonic			$\lambda_2 = \ell$ $f_2 = 2f_1$
2nd overtone 3rd harmonic			$\lambda_3 = \frac{2}{3}\ell$ $f_3 = 3f_1$
3rd overtone 4th harmonic			$\lambda_4 = \frac{1}{2}\ell$ $f_4 = 4f_1$
4th overtone 5th harmonic			$\lambda_5 = \frac{2}{5}\ell$ $f_5 = 5f_1$

WORKED EXAMPLE 10.4

Calculate the length of a pipe open at both ends whose fundamental frequency is 320 Hz, when the temperature is such that the speed of sound in the air is 340 m s^{-1}. (3 marks)

Answer	Logic	
$f_1 = \frac{v}{2\ell}$	Select the relationship.	1 mark
$\ell = \frac{v}{2f_1}$ $= \frac{340 \text{ m s}^{-1}}{2 \times 320 \text{ s}^{-1}}$	Rearrange the formula and substitute the values.	1 mark
$\ell = 0.53$ m	Solving for the answer	1 mark

Try these yourself

1 The fundamental harmonic of a pipe open at both ends is 640 Hz. What is its length? (3 marks)

2 A 2.2 m pipe open at both ends is resonating at its second harmonic frequency.

a What is the wavelength of the standing wave? (2 marks)

b What is the frequency of the second harmonic? (2 marks)

Stationary waves in closed pipes

Consider an air column that is in a tube closed at one end. When it is resonating, a pressure antinode (displacement node) occurs at the closed end. Figure 10.43 represents the particle displacement and pressure variations in an pipe closed at one end.

For a closed pipe to resonate, the length of the pipe must be related to the resonance wavelength by the relationship:

$$\ell = (2n - 1)\frac{\lambda}{4}$$

The fundamental resonance mode when $n = 1$:

$$\ell = \frac{\lambda}{4}$$

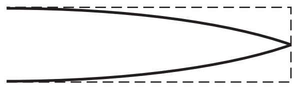
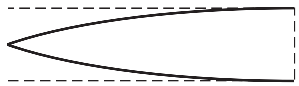
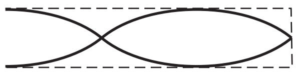
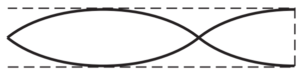
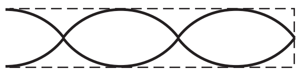
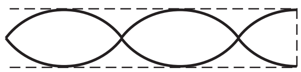
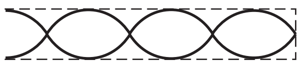
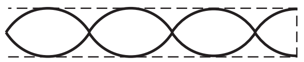
The resonance mode when $n = 2$ (third harmonic)

$$\ell = \frac{3\lambda}{4}$$

The resonance mode when $n = 3$ (fifth harmonic)

$$\ell = \frac{5\lambda}{4}$$

Only the odd harmonics are present in a closed pipe. This means that musical instruments with closed pipes can only play half the notes that those with open pipes can play.

Vibration mode	Particle displacement	Pressure variation	f & λ
Fundamental mode of vibration 1st harmonic	 Antinode Node	 Node Antinode	$\lambda_1 = 4\ell$ $f_1 = \frac{v}{4\ell}$
1st overtone 3rd harmonic			$\lambda_2 = \frac{4\ell}{3}$ $f_2 = 3f_1$
2nd overtone 5th harmonic			$\lambda_3 = \frac{4\ell}{5}$ $f_3 = 5f_1$
3rd overtone 7th harmonic			$\lambda_4 = \frac{4\ell}{7}$ $f_4 = 7f_1$
4th overtone 9th harmonic			$\lambda_5 = \frac{4\ell}{9}$ $f_5 = 9f_1$

◀ **Figure 10.43**
A particle displacement and a pressure variation representation of the resonant frequencies of a tube closed at one end. All have maximum particle displacement at the open end and none at the closed end.

WORKED EXAMPLE 10.5

When the air is at 0°C, what will be the fundamental frequency and the frequency of the first two overtones for an organ pipe 2.4 m long if it is:

a open at both ends? (5 marks) **b** closed at one end? (5 marks)

The speed of sound is 331 m s^{-1} at 0°C.

Answers	Logic	
a $f_1 = \frac{v}{2\ell}$	Use the correct relationship.	1 mark
$f_1 = \frac{331\ \text{m s}^{-1}}{2 \times 2.4\ \text{m}}$	Substitute the known variables.	1 mark
$f_1 = 69$ Hz	Calculate the answer.	1 mark
The frequencies of the first two overtones are $2f_1$ and $3f_1$, that is 138 Hz and 207 Hz.	State the answer.	2 marks

▶

b	$f_1 = \frac{v}{4\ell}$	Use the correct relationship.	1 mark
	$f_1 = \frac{331\ \text{m s}^{-1}}{4 \times 2.4\ \text{m}}$	Substitute the known variables.	1 mark
	$f_1 = 34.5$ Hz	Calculate the answer.	1 mark
	As only the odd-numbered harmonics are present, the frequencies of the first two overtones are: $3f_1$ and $5f_1$, that is 103.5 Hz and 172.5 Hz.	State the answer.	2 marks

Try these yourself

A girl blows across the mouth of a bottle and sets the air column inside it resonating at its third harmonic of 930 Hz. The speed of sound in the air is 340 m s^{-1}.

a What is the length of the bottle? (5 marks)

b What is the frequency of the next harmonic? (5 marks)

EXPERIMENT 10.2

FINDING THE SPEED OF SOUND BY AIR COLUMN RESONANCE

The speed of sound in air varies with temperature, moisture content and the local atmospheric pressure.

Aim

To find the speed of sound in air in your classroom

Materials

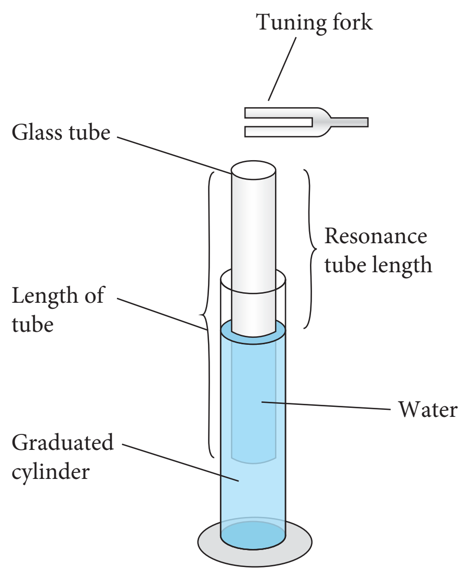
- a large graduated cylinder
- a glass tube of about 2–3 cm in diameter and about 30 cm long
- a tuning fork with a fundamental frequency of 440 Hz
- a ruler
- a marking pen that can write on wet glass

Prepare a risk assessment for your experiment and have your teacher approve it before proceeding.

What are the risks in doing this experiment?	How can you manage these risks to stay safe?

Procedure

1 Set up the equipment as shown in Figure 10.44.
2 Fill the graduated cylinder with water to just below the top.
3 Lower the glass tube until the top is just above the surface of the water.
4 Strike the tuning fork on a surface that will get it oscillating strongly and hold it near the open end of the tube.
5 Raise the tube slowly until it begins to resonate.
6 With the pen mark the base of the tube where it contacts the water when the loudest resonance is detected.
7 Measure the length from the pen mark to the top and record it.
8 Repeat the procedure twice and average the three values.
9 Analyse the results, given that for the fundamental frequency the resonance length is $\ell = \frac{\lambda}{4}$.



▲ **Figure 10.44**
Experimental set-up

Results

Record your results in a table like the one below.

Data	Trial 1	Trial 2	Trial 3	Average
Frequency of the tuning fork				
First resonance length				

Analysis of results

1 Use the data to calculate your best estimate of the speed of sound. Include the uncertainty in this value.
2 A commonly accepted value for the speed of sound is (341 ± 1) m s^{-1}.
Compare your best estimate with this accepted value.
  a To what extent do the two values overlap?
  b Calculate the percentage difference between the best estimate of the accepted value and your best estimate of the value.
  c Comment on the accuracy and precision of your value compared with the accepted value.

Discussion

1 What atmospheric conditions may have affected your result?
2 If you did the experiment on another day when the temperature in the room was hotter how would your result change? Explain your answer.
3 What changes could you make to the experiment to make the data gathered more accurate?
4 If you replaced the air in the glass tube with carbon dioxide would the value of the speed of sound change. Explain your answer.
5 Draw a diagram showing the pressure node and antinode in the closed tube when it was resonating at the tuning fork's fundamental frequency.
6 If you repeated the experiment with a tuning fork that had a higher fundamental frequency, would the first resonance occur in a longer or shorter air column? Justify your answer.

Conclusion

Write a conclusion to answer your aim.

QUESTION SET 10.6

Unless otherwise stated, take the velocity of sound in air at room temperature to be 340 m s^{-1}.

Remembering

1 A tuning fork sounds louder when its stem is touched to a bench. Why is this?

Understanding

2 A tube of length ℓ is open at both ends. A stationary wave is set up in the tube. Only waves with certain frequencies will cause resonance within the tube. Which of the following gives the set of wavelengths that can exist in a tube open at both ends (n = 1, 2, 3, . . .)?

a $n\ell$

b $\frac{\ell}{n}$

c $\frac{4\ell}{2n-1}$

d $\frac{2\ell}{n}$

Applying

3 Water is poured into a long metal tube closed at one end until the shortest resonant length is found for a fork of frequency 256 Hz. If the length of the air column in the tube is 31.0 cm, what is the velocity of sound in the air at the time?

4 A vertical pipe of length 1.40 m is filled with water, which is allowed to run out slowly from the lower end, while a vibrating tuning fork of frequency 512 Hz is held over its open end. How many positions of resonance will be obtained?

5 The human ear is most sensitive to sounds of a frequency of about 5000 Hz. The outer ear canal can be modelled as a tube closed at one end (Figure 10.45). Assuming that this frequency corresponds to the fundamental frequency, what is the length of the outer ear canal of a human?

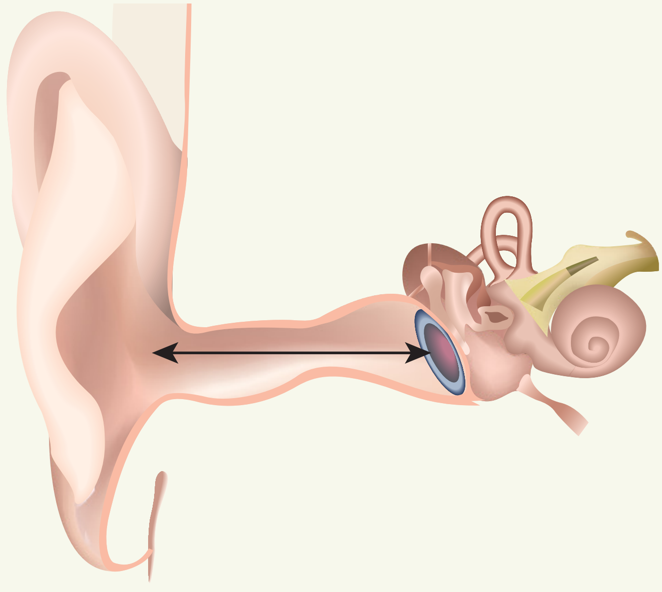


Figure 10.45 ▶ The human outer ear canal

Analysing

6 The frequency of the maximum sensitivity for a domestic cat is different from that of humans. The frequency depends on the length of the ear canal. Will the frequency of maximum sensitivity for the cat be higher or lower than that for humans? Explain your answer.

Objective and subjective measures of sound

In Figure 10.15, changes to amplitude are seen as higher peaks on a screen. Amplitude can be increased without changing frequency. Amplitude change causes changes in **loudness**, but so do changes in frequency. Loudness is a subjective experience that is affected by both amplitude and frequency. The ear is more sensitive to sounds at 3 kHz than at 1 kHz for example.

The **intensity** of the sound is an objective measure of the rate at which the wave motion carries energy. Energy E (joule, J) per unit time t (s) is power P (watt, W), so sound intensity is power transmitted through a unit area, A. The unit of intensity is W m^{-2}.

$$P = \frac{E}{t}$$

$$I = \frac{P}{A}$$

$$\Rightarrow I = \frac{E \div t}{A}$$

$$I = \frac{E}{At}$$

Experiments show that intensity is directly proportional to the square of the amplitude of the wave: $I \propto \text{amplitude}^2$. Thus, when the wave amplitude is doubled, the wave carries $2^2 = 4$ times more energy than before.

The intensity of a sound wave is measured by the energy passing per unit time through a unit area taken at right angles to the direction of propagation of the wave.

$$I = \frac{P}{A}(\text{Unit : W m}^{-2})$$

Intensity $\propto$ amplitude^2

The decibel scale

The human ear can detect sound pressures from as low as 2×10^{-5} Pa – the threshold of hearing – to ten million times greater at 100 Pa – the threshold of pain. The **decibel scale** was developed to make these numbers more manageable. The scale is changed to a logarithmic scale. That is, sound level, L (dB), does not directly translate to intensity, I (W m^{-2}).

A decibel (dB) is one tenth of a bel. It is named after Alexander Bell (1847–1922), a prolific inventor who is credited among many things with inventing the telephone. It is a logarithmic scale based on the ratio of the actual sound intensity (I) and the sound intensity of the threshold of hearing (I_0).

$$L\ (\text{dB}) = 10 \log \frac{I}{I_0}$$

The human brain hears in a type of logarithmic scale. A tenfold increase in intensity causes a 10 dB increase in loudness. This is perceived as a doubling of the loudness of the sound. In other words a 10-decibel (1 bel) sound delivers 10 times as much energy as a 0 decibel sound, and a sound of 20 decibels (2 bel) delivers 10^2 times as much energy. The human ear can detect changes in sound intensity of about 3 dB. Approximate sound intensities and sound intensity levels of common sounds are given in Table 10.3.

Table 10.3 Approximate sound intensities and sound intensity levels of common sounds

Sound	Intensity ($W m^{-2}$)	Sound level (dB)
Threshold of hearing	10^{-12}	0
Rustle of leaves	10^{-11}	10
Whisper	10^{-10}	20
Quiet radio	10^{-8}	40
Normal conversation	10^{-6}	60
Busy street traffic	10^{-5}	70
Loud radio	10^{-4}	80
Noisy factory	10^{-4}	85
Motor car horn	10^{-3}	90
Jackhammer	10^{-2}	100
Thunder	10^{-1}	110
Rock concert	1	120

The phon: a comparative measurement of loudness

Loudness is measured in **phon**. The phon depends on both the intensity and the frequency of the sound. To determine the loudness of a sound, the intensity of a 1000 Hz sound is changed until it is sensed by a listener to be the same loudness as the sound being measured. The loudness of the measured sound in phon is then equal to the intensity level of the adjusted reference sound in decibel above the threshold of hearing. A note of frequency 1000 Hz with a sound intensity level of 50 dB has a loudness of 50 phons. Sounds at other frequencies are then compared to these standard loudnesses. For example, all sounds that are perceived to be as loud as a 1000 Hz, 50-phon sound are plotted on an equal 50 phon (isophon) line.

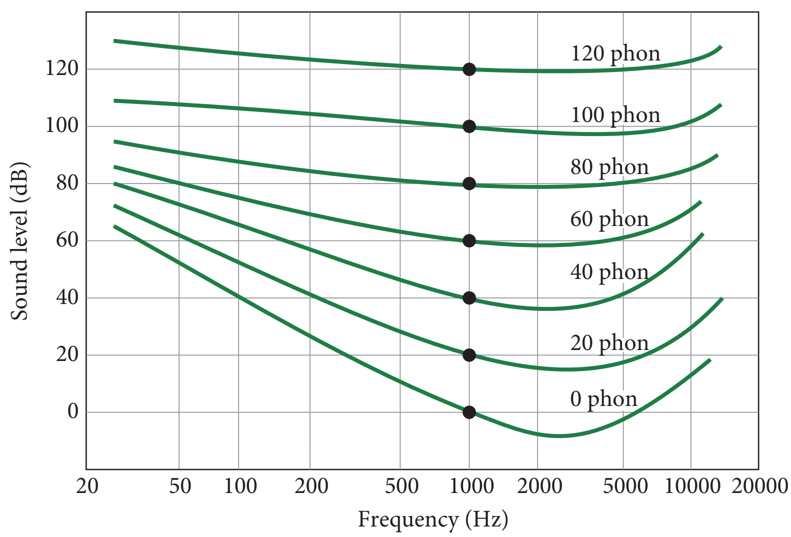


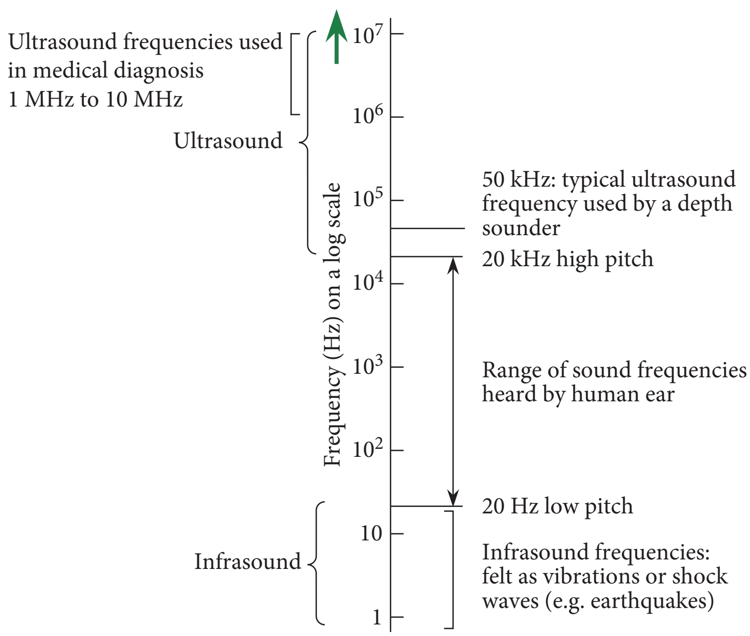
Figure 10.46 ▶ Graph of the sensitivity of the human ear. The large dots represent the *reference loudness*, which is the perceived loudness of a 1000 Hz sound at different sound levels. The curves indicate the sound levels that produce the same perceived loudness at other frequencies.

Range of human hearing

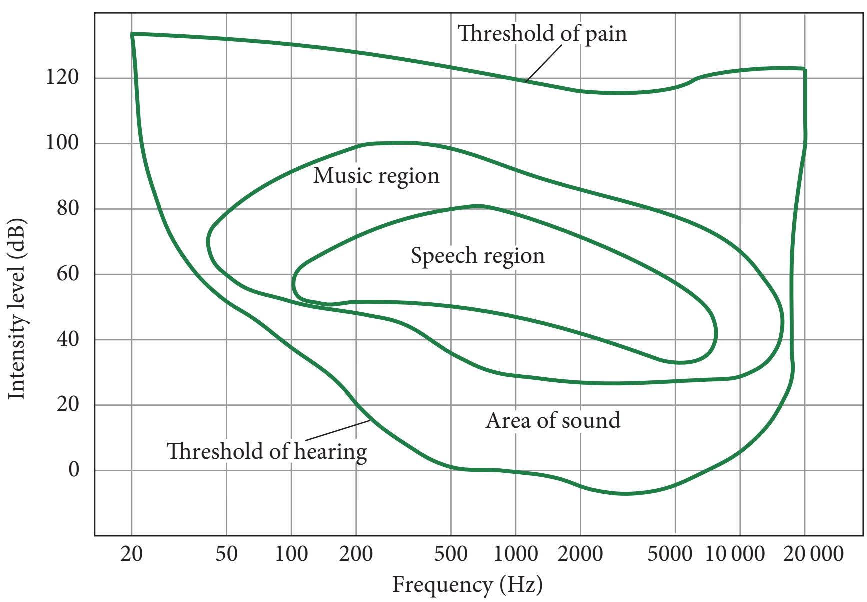
The average human ear can hear sounds with frequencies from about 20 Hz (**lower limit of audibility**) to 20 000 Hz (**upper limit of audibility**). **Infrasound** is any sound with a frequency less than 20 Hz. Some low frequency sounds, infrasound, can be sensed by the human body, but not as audible sounds. They have been associated with car and industrial accidents. Any sound of frequency greater than 20 kHz (20 000 Hz) is called ultrasound. Ultrasound imaging is used in medicine.

Figure 10.48 shows the range of response of the human ear. This is an average result obtained by testing many people. Above the curve, sound causes pain. Below the curve, hearing is difficult. An increase in intensity level can assist hearing.

Figures 10.47 and 10.48 show that frequency and intensity together influence human experience of hearing.



◀ **Figure 10.47**
Frequency spectrum



◀ **Figure 10.48**
Audibility range for the human ear

Sound quality

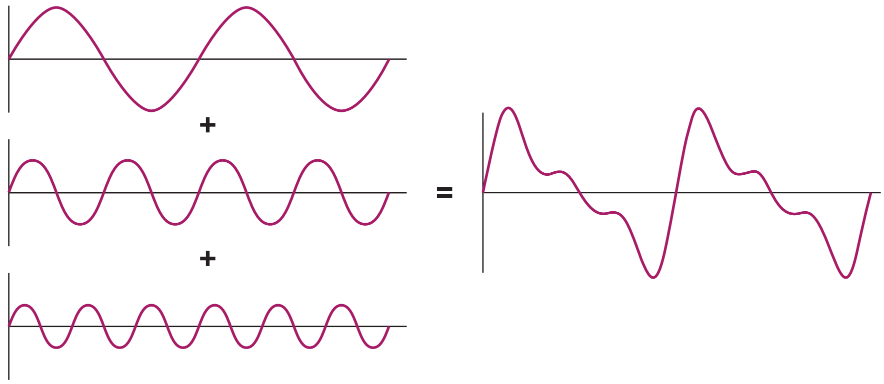


Figure 10.49 ▶ The sound spectrum of the fundamental and the next two harmonics combine to produce the distinctive shape of the sound waveform that gives the sound its timbre.

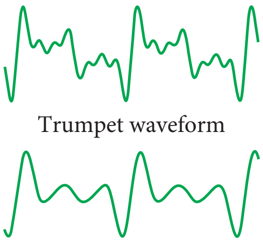


Figure 10.50 ▲ Sound spectrums produced by a clarinet and a trumpet

Imagine you are in a room with your eyes closed. A note is played on a saxophone and then the same note is played on a piano. You can tell which is the piano and which is the saxophone. Their **timbres** are different. Timbre, **tone colour** or **tone quality**, is a subjective experience. It is based on the proportions of different harmonics – the frequency spectrum – produced by an instrument.

The timbres of instruments are affected by their different shapes, the materials used, and how the sound is produced.

Sound measures can be objective or subjective.

Objective measures: frequency, wavelength, speed, intensity

Subjective measures: pitch, timbre, loudness

QUESTION SET 10.7

Remembering

1 What is the difference between frequency and:
  a pitch? b timbre?

2 What is the difference between an objective measure of sound and a subjective measure of sound? Explain using intensity and loudness as examples.

Understanding

3 Why do different musical instruments sound different when playing the same note?

4 In Figure 10.46:
  a what does the 20 phon line mean?
  b at what sound level is a 200 Hz signal heard as 20 phon?

Applying

5 The threshold of hearing is defined as the least intensity of a 1 Hz signal that can just be heard by a 'normal' ear.
  a Use the information in Figure 10.46 to determine the frequency at which the 'normal' human ear is most sensitive.
  b What is the intensity of this sound?

Analysing

6 At what sound level, must a frequency of 50 Hz be produced to be perceived to be as loud as a sound of frequency 1000 Hz produced at a sound level of 60 dB? (Use Figure 10.46 to answer the question.)

Reflecting

7 How do you now understand the nature, production and hearing of sound? Give specific details.

Imaging with mechanical waves

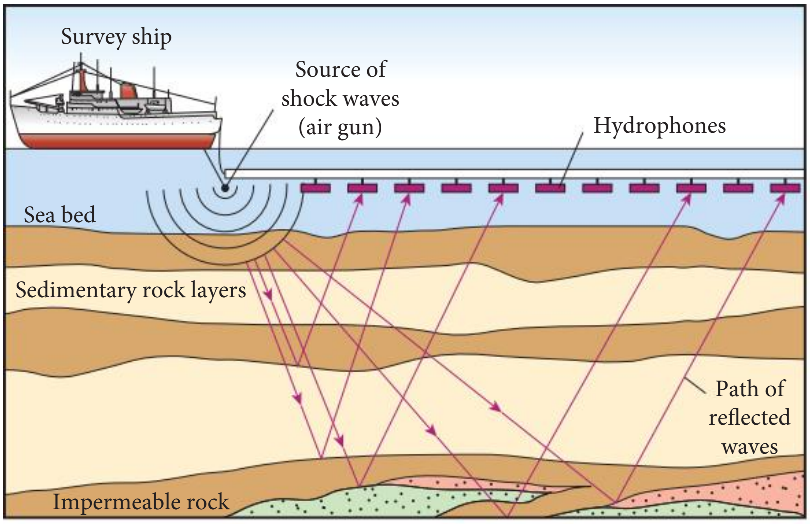
Interactions of mechanical waves with matter provide useful information or images. For example, the shock waves created by earthquakes provide images of the structure of Earth's interior, and ultrasound waves are used to provide images of the structure of parts of the human body.

Looking below the surface of Earth

Scientists use mechanical waves to probe the nature of the subterranean structure of Earth. **Seismic reflection** is a technique developed by groups of collaborating scientists for mapping rock layers underneath the ground. An artificial seismic wave is generated at a particular location on land or at sea. These waves travel into the ground and are reflected and refracted at boundaries between different rock layers of differing densities. Seismic detectors are carefully placed to record the arrival times of these waves.

Analysing the elapsed time data enables scientists to construct two- and three-dimensional geological maps with the time and directional data. These maps are used extensively in exploration for minerals, natural gas and petroleum.

a



b

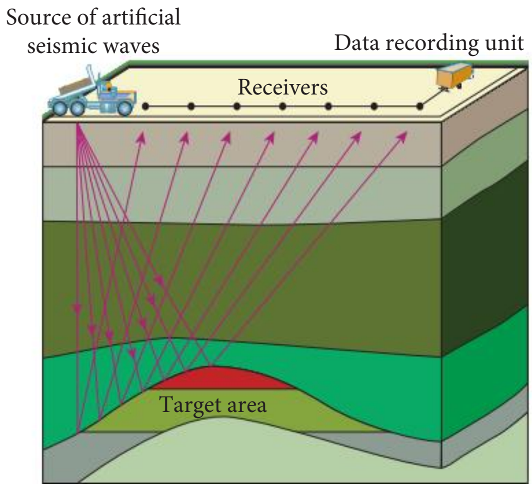
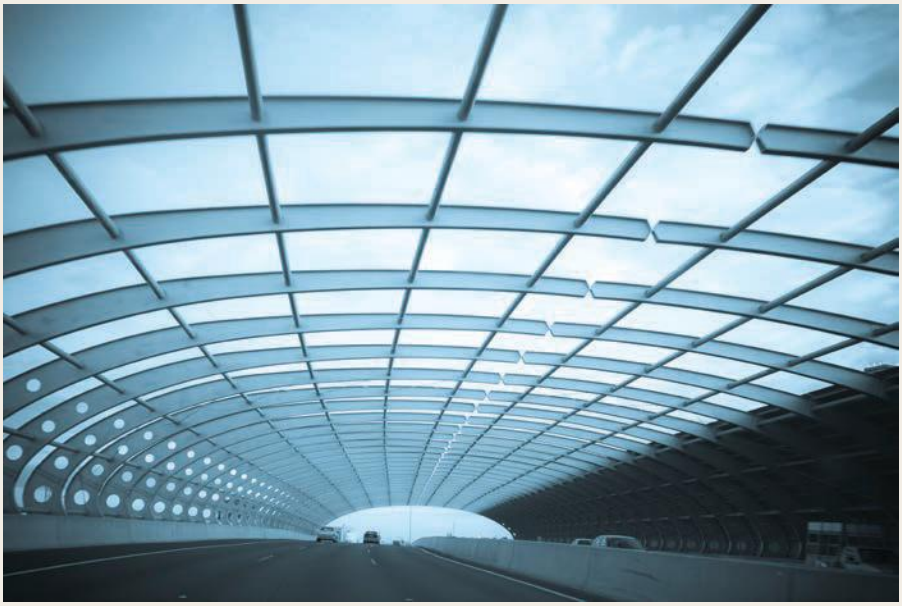


Figure 10.51
a) An array of detectors towed behind a boat gathers seismic data in a marine setting.
b) An array of detectors embedded in the ground gathers seismic data in a terrestrial setting.

Scientific literacy: Road noise pollution and acoustic design



▲ Figure 10.52
The noise-reducing sound tube tunnel on the Tullamarine Freeway in Melbourne, Victoria

Noise is among the most pervasive pollutants today and is a growing problem. The word noise is derived from 'nausea,' meaning seasickness. It has a negative effect on human health and wellbeing. Noise in today's busy world is unavoidable but with clever design, regulation and self-control it can be kept to safe levels.

Road noise is an example of a 'line source' of noise pollution; it is not emanating from a point source, but a continuous line of moving traffic. The level of noise is a function of volume of traffic, type of vehicle, and speed. Regulations limit the amount of noise some vehicles can produce and some require vehicles to be properly operated and maintained. Despite regulations, the noise levels are usually only reduced by 5 to 7 dB.

Solutions to controlling road noise include using buffer zones between the traffic and the residential areas, as well as barriers in the form of fences, earthworks, walls and thick vegetation.

Barrier designs consider how to maximise blocking the unwanted noise from reaching the areas that need to be protected. Diffraction of sound must be addressed. Sound waves bend downwards when they pass over the top of a noise barrier. This means the height of the noise barrier becomes important. Refraction complicates things even further. If there are temperature gradients created within the buffer zones, the sound can be refracted into the areas the barriers are trying to protect.

The materials a road surface is made of and its texture (rough or smooth) affects the intensity and spectrum of sound emanating from the tyre–surface interaction. Roadway surfacing choices have been shown to contribute a decrease in road noise of up to 4 dB. Because sound levels are measured using a logarithmic scale, a reduction of 9 dB is equivalent to eliminating approximately 80 per cent of the unwanted sound.

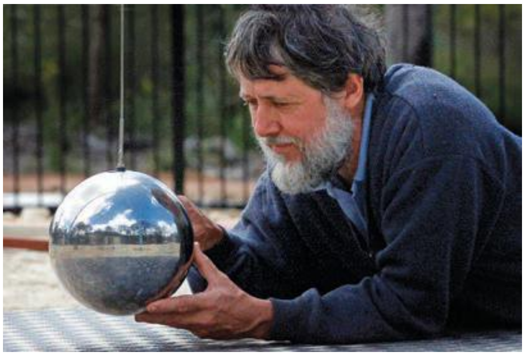
In built-up urban areas, space is limited and simple buffer zones and barriers are not a workable solution. More complex and more expensive solutions, such as sound tubes or putting the roads underground, are required. Sound tubes can be very effective in overcoming the complicated problems of refraction and diffraction and be aesthetically pleasing (Figure 10.52). The more expensive solution is to put the roads underground by cutting a trench, laying the road, then covering with a concrete slab or tunneling. This contains the sound pollution within the tunnel. Normally, the benefits of noise reduction far outweigh the aesthetic impacts for residents protected from unwanted sound.

Questions

1 Road noise is what form of noise pollution?
2 Why is diffraction a problem with 'barrier noise abatement'?
3 What urban situation would make a sound tube tunnel a good solution to traffic noise pollution?
4 How can only a 9 dB reduction in noise produce an 80% decrease in the noise?
5 What do you need to consider when building systems to protect residents from road noise?
6 Do you think that this article gives a comprehensive overview of solutions to road noise pollution? What is one other piece of information you would like to have had provided?

Case study

Professor David Blair



▲ Figure 10.53
Professor David Blair

David Blair is a Winthrop Professor at the University of Western Australia, and Director of the Australian International Gravitational Research Centre. He has spent many years working on the challenge of detecting Einstein's gravitational waves. Gravitational waves are wave disturbances in the fabric of space–time and are a bit like sound waves except they travel through empty space at the speed of light. When he began his research there were just a few dozen researchers in a few universities in Europe and USA trying to devise and test the best designs for gravitational wave receivers. Now there are many research groups collaborating worldwide.

Professor Blair pointed out that we know of only two types of waves that travel through empty space at the speed of light. The first are electromagnetic waves, which were predicted in about 1860 and first demonstrated 25 years later. This discovery led to the successful exploitation of electromagnetic waves during the 20th century, and has revolutionised our lives.

The second type of wave is gravitational waves. Nature gives us another spectrum like another sense, and one that so far is known only theoretically. Today after 40 years of effort, scientists around the world are confident that their latest receivers will lead to the successful detection of gravitational waves. Just as exploration of the universe with electromagnetic waves has revolutionised our understanding of the universe, physicists think that gravitational waves will provide us with new insights and will probably bring surprises.

The latest detectors of gravitational waves use very powerful laser beams to detect the tiny vibrations in the distance between distant mirrors as gravitational waves go past. Because gravity is just curved space, gravitational waves are just ripples in the curvature of space. When space ripples, the distances between objects change. But because it takes vast amounts of energy to bend space, the waves are hard to detect because the ripples are infinitesimal. So far scientists have been able to measure mind-bogglingly small vibrations of about 10^{-20} of a metre changes in distance between mirrors 4 km apart, but this is still about 10 times too big for confident detection. The latest receivers are expected to overcome this barrier and begin exploring the new spectrum. It will be headline news when the first detection happens.

Professor Blair explained that gravitational wave detection involves a huge range of technologies. At the heart of the detectors are very powerful lasers and the most perfect mirrors ever created. Complex optical circuits manipulate and control the light to enable sufficient measurement sensitivity. But detectors cannot work unless vibrations from the outside world are suppressed. So the mirrors need to be suspended from sophisticated vibration isolation systems. Gravitational wave data come from changes in the light intensity coming out of the detector. These data then need to be very carefully analysed using supercomputers to distinguish between possible signals and noise entering the detector.

Professor Blair has led the development of a gravitational wave research centre at Gingin, about 80 km north of Perth. The WA government provided a 50 km^2 site for a future large detector, but at present a research team is working on solving technical problems that are likely to create difficulties in the new detectors being constructed in the northern hemisphere. But Professor Blair's goal is to see the creation of a large Australian detector. This is needed because without one in the southern hemisphere it is impossible to pinpoint the sources of signals.

Professor Blair said that the most exciting signal they hope to see will be the coalescence of two black holes. These signals make a characteristic 'chirp' sound that will reveal all the complexities of black holes and may lead to new understandings about the creation of the universe.

Questions

1 What evidence is there that gravitational waves are or are not mechanical waves?

2 What combinations of technologies are being refined to make the detection of gravity waves possible?

3 Professor Blair stated that the discovery of electromagnetic waves revolutionised our lives. What changes can you think of that have been brought about by their discovery?

4 What are the gravitational wave detectors attempting to actually measure?

5 The information electromagnetic radiation carries has allowed astronomers using telescopes to gain an understanding of the structure and history of the universe. What benefits do you think the information and energy contained in gravitational waves may bring in the future?

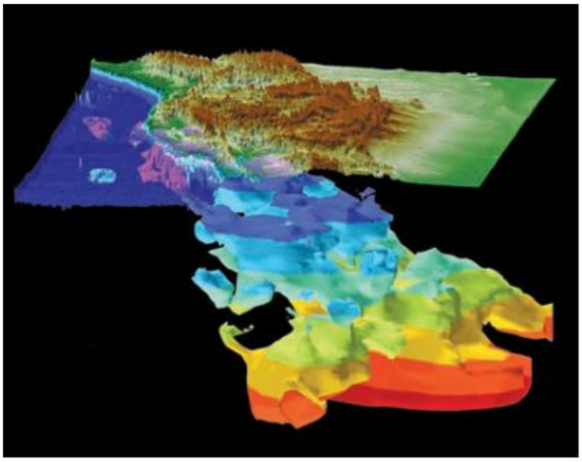


Figure 10.54 ▲
A three-dimensional image of a seismically active slab of Earth's mantle under western North America known as the Farallon plate. The data for this image were obtained from hundreds of ground-based sensors that measure seismic activity. Colours represent 200-kilometre depth intervals.

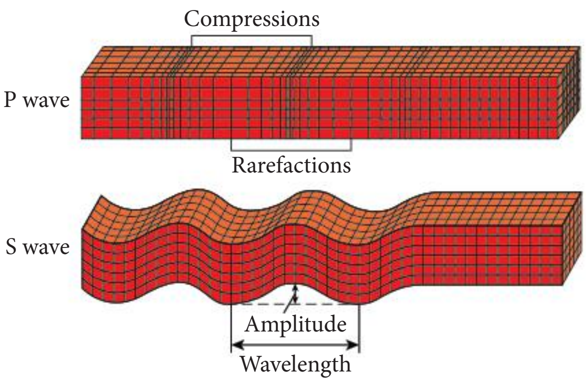


Figure 10.55 ▲
Propagation of longitudinal primary (P) and transverse secondary (S) shear waves

Earthquakes reveal Earth's inner structure

The outer layer of Earth is made up of tectonic plates. Earthquakes happen when these plates catch when slipping past one another. The pressure builds up at the catch points as the plates continue to press on each other. When the pressure gets too great, the rock gives way and the plates suddenly slip alongside each other with a jolt. The stored energy is released abruptly. Vibrations travel as shockwaves (seismic waves) through Earth's interior.

These seismic waves radiate in all directions from the point underground where the energy was released. This point is known as the **seismic focus**. Directly above this is the earthquake's **epicentre** – the point on Earth's surface where the earthquake will be experienced most strongly. If this is in an inhabited area it is the point at which the most damage is done.

Waves that propagate within the ground are called **body waves**. There are two types of body waves, the primary or **P wave** and the secondary or **S wave**.

Primary (P) waves are longitudinal compression waves (sound waves). Secondary (S) waves are transverse waves; they are also called **shear waves**. Their velocities vary with the density of the rock they pass through. The density of rock increases with depth due to the pressure of the rock above. This means the speed of the waves increases with depth. Different rock composition also affects the speed of the waves. This density gradient refracts the waves back up to the surface in a curved concave path, where they can be recorded on a **seismograph**.

Seismographs record the amplitude and frequency of seismic waves and yield information about Earth and its subsurface structure.

When an earthquake occurs, the seismic P and S waves spread out in all directions through Earth's interior. These waves

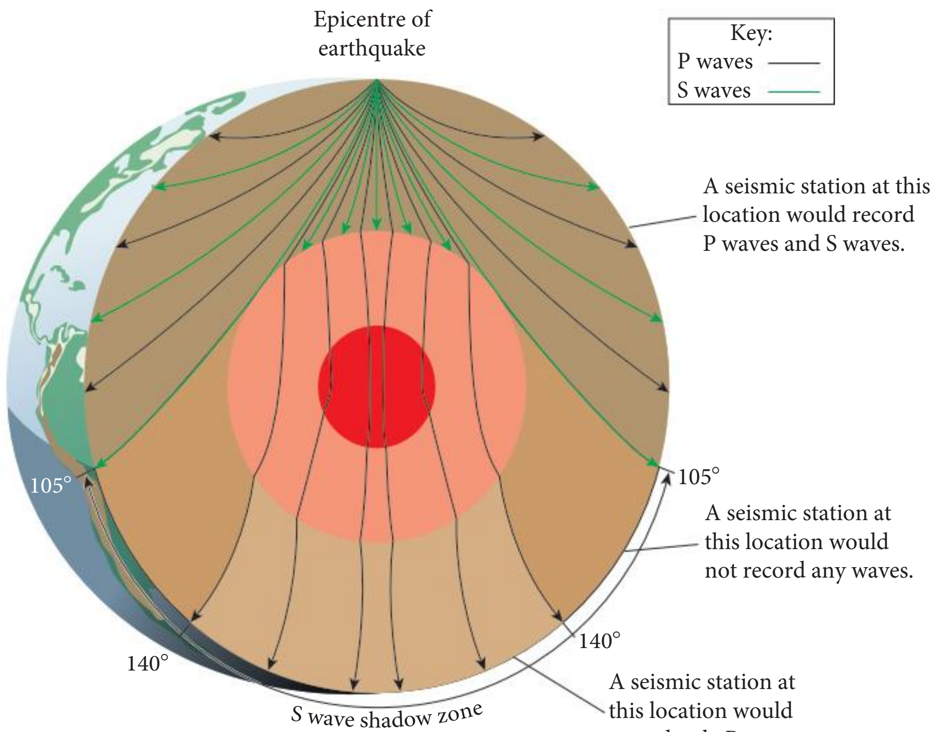


Figure 10.56 ▶
Recording stations around the world use the mechanical waves generated by earthquakes to probe the structure of Earth.

propagate all the way from one side of Earth to the other, so we can look all the way to Earth's centre. Studying these waves has also revealed the relative dimensions and densities of the other parts of Earth's internal and surface structure.

This is how we know that Earth's core is liquid. S waves are unable to travel through liquids, and this particular wave trait indicates a boundary between the solid mantle and the liquid outer core. The boundary between the inner core and outer core is also detected by a difference in velocities of P waves. As S waves do not propagate through the outer core, a 'shadow' zone is created on the other side of Earth from the source of the seismic waves. In this zone no S waves are detected.

Ultrasound in medicine

Ultrasound is a high frequency sound wave (>20 kHz). It is used to detect objects and measure distances. For example, it is used as a range finder in cameras with automatic focusing systems. It is also used to detect structural flaws in materials that are invisible to the eye, for accelerating chemical processes, and in medical therapy to enable accurate injection into muscles or to break up kidney or bladder stones (calculi). **Ultrasonography** is used extensively in medicine to image organs and tissue anomalies such as cancers, and to monitor the health of babies in the womb.

The ultrasound CPU (central processing unit) calculates the distance from the probe of the reflections from the tissue boundaries. It does this by using the speed of sound in tissue, which is about 1540 m s^{-1}, and the time for each echo to return. The distances and intensities of the echoes are displayed on a screen as a multidimensional image. In a typical ultrasound, thousands of pulses and echoes are sent and received each second. The probe can be moved along the surface of the body and be angled to obtain various views.

Newspix/Brad Newman

▲ Figure 10.57
A three-dimensional image of an eight-month-old foetus

Ultrasound machines capable of three- and four-dimensional imaging have been developed. These devices take several two-dimensional images by moving the probes across the body surface or rotating inserted probes. The two-dimensional scans are then combined by computer software to form these images. The three-dimensional image is a delayed image and a four-dimensional image is a real-time image.

QUESTION SET 10.8

Remembering

1 What are seismic waves?

2 Which mechanical wave, the P wave or the S wave is a sound wave? Justify your answer.

Understanding

3 Earthquakes release large shockwaves that travel through and around Earth. How can these be used by scientists to determine Earth's internal structure?

4 Why do P waves and S waves refract back up towards Earth's surface?

5 Why are sound waves with frequencies greater than 20 000 Hz called ultrasound?

Applying

6 An ultrasound transmitter/receiver produces a 2 MHz signal and is placed on the skin of a patient. A tumour 5.1 cm from the surface reflects the sound wave back to the receiver.

a How long will it take the transmitted ultrasound signal to return to the transmitter/receiver?

b What effect on the image would using a ultrasound transmitter/receiver that produced a 4 MHz signal have?

Analysing

7 Explain how the sonar 'pings' sent out from a submarine are used to determine:

a the separation distance to a nearby submarine.

b the relative motion of another submarine.

Reflection

8 What effects do you think improving the science of seismography (the study of Earth's interior using seismic waves) will have on the ability to predict earthquakes?

CHAPTER SUMMARY

- Mechanical waves need a medium for transmission.
- Mechanical waves can be longitudinal or transverse vibration modes.
- Mechanical waves transfer energy without the transfer of the medium it is passing through.
- Mechanical waves can be reflected, refracted and diffracted.
- The period and frequency of a wave or vibration are related: $f = \frac{1}{T}$
- The velocity of a wave is related to the wavelength and frequency: $v = \frac{\lambda}{T} = f\lambda$
- Reflection is the bouncing of waves off an interface.
- Refraction is the bending of a wave at an interface between two media.
- Refraction is caused by the way in which different media affect the speed of waves.
- Sound waves are longitudinal.
- Waves can pass through each other and remain unchanged when they have passed (superposition principle).
- At any instant when waves are passing through each other, constructive or destructive interference can occur.
- Constructive interference occurs when two waves passing through each other result in a wave of greater amplitude.
- Destructive interference occurs when two waves passing through each other result in a wave of lesser amplitude.
- Resonance occurs when an external driving vibration transfers maximum energy to another system with the same natural frequency.
- Standing waves result from the superposition of two continuous waves travelling in opposite directions with the same frequency and amplitude.
- The distance between nodes in a standing wave is half a wavelength.
- The fundamental frequency of a system is the lowest frequency at which a standing wave can form.
- Harmonics are whole number multiples of the fundamental.
- Two notes of the same frequency from different sources sound different because of the presence of overtones.
- Mechanical waves generated by earthquakes give us information about the structure of Earth's interior.

CHAPTER GLOSSARY

amplitude the maximum displacement of the particle of a wave from its mean position

antinode point along a standing wave at which the wave has maximum amplitude; it is the result of a crest overlapping a crest or a trough overlapping a trough

body waves seismic waves that pass through the body of Earth

compression region of high pressure in a mechanical wave

constructive interference waves interacting and increasing in amplitude at the point of interaction

continuous wave continuous wavefronts passing through a medium

crest the positive peak of a wave

decay the decrease in amplitude when a vibrating force has been removed

decibel scale a logarithmic scale of sound level

density the mass per unit volume of a substance under specified conditions of pressure and temperature

destructive interference waves interacting and decreasing in amplitude at the point of interaction

diffraction the bending of a wave around objects or the spreading of the wave after passing through a gap

Doppler effect the shift in frequency and wavelength of waves that results from the relative motion of source and receiver

epicentre the point on the surface directly above the seismic focus

first harmonic the simplest mode of vibration and accounts for the fundamental tone

forced vibration vibration occurring when an object is forced to vibrate by another vibrating object

free or natural vibrations vibrations occurring when an object is vibrating by itself

frequency the number of whole waves or oscillations generated in one second

infrasound a sound with a frequency below the range of human hearing

intensity a measure of the energy passing per unit time through a unit area taken at right angles to the direction of propagation of the wave; energy per unit area per unit time

longitudinal wave a wave whose particles oscillate about a mean position in the same line as the direction of travel of the wave

loudness a subjective quality of perception of the amount of sound energy arriving at a person's ear

lower limit of audibility the lowest frequency a human can hear (approximately 20 Hz)

mechanical wave a disturbance that requires a material medium (solid, liquid or gas) for its propagation

node point along a standing wave at which the wave has minimal amplitude; it is the result of the overlap of a crest with a trough

P wave longitudinal compression waves that pass through the body of Earth

period (*T*) the time it takes before a wave repeats itself

phon a measure of loudness; it depends on both intensity and the frequency

pitch a subjective sensory characteristic related to frequency

pulse a single wavefront passing through a medium

rarefaction region of lower pressure in a mechanical wave

ray a line drawn at right angles to the wavefront and in the direction of propagation

reflection (law of) the angle of reflection (r) equals the angle of incidence (i)

refraction bending of waves as they pass through different media

resonance oscillation induced in a physical system when it is affected by another system that is itself oscillating at the right natural frequency

reverberation the effect that occurs when too many sound wave reflections arrive at your ear for you to distinguish between the sounds

S wave a transverse earthquake wave that shakes the ground back and forth perpendicular to the direction the wave is moving; also known as a shear wave

seismic focus the underground point from which earthquake energy is released

seismic reflection a technique developed for mapping the rock layers underneath the ground

seismic waves mechanical waves of energy that travel through Earth's layers that result from earthquakes, explosions or volcanic activity and are propagated within Earth or along its surface

seismograph a device that records the amplitude and frequency of seismic waves and yields information about Earth and its subsurface structure

shear waves one of two types of seismic waves; also known as S waves, they are transverse waves whose velocities vary with the density of the rock they pass through

sound wave movement of energy by longitudinal vibration through a medium

stationary wave or standing wave a wave that oscillates in place, without transmitting energy along its extent. Standing waves have stable points, called nodes, where there is no oscillation

superposition when two or more waves of the same nature travel past a point of the medium, the resultant displacement of the medium at that point is given by the sum of the individual displacements due to the waves

temperature gradient the gradual change in temperature within a medium

timbre the combination of qualities of a sound that distinguishes it from other sounds of the same pitch and volume; it is a subjective sound quality

time-base scale a scale based on time

tone colour/quality the aspect of sound that allows the listener to identify the sound source or combination of sound sources; also called timbre

total internal reflection the sound wave is contained within the tube and is coherently reflected back and forth along the tube

transverse wave a wave whose particles oscillate about a mean position perpendicular to the line of travel of the wave

trough the negative peak of a wave

ultrasound a sound with a frequency above the human hearing range

ultrasonography using ultrasound for medical purposes

upper limit of audibility the highest frequency a human can hear (approximately 20 000 Hz)

wavefront an imaginary surface joining all points in space that are reached at the same instant by a wave propagating through a medium

wavelength the distance between corresponding points on a continuous wave (e.g. two crests or two troughs)

CHAPTER REVIEW QUESTIONS

In all these review questions use 340 m s^{-1} as the value for the speed of sound.

Remembering

1 Give three examples of mechanical waves.

2 Mechanical waves are also called elastic waves. Why do you think this is so?

3 A change in the direction of propagation of a sound wave can be caused by three processes. Name and illustrate each of the three processes.

4 What do you call waves that travel through Earth's layers that are the results of earthquakes, explosions or volcanic activity?

Understanding

5 What is the difference between a forced vibration and a free vibration?

6 The velocity of a transverse wave in a wire stretched between two points is 240 m s^{-1}. When the wire is forced to vibrate, it does so with a fundamental frequency of 480 Hz.

a What is the wavelength of the standing wave in the wire?

b What is the length of the wire between the two points?

7 What length of air column, closed at one end, will have a fundamental frequency of 256 Hz?

8 How does the slide on a trombone enable different notes to be played?

9 Pipe organs make use of both open and closed pipes. Calculate the frequencies of the first three harmonics of an organ pipe of effective length 0.50 m:

a closed at one end.

b open at both ends.

10 The length of a vibrating guitar string is 60 cm.

a What is the wavelength of the fundamental mode of vibration?

b If the speed of the wave in the guitar string is 360 m s^{-1}, what is its fundamental frequency?

c If the vibrating length of the guitar string is increased, does the fundamental frequency increase or decrease? Give a reason for your answer.

11 The velocity of the transverse waves in a guitar string is 350 m s^{-1}. When the guitar string is plucked, it vibrates with a fundamental frequency of 330 Hz.

a What is the wavelength of the standing waves in the plucked guitar string?

b What is the length of the guitar string?

c How far apart are the nodes in the standing waves in the string?

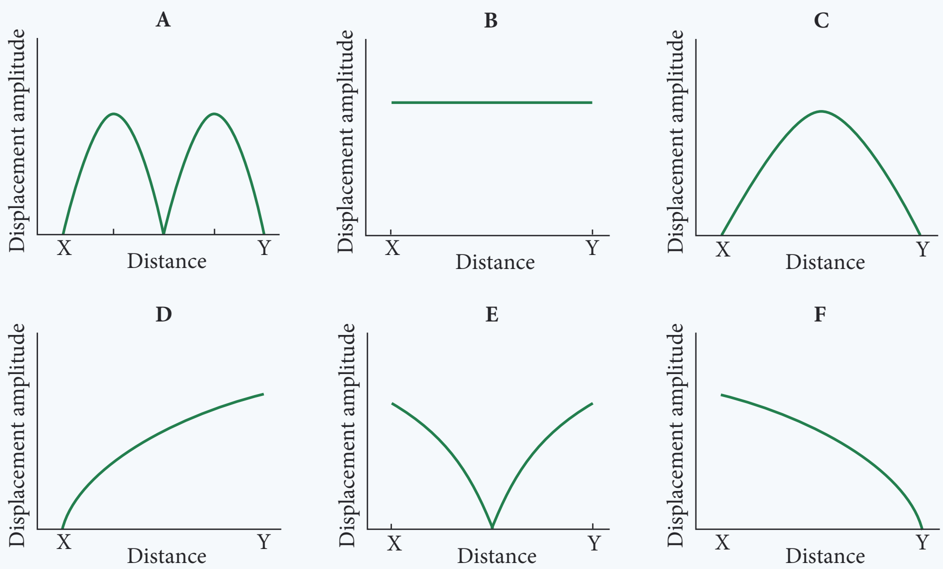
12 A clarinet acts as a tube closed at one end (the mouthpiece) and open at the other end. Vibration of the air column is produced in the mouthpiece. A particular clarinet has a fundamental frequency of 150 Hz.

a What is the wavelength of this sound?

b Which one (A–F) of the graphs of displacement amplitude plotted against distance from the open end best illustrates the amplitude of vibration of air particles at the fundamental frequency of 150 Hz? (On the distance axis, X is the open end and Y is the mouthpiece.)

c What is the length of this model of clarinet?

d What other frequencies are possible in this model?



13 The speed of sound in water is more than four times the speed of sound in air; water is much denser than air. Why does sound travel faster in water than in air?

14 Classify the following examples of waves or pulses as mechanical or electromagnetic and as transverse or longitudinal.

a P wave

b S wave

c Waves produced by the wind on the water

d Visible light from the Sun

e Waves produced by the vibrating air column of a trombone

15 A plastic tube with corrugated sides and open at both ends can produce a musical note if whirled in a horizontal circle.

a The effective length of one such tube is 1.2 m. What would be the fundamental frequency of the note produced when the tube was whirled in a horizontal circle?

b What is the frequency of the first two overtones?

c Which one of the following best describes the displacement nodes for the fundamental mode of the vibration?

A There is only one displacement node in the tube and this is at one end.

B There is only one displacement node in the tube and this is in the centre.

C There are displacement nodes at each end and none in between.

D There are two displacement nodes in the tube, at 0.40 m and 0.80 m from one end.

d There will be a range of frequencies produced by the tube but only a few frequencies are audible. Explain why. What difference would it make if a longer tube was used?

Applying

16 Figure 10.58 shows the path taken by sound waves travelling through calm air above the surface of Earth. What must the temperature profile of the air be for the sound to be refracted as shown?

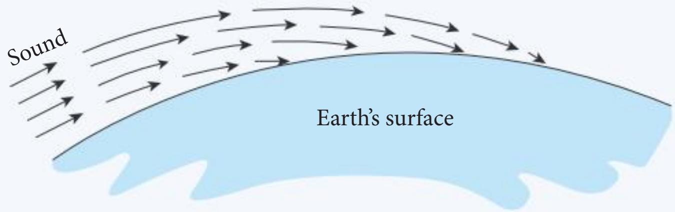


Figure 10.58

17 Estimate how long a didgeridoo is.

- a Find its fundamental wavelength.
- b What is its fundamental frequency?

18 a Draw a diagram of a didgeridoo resonating at its fundamental frequency and on the diagram show the:

- i pressure variation.
- ii displacement variation.

b Explain why the air at the closed end is a pressure antinode and a displacement node.

19 A sound wave travels from a substance in which it is traveling at 345 m s^{-1} to a second substance in which it is traveling at 1025 m s^{-1}. The wavelength of the sound in the first substance is 30 cm.

- a What is the frequency of the wave in the first substance?
- b What is the wavelength of the wave in the second substance?

Analysing

20 A student finds the speed of sound by measuring how long it takes the echo of a short clap to return from a wall (25.0 ± 0.1) m away. The time measured is (0.15 ± 0.02) s.

- a What is the speed of sound measured by the student?
- b What is the absolute uncertainty in the result?
- c What result should the student report?
- d Is the student's result consistent with the reported value for the speed of sound in air of (341 ± 5) m s^{-1}?

21 The Graham Farmer tunnel in Perth is 1.6 km in length.

- a What is the fundamental resonance frequency of the tunnel?
- b Did the designers of the tunnel need to take the resonance into consideration in their design? Explain your reasons.

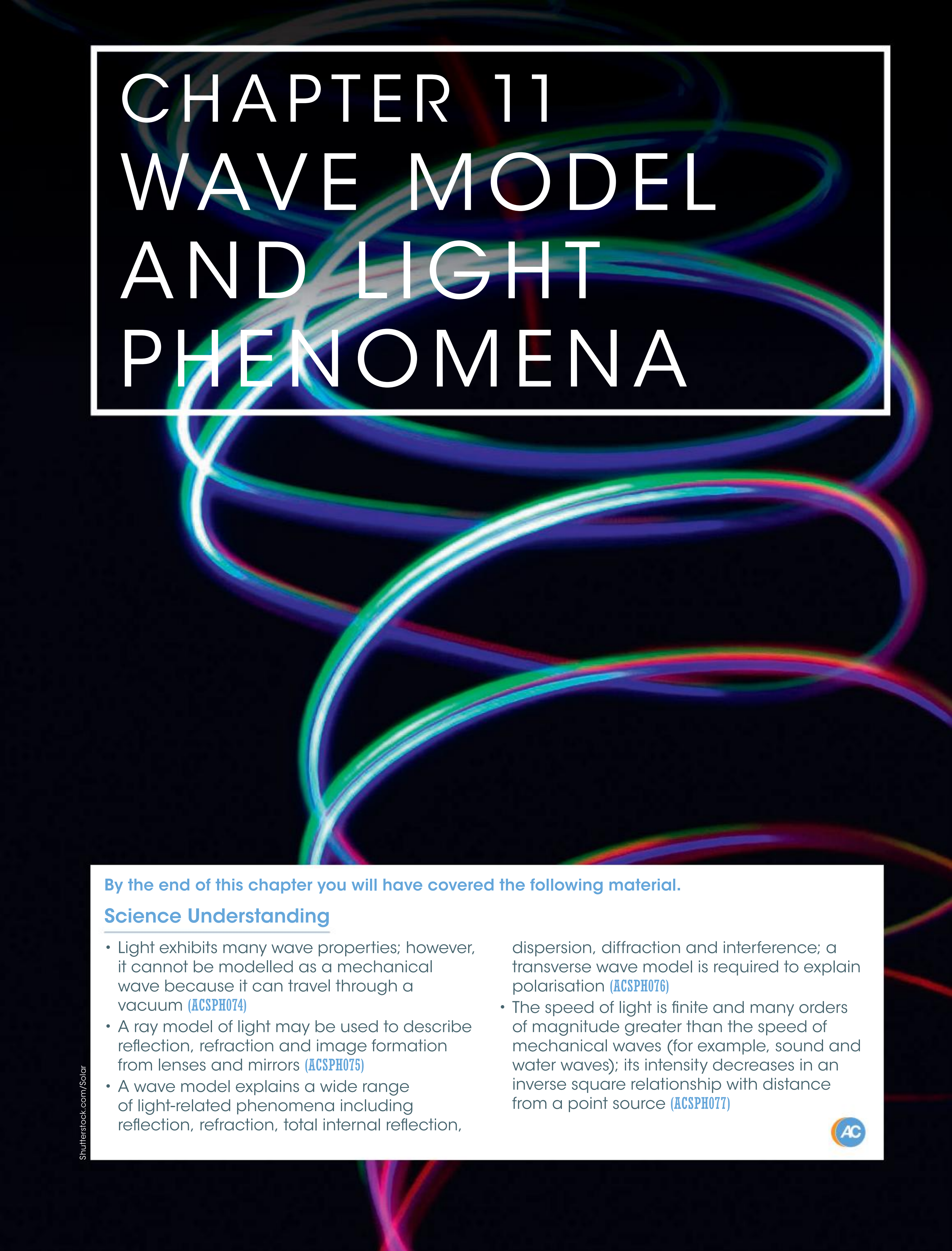
22 An adolescent person can hear sound from 20 Hz to 20 kHz. A telephone speaker can only produce sounds with wavelengths ranging from 20 to 30 cm.

- a What is the range of frequencies that the speaker can produce efficiently?
- b Why does the voice of a friend on the phone sound different from when you are talking to them face to face?

23 What are the advantages of using ultrasounds of very high frequency compared to ultrasounds of lower frequencies in medical imaging?

Reflecting

24 What role do you think the study of earthquakes has had on our understanding of the internal structures of Earth?



CHAPTER 11 WAVE MODEL AND LIGHT PHENOMENA

By the end of this chapter you will have covered the following material.

Science Understanding

- Light exhibits many wave properties; however, it cannot be modelled as a mechanical wave because it can travel through a vacuum (ACSPH074)
- A ray model of light may be used to describe reflection, refraction and image formation from lenses and mirrors (ACSPH075)
- A wave model explains a wide range of light-related phenomena including reflection, refraction, total internal reflection, dispersion, diffraction and interference; a transverse wave model is required to explain polarisation (ACSPH076)
- The speed of light is finite and many orders of magnitude greater than the speed of mechanical waves (for example, sound and water waves); its intensity decreases in an inverse square relationship with distance from a point source (ACSPH077)

Shutterstock.com/Solar

Introduction

'What is light?' This question is a somewhat difficult question to answer. Light reflects from surfaces, goes through transparent materials and produces electricity in solar cells. If we illuminate a green surface with orange light it appears black. It is the interaction between light and matter that must be explained. The better question, therefore, is 'What does light act like when it interacts with matter?'

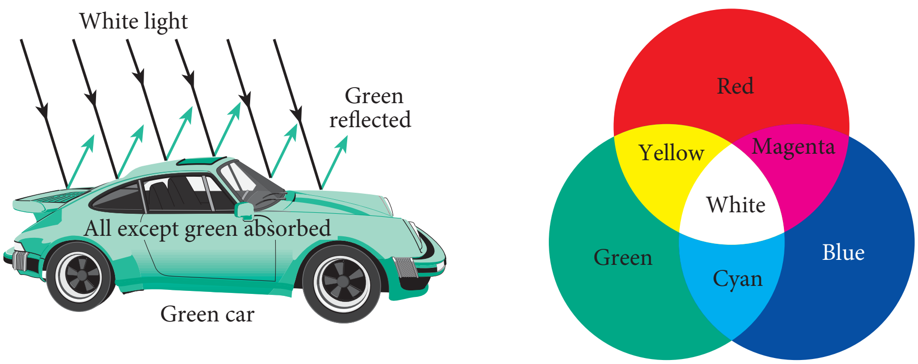


Figure 11.1 ▶ Light and matter interact to produce colour

In this chapter, we look at the way light interacts with materials to produce a range of effects: reflection, refraction, polarisation, diffraction and interference. These interactions can all be explained by a wave model. In the process we might be tempted to think that light *is* a wave. Be careful! Light can be shown to *act like* a wave in some of its interactions with matter. But there are other interactions where the wave model breaks down. It is always best to think about how and under what conditions models, such as the wave model, are useful explanatory tools.

The behaviour of light

Light exhibits many different phenomena. Without light we cannot see. Shadows form when light is blocked from view. A medium is the material through which light travels. It may be empty space or a transparent material such as water, glass or diamond. Light changes direction by bouncing off surfaces (reflection). Diffuse reflected light comes from most surfaces but images are formed by reflection from very smooth surfaces or mirrors. We use mirrors for grooming, for safe driving and, in dentistry, to inspect teeth.

Light changes direction as it is transmitted from one transparent material to another (refraction). Rainbows, twinkling stars and sparkling jewels show light of different colours. Refraction enables image formation by magnifying lenses, binoculars, refracting and reflecting telescopes and microscopes. Lenses are used in spectacles to correct vision defects such as myopia (short-sightedness) and presbyopia (long-sightedness). Light coming through small openings or around small objects forms predictable diffraction and **interference** patterns.

Sources of light

Luminous light sources (light bulbs, the Sun, lasers) produce light directly by internal processes. **Non-luminous** (the Moon, a photographer's silver umbrella) reflect light.

Models of the interactions of light with matter

In earlier years, you may have been told that light is a wave. No, it is not! You may also have been told that light is a particle. No, it is not! The wave explanation and the particle explanation are both models. A model is not the thing itself; therefore, it is possible to use different models to explain different phenomena. Light itself cannot be adequately explained: light is light, matter is matter, models of light and matter are models of light and matter.

There are three current models used to explain propagation of light and the interactions between light and matter.

- The ray model: This is the simplest model in which light is modelled as travelling in straight lines (rays) from any source. The rays change direction whenever light interacts with a material. The ray model is useful for analysing the interaction of light with large objects or surfaces, such as lenses and mirrors. We can model reflection and refraction using the ray model.
- The wave model: In the wave model, light is treated as an electromagnetic wave that propagates through vacuum or a medium with a speed that depends on the electric and magnetic properties of the medium. The wave model is useful for analysing the interaction of light with objects that are a similar in size to the wavelength of the light; for example, small apertures and obstacles, and the edges of objects. We can model interference and diffraction using the wave model.
- The photon (particle) model: Some interactions of light with matter cannot be explained by treating light solely as a wave. To understand these interactions, we model light as consisting of particles called photons, each with a characteristic energy. When light interacts with matter, an entire photon, but not part of a photon, may be absorbed (or emitted). This is the quantum model of light, as the photons are discrete quanta of light energy.

iStockphoto/teddymeeks

iStockphoto/Eloi_Omella

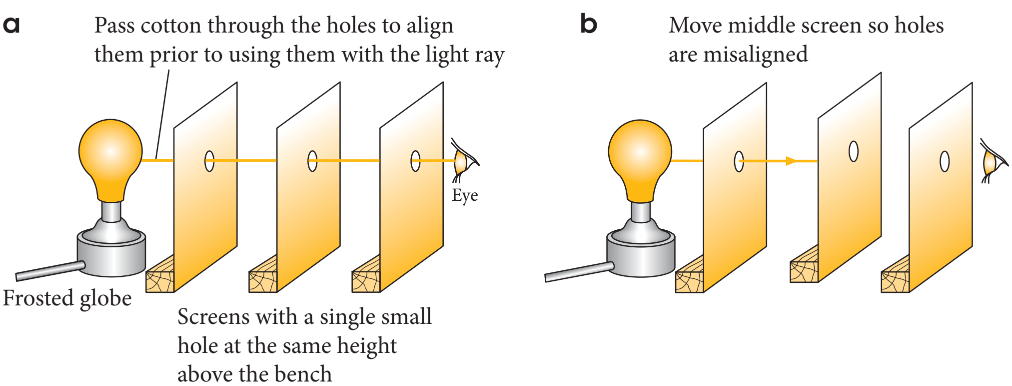
▲ **Figure 11.2**
The Sun produces its own light; the Moon reflects sunlight.

The photon model and the idea of wave–particle duality is discussed in Unit 4, Quantum theory.

In some experiments, light seems to travel as a wave, but to interact with matter as a particle. These experiments cannot be explained without the use of both the wave and particle models together. We call this need for these two apparently quite different models the wave–particle duality. In this chapter we shall be using the ray and wave models.

Light travels in straight lines

Light from a globe can be shown to travel along a straight line ray.



▲ **Figure 11.3**
a) Light travels along a straight line through three aligned holes. b) It cannot reach the eye if the holes are misaligned.

Shadows can be explained by the ray model. Light from a point source produces a sharp shadow. The shadow produced by an extended, wider light source has two sections: the **umbra** (darker part) and the **penumbra** (some light blocked). An **eclipse** of the Sun or the Moon demonstrate umbra (**total eclipse**) and penumbra (**partial eclipse**).

Never view the Sun directly. Never view a solar eclipse directly because the light from the Sun can cause permanent injury to eyes. The only safe way to view a solar eclipse is to look at an image projected onto a screen.

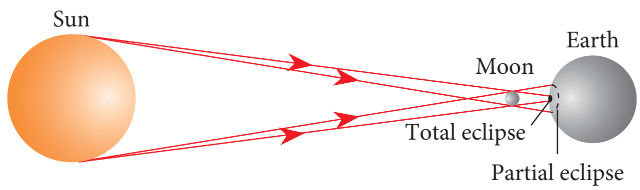


Figure 11.4 ▲
In an eclipse of the Sun, the Moon covers the Sun partially at some places (penumbra) and totally at others (umbra).

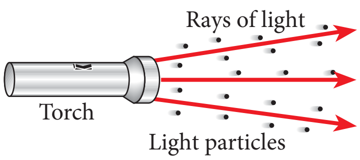
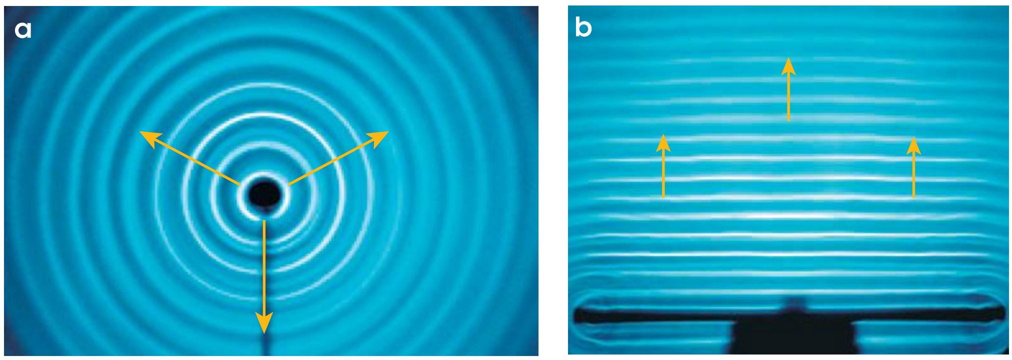


Figure 11.5 ▲
Light acts like a stream of particles, represented by a straight-line ray.

Straight-line propagation of light: ray, particle and wave models

Rays represent a stream of particles (particle model) or the direction of travel of a wave (wave model). Each direction is represented by a straight-line ray.

The light source acts like the waves produced when a stone is dropped into a pond. Wavefronts far from the source become parallel, to a good approximation.



The Picture Source

The Picture Source

Figure 11.6 ▶
Rays are at right angles to the wavefront. a) Circular waves from a point source showing radial rays at right angles to the wavefront; b) Straight waves showing rays at right angles to the wavefront

The speed of light

MEASURING THE SPEED OF LIGHT

Use diagrams and words to explain how Roemer showed, via extra-terrestrial observations that the speed of light is finite. Calculate the speed of light based on Roemer's observations.

In the early 17th century, Galileo recognised that sound travels at least ten times slower than light. He proposed an experiment: place two people with covered lighted lanterns a measured distance apart. One person reveals the light and the other responds as soon as they see the first lantern. His method was tried over a 1.6 km distance and failed. The speed was so great and the timing procedures so elementary that no useful data was recorded.

In 1676, Olaf Roemer (1644–1710) used observations of the moons of Jupiter and Earth's orbital diameter around the Sun to show that light speed was large but finite. Using his data, later calculations suggested Roemer's value for the speed of light was about $2.2 \times 10^8\,\text{m s}^{-1}$. Between 1848 and 1862, Hippolyte Fizeau (1819–96) and Léon Foucault (1819–68) used precision clocks and clockwork motors to make the first terrestrial measurements of the speed of light in air and water. They showed that light travels slower in water than in air. The medium affects the speed of light.

The current value for the speed of light in vacuum is $2.997\,924\,58 \times 10^8\,\text{m s}^{-1}$ ($3.0 \times 10^8\,\text{m s}^{-1}$). Foucault's best result, in 1862, was $2.998 \times 10^8\,\text{m s}^{-1}$.

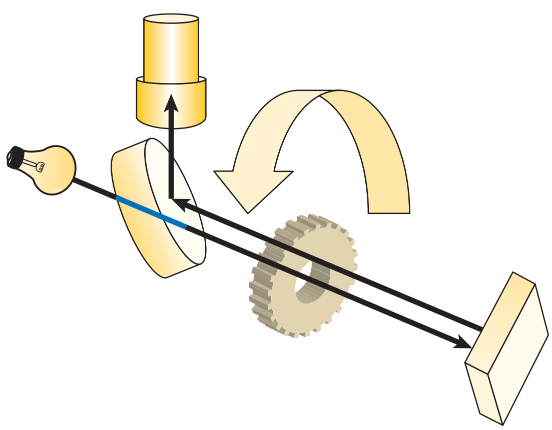


Figure 11.7 ▶
Fizeau's method for measuring the speed of light. A ray of light passes between adjacent teeth of a toothed wheel and is reflected. When the wheel is turned fast enough, the reflected light is blocked by the next tooth.

QUESTION SET 11.1

Remembering

1 Define these terms.
 a Shadow b Medium
 c Reflection d Refraction
2 What is the difference between a luminous and a non-luminous source of light?
3 Name the three models of light.
4 Is the speed of light in air slower than, equal to, or greater than its speed in water?

Understanding

5 Why is the question, 'What is light?' a misleading question? How should the question be stated? Why?

Applying

6 Draw a diagram to show partial and complete eclipse regions for a lunar eclipse.

Analysing

7 Is grass really 'black' when illuminated by 'orange' light?
8 How could you use Fizeau's experiment to measure distances to objects?

Reflecting

9 What have you learnt about:
 a light and its interactions with matter?
 b how models are used in physics?
 c technology and knowledge building?

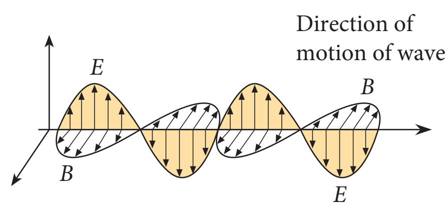
Electromagnetic wave model

By 1855, research into electromagnetism, pioneered by Michael Faraday (1791–1867), showed that every medium has two properties that affect electromagnetic radiation: **electrical permittivity, ε**, and **magnetic permeability, μ**. An electric charge oscillating in one direction causes a magnetic effect at right angles. An electromagnetic wave then speeds away in the third dimension.

In 1861, James Clerk Maxwell demonstrated that the speed of electromagnetic waves in a medium, *c*, is related to the electric permittivity and magnetic permeability of the medium:

$$c = \frac{1}{\sqrt{\mu \varepsilon}}$$

At that time, Maxwell was the first to make the direct connection between light and electromagnetic waves. Experimental confirmation came later. Between 1884 and 1889, Heinrich Hertz (1857–94) produced radio waves, known to be a form of electromagnetic wave. He measured their speed for the first time. He showed that radio waves and light both travelled at the same speed: light behaved like electromagnetic waves. Einstein used the relationship between speed of light, permittivity and permeability in his theory of relativity in 1905.



▲ **Figure 11.8** The electric and magnetic effects oscillate at right angles to each other, while the wave travels in the third dimension at right angles to both effects.

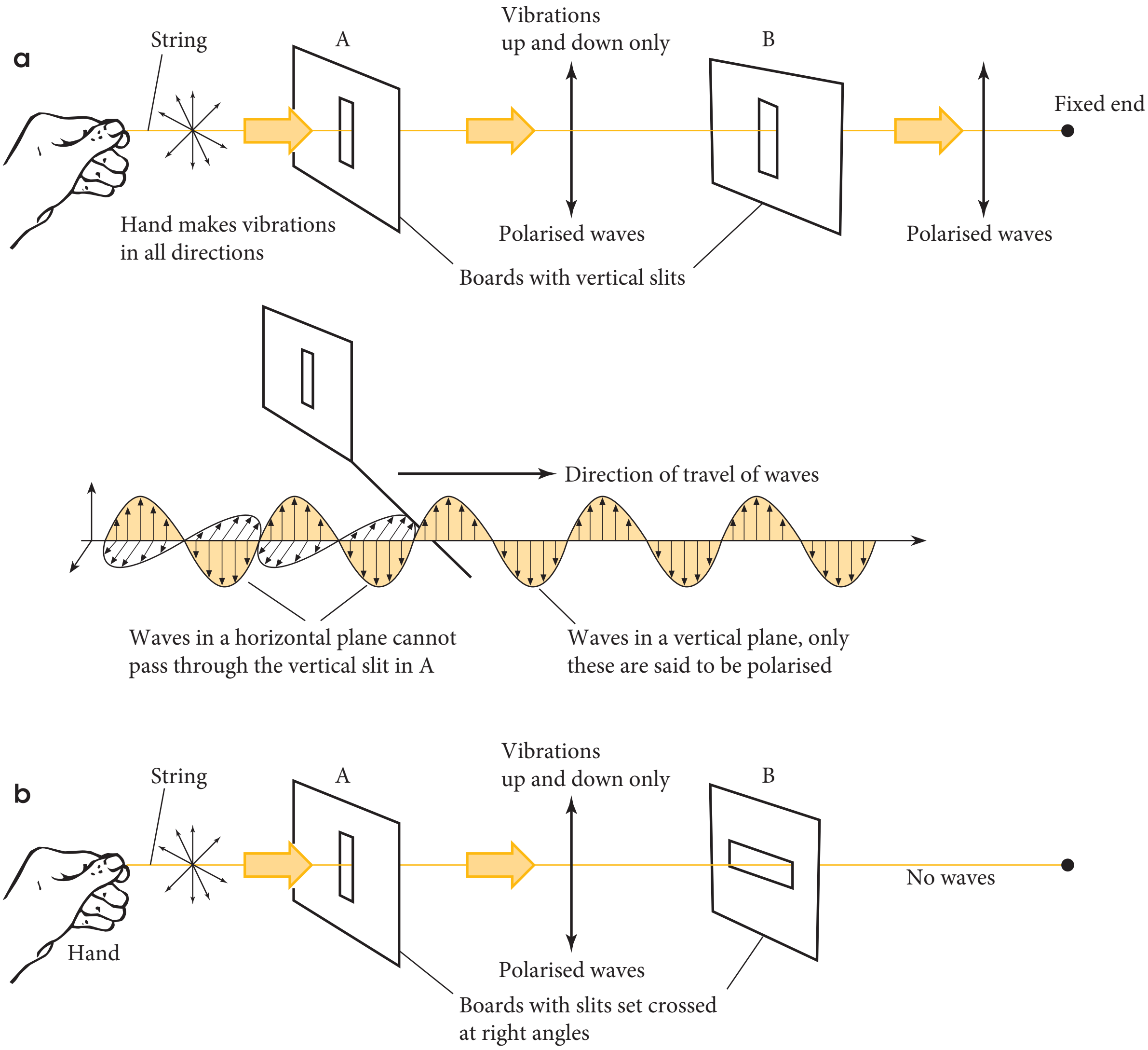
See Unit 4 Relativity for more information.

The **electromagnetic wave model** of light states that, in its interactions with matter, light *acts like* a three-dimensional transverse electromagnetic wave.

Polarisation and transverse waves

Polaroid material is made from many small, naturally polarising and transparent crystals on a polyvinyl plastic base. When two sheets are arranged so that their polarising planes are parallel, light is transmitted. However, when one sheet is rotated through 90°, no light is transmitted. **Polarisation** can be observed in natural and human-constructed environments. It shows that electromagnetic waves are transverse waves. A mechanical analogy or model can be used to explain polarisation. In Figure 11.9(a), both slits A and B are arranged vertically. If slit B is placed horizontally, the vertically polarised waves from slit A (**polariser**) cannot pass through slit B (**analyser**), as seen in Figure 11.9(b).

Longitudinal waves cannot be polarised. They travel in the same plane, so the oscillations can always go through slits A and B.



▲ Figure 11.9
A mechanical model for explaining polarisation. a) Vertically orientated transverse waves pass through slits A and B. b) Vertically polarised waves pass through slit A but cannot pass through slit B, which is perpendicular to A.

INVESTIGATION 11.1

POLARISATION AROUND YOU

Conduct an investigation into one or more aspects of polarisation in the environment. Polaroid sunglasses, lasers, 3D images and computer screens all use polarised light. Whenever light is reflected from a surface it is partially polarised. This is most obvious at **Brewster's angle**, which differs for different materials. You can see the change in intensity if you point a polarising filter to the sky at the Brewster angle and then rotate it.

Wikimedia Commons/PiccoloNamek

▲ **Figure 11.10**
The same object photographed in (left) direct and (right) polarised sunlight.

What is your aim?

What do you hope to know and be able to do as a result of your investigation? Be specific. How do you intend to show the results of your investigation? Will you prepare a webpage, a written report, a video, a slideshow or a Voki?

What will you need?

Consider the materials you will need in order to achieve your aim and make it possible to submit your final product. Think about the background information you might need to collect before you get started.

Search YouTube for videos on polarisation, birefringence and Brewster's angle for more information.

What are the risks?

Construct a table similar to the one below. Identify specific risks involved in the investigation and ways that you will manage the risks to avoid injuries or damage to equipment.

What are the risks in doing this investigation?	How can you manage these risks to stay safe?

If you intend to use a laser or bright lights, make sure you have considered all relevant safety procedures. Have your teacher check your procedures and risk assessments before beginning work.

How will you carry out your investigation?

Plan your work so that you are as efficient as possible. There are always glitches in investigations. Be aware of these and plan to respond quickly.

What results will you collect?

Consider whether you will collect quantitative or qualitative data. The type of data you collect will affect your analysis.

How will you analyse your results?

You will need to consider the best way to describe the data you collect.

What have you found?

This is where you go back to the environment you have investigated and decide what you can say about polarisation in that environment.

What do you conclude?

In this section bring together everything you have discovered.

Ideas for improvement or further investigation

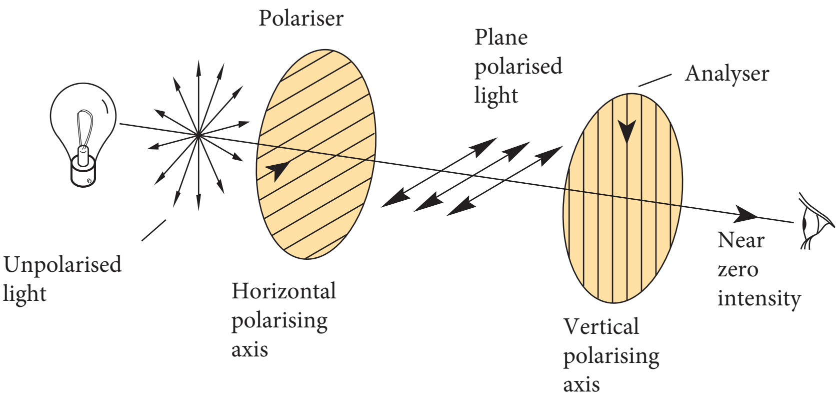
Reflect on the experience and show you are brimming with ideas for more work in this area!

WORKED EXAMPLE 11.1

Light from an electric globe passes through a polariser. An analyser is placed over the polariser, making the globe look dark. With every quarter turn, the transmitted light goes from a dark minimum to a bright maximum. Use the electromagnetic wave model of light to explain this phenomenon. (5 marks)

Answer	Logic	
The electromagnetic wave model of light states that light is transmitted as transverse electromagnetic waves.	Define the electromagnetic wave model.	1 mark
The light from the globe comes out with the electric part of the wave orientated (polarised) in all directions.	Explain the initial state of the light.	1 mark
The polariser allows most light through when the electric part of the wave is aligned with it.	Explain what the polariser does.	1 mark
At the analyser, these aligned, polarised waves are absorbed or transmitted. With each quarter turn the plane of polarisation of the analyser goes from right angles to the polariser to parallel to the polariser. The intensity goes successively from minimum transmission to maximum transmission.	Explain what the analyser does.	2 marks

A diagram can achieve the same result.



▲ Figure 11.11
Light from a globe is randomly polarised. It is polarised at the polariser and analysed at the analyser. In this case, most of the light is stopped at the analyser. When the analyser is rotated by 90°, maximum transmission occurs.

Try this yourself

Light from a particular part of the sky is quite bright when looked at through polarising sunglasses. When the sunglasses are rotated a quarter turn, the light becomes noticeably dimmer. Use the electromagnetic wave model of light to explain this phenomenon. (5 marks)

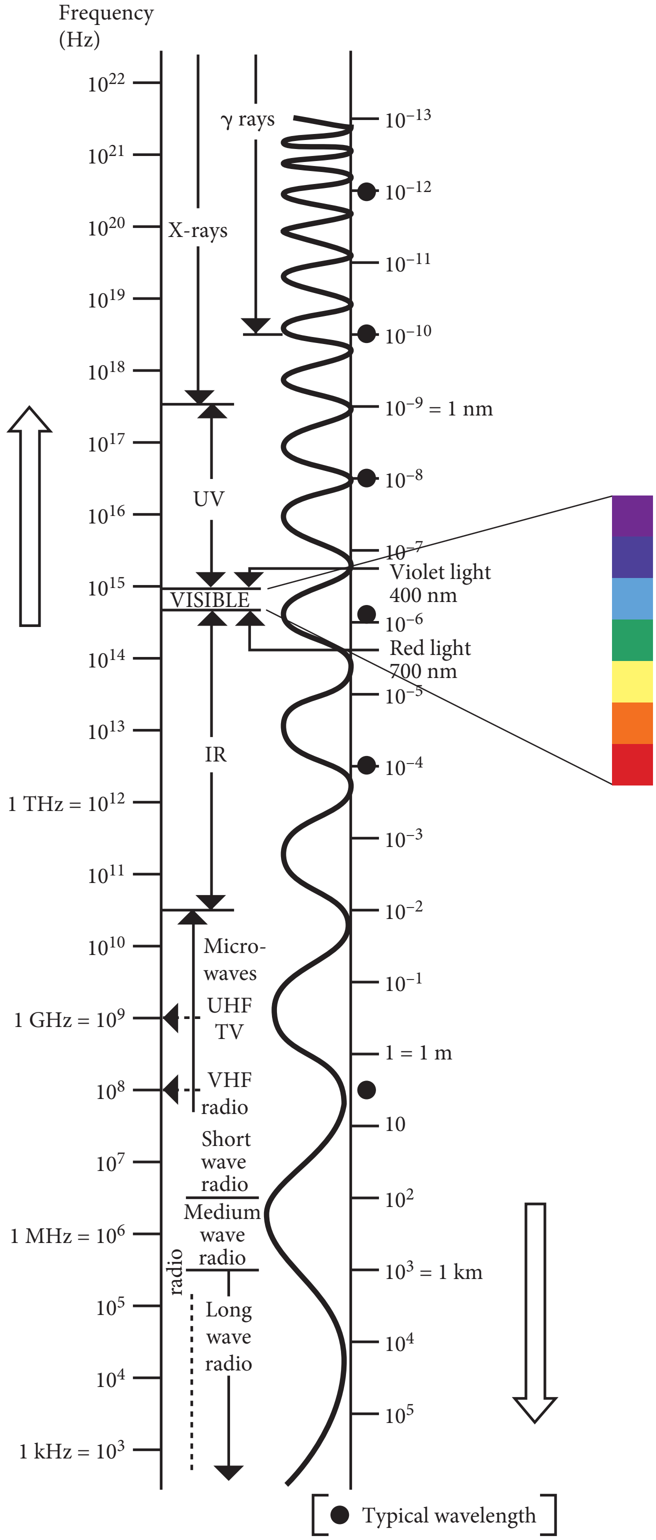
The electromagnetic spectrum

Experiments on visible light, γ rays, X-rays, ultraviolet, infrared, microwaves and radio waves enable us to find a wavelength. The frequency can be deduced from $v = f\lambda$.

Table 11.1 Wavelengths and frequencies of coloured light

Colour	Wavelength (nm)	Frequency (10^{14} Hz)
Red	750–620	4.00–4.84
Orange	620–590	4.84–5.08
Yellow	590–580	5.08–5.17
Green	580–500	5.17–6.00
Blue	500–460	6.00–6.52
Indigo	460–450	6.52–6.67
Violet	450–400	6.67–7.50

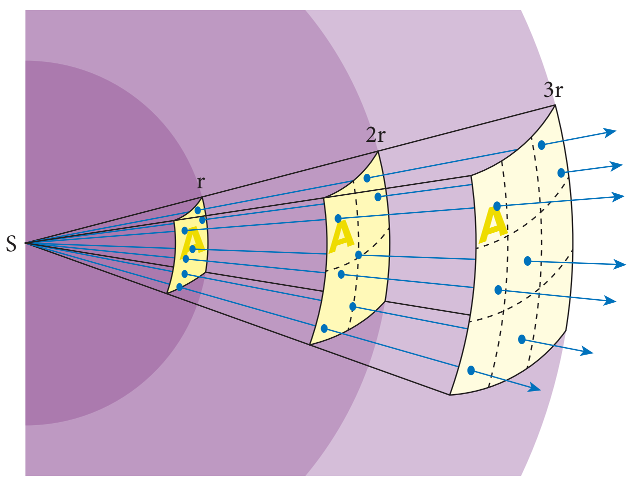
The electromagnetic spectrum comprises all of the known types of electromagnetic radiations.



▲ **Figure 11.12**
The electromagnetic spectrum goes from γ rays through visible light to radio waves.

Light intensity and the inverse square law

Light from a point source spreads uniformly into the surrounding space. The energy at the source becomes spread out over larger and larger areas as the light travels away from the source.



▲ **Figure 11.13**
Light from a point source spreads uniformly into the surrounding space.

This form of equation also works for gravitational and electrostatic effects, which are studied in Unit 3.

The energy is spread over the area of a sphere of radius r:

$$A = 4\pi r^2$$

If the source strength, S, is the energy per second being emitted, then the intensity, I, at distance, r, from the source, is the energy per second per area:

$$I = \frac{S}{4\pi r^2}$$

$$\Rightarrow I = \frac{1}{4\pi}\left(\frac{S}{r^2}\right)$$

Thus, the intensity at any point is proportional to the source strength, $I \propto S$, but inversely proportional to the square of the distance from the source, $I \propto \frac{1}{r^2}$.

The constant, $\frac{1}{4\pi}$, tells us that a sphere is involved in the calculations.

WORKED EXAMPLE 11.2

Three metres from a light source the intensity is 240 W m^{-2}. What is the intensity at 1.5 m from the source? (3 marks)

Answer	Logic	
$I \propto \frac{1}{r^2}$	Show proportionality.	1 mark
$\Rightarrow \frac{I_2}{I_1} = \left(\frac{r_1}{r_2}\right)^2$		
$\Rightarrow I_2 = \left(\frac{r_1}{r_2}\right)^2 \times I_1$		
$\Rightarrow I_2 = \left(\frac{3.0\text{m}}{1.5\text{m}}\right)^2 \times 240\ \text{W m}^{-2}$	Substitute correct values.	1 mark
$\Rightarrow I_2 = 960\ \text{W m}^{-2}$	Calculate the answer.	1 mark

Try these yourself

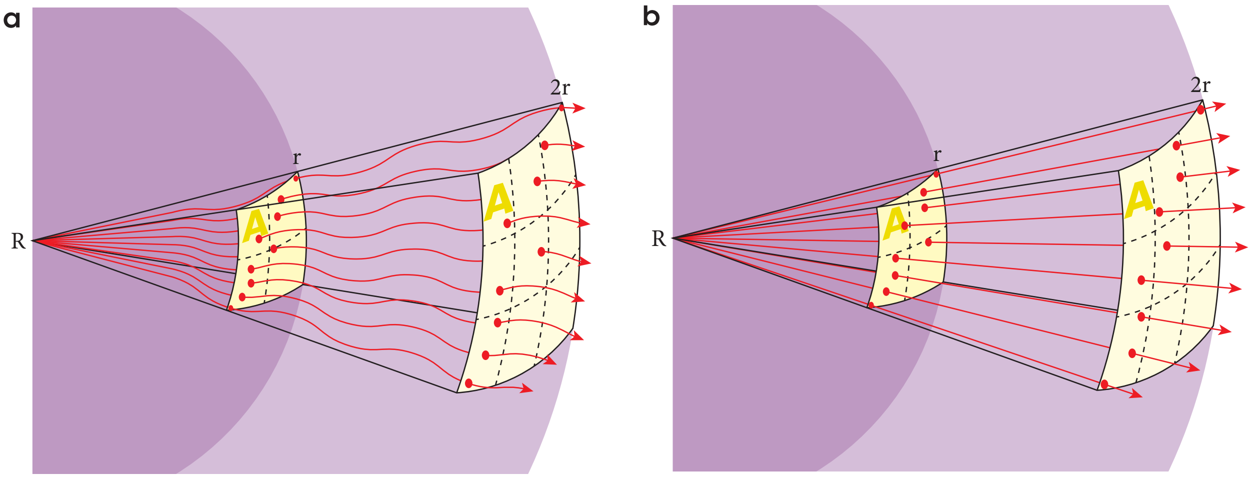
Ten metres from a light source the intensity is 400 W m^{-2}. What is the intensity at:

a 40 m? (3 marks)

b 8.0 m? (3 marks)

Wave and particle models can both be used to explain this relationship. For the wave model, a wavefront spreads into the space equally in all directions. At any particular distance from the source, the wave carries all the energy to the whole spherical area defined by the wavefront (see Figure 11.14(a)).

Similarly, for the particle model, the source sends particles uniformly in all directions. At any particular distance from the source, the combined energies of all the particles are spread onto the whole spherical surface (see Figure 11.14(b)).



▲ Figure 11.14
a) Waves carry energy uniformly from source onto a sphere of radius *r*.
b) Particles carry energy uniformly from source onto a sphere of radius *r*.

QUESTION SET 11.2

Remembering

1 What are the two properties of a medium that affect the propagation of light?

2 a State the electromagnetic wave model as it relates to light.
  b Which end of the electromagnetic spectrum has the highest frequency?
  c Which end of the electromagnetic spectrum has the longest wavelength?

3 What is polarisation? How can you demonstrate the polarisation effect?

Understanding

4 Draw and annotate a series of diagrams to explain polarisation in terms of the transverse electromagnetic wave model.

5 Explain how the wave model can explain the intensity law for point sources of light.

Applying

6 A single analyser, without a polariser, is used to show polarisation. How can this be done?

7 Copy and then complete the following table.

Colour	Wavelength (nm)	Frequency (10^{14} Hz)
	650	
		5.50
		7.00

Analysing

8 High energy γ-rays, visible light and radio waves have frequencies of order of magnitude 10^{22}, 10^{15} and 10^{3} Hz respectively. What is the order of magnitude comparison for their wavelengths?

9 Light spreads out uniformly from a 100 W source. What is the intensity of the light at:
  a 1.0 m?
  b 2.0 m?
  c 4.0 m?
  d 5.2 m?

Reflecting

10 What aspects of the electromagnetic wave model help to explain phenomena associated with light?

The ray model: reflection

When a beam of light is incident on a smooth, polished surface such as a plane mirror or a very still water surface, the rays of light forming the beam are reflected in a predictable way. This is **regular** or **specular reflection**.


Shutterstock.com/puttsk

Figure 11.15 ▶ An almost perfect reflection in a still pool of water

Most surfaces reflect incident light in all directions. This is known as **diffuse reflection** or **scattering**. For example, a sheet of paper or a painted wall appears smooth, but a microscopic examination of the surface will show it to be rough. Parallel rays incident on a rough surface are scattered in all directions. This is a particularly important property – opaque objects are visible from many different angles.

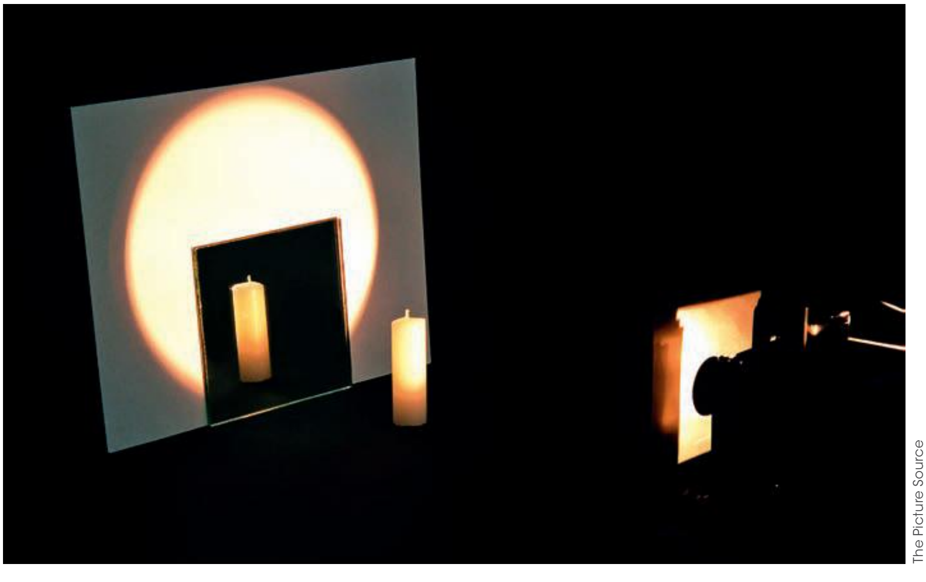

The Picture Source

Figure 11.16 ▶ Diffuse and regular reflection from a mirror in front of a white piece of paper. The mirror is mainly dark because light is not reflected to the camera, while the paper reflects light in all directions, including towards the camera. The candle and its image are recorded by diffuse reflection to the camera.

The quantity of light reflected from a surface depends on the nature of the surface and the direction of the incident light. A good quality mirror, made by backing a sheet of glass with a thin layer of metal, reflects about 95% of the incident light. **Optical fibres** totally internally reflect more than 99% of the incident light.

Law of reflection

Reflection from surfaces always follows the law of reflection. This is true for specular and diffuse reflection; however, it is easier to observe specular reflection.

1 The incident ray, the **normal** perpendicular to the surface, and the reflected ray all lie in the same flat surface (they are **coplanar**).
2 The angle between the incident ray and the normal is equal to the angle between the normal and the reflected ray: $\angle i = \angle r$.
The law of reflection applies at each point on a surface.

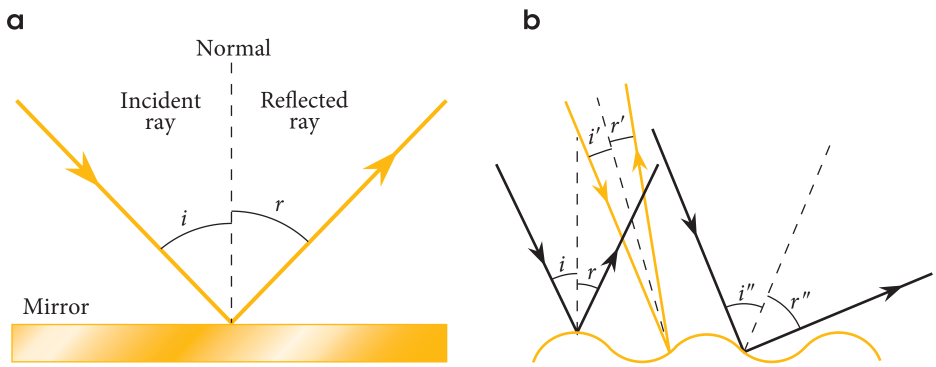


Figure 11.17
a) Specular reflection;
b) Diffuse reflection

Reflection: the ray model

When a ray of light is incident on a surface at some angle to the surface, some of that light will be reflected and some will be absorbed or transmitted. The part of the incident light that is reflected leaves the surface at some angle to the surface. The two rays define a plane. (Any two lines define a plane, just as any two points define a line.)

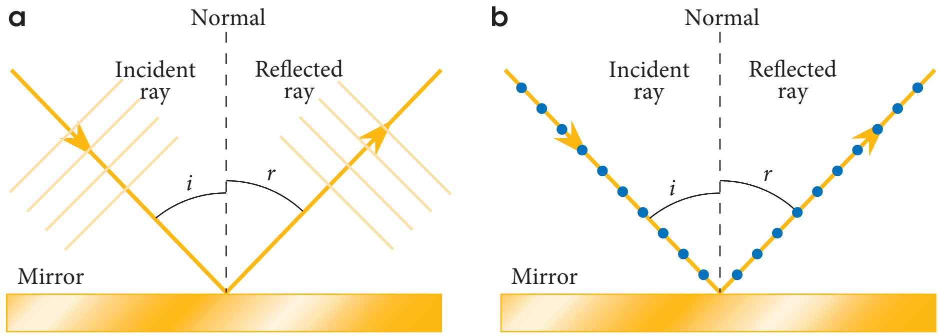
If we measure the angle between the surface and the reflected ray, and the surface and the incident ray, we find that the angles are equal. This applies at any point on a surface, as shown in Figure 11.17. Figure 11.17(a) shows the simple case of a flat surface, and hence is specular reflection. Figure 11.17(b) shows the case of a rough surface, and hence shows diffuse reflection. In both cases, the angle between the surface and the incident ray and the surface and the reflected ray are equal at the point where reflection occurs.

The angles of the rays are usually measured from the normal to the surface rather than the surface itself. The normal is drawn perpendicular to the surface, and in the plane of the two rays. (This is really just to make drawing ray diagrams on paper, which is a flat plane, easier.) The angle of the incident ray to the normal (the **angle of incidence**), i, must also be equal to the angle of the reflected ray to the normal, r, the **angle of reflection**.

> The law of reflection states that the angle of reflection is always equal to the angle of incidence.

Figure 11.18 shows this law using light represented as a wave. The directions of travel of the incoming and reflected waves are shown by the rays drawn perpendicular to the wavefronts.

The ray model is used as a model of both the wave model and the particle model. Rays are drawn perpendicular to the wavefronts (see Figure 11.18(a)) or along the line of particles (see Figure 11.18(b)).

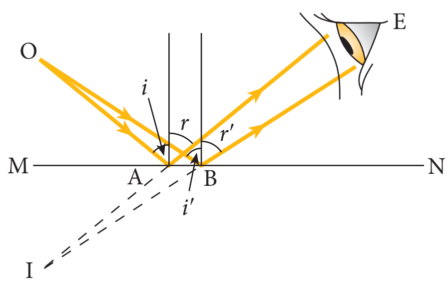


▲ **Figure 11.18**
a) Using the wave model, the ray is perpendicular to the wavefront. For simplicity, the waves interacting at the surface are not shown. b) Using the particle model, the ray is along the line of particles.

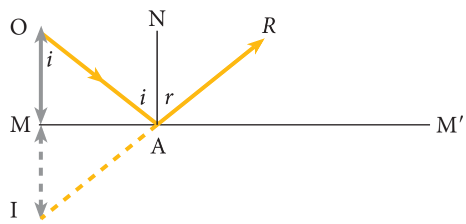
Forming images in plane mirrors

Light radiates from a **point source** in all directions. When the rays strike a plane mirror, they reflect ($\angle i = \angle r$). They appear to come from an **image** point, a **virtual image**, behind the mirror. The rays that enter our eyes must affect our retinas. Reflected rays form a **real image** in our eyes. Psychologically, we perceive a virtual image of the object to be where it is not physically present.

The virtual image appears at the same distance behind the mirror surface as the object is in front. The image is the same size as the object. The **magnification**, $M = 1$.



▲ **Figure 11.19**
Reflected rays are perceived to be coming from behind the mirror. The image is virtual because the rays do not pass through the image. A real image is formed on the retina of the eye.



▲ **Figure 11.20**
Geometric construction to show the law of reflection

Figure 11.19 shows how the image is formed and seen by an observer. Rays of light from the object, O, travel to the mirror and reflect such that the angle of incidence is equal to the angle of reflection. Two rays are shown, which reflect at points A and B. When we look towards points A and B on the mirror, it appears that light is coming from these points. If we extend the rays behind the mirror, they intersect at point I behind the mirror. Point I is the position of the image.

Figure 11.20 shows a geometric construction using the law of reflection that allows us to find the magnification and position of the image. This sort of diagram is called a **ray diagram**. We draw our object as having some actual size, such as the arrow in Figure 11.20. We draw two rays coming from the top of the object and reflecting from the mirror. The rays must obey the law of reflection as shown. We again extend the reflected rays behind the mirror to the point at which they intersect. This point corresponds to the top of the image, the arrowhead. Our object has a height equal to the distance between the mirror, M, and point O, the image has a height equal to the distance between the mirror, M, and point I. The ratio of these distances is the magnification, $M = \frac{h_i}{h_o}$. For a plane mirror, $M = \frac{h_i}{h_o} = 1$. For curved mirrors, the magnification may be greater or less than 1.

QUESTION SET 11.3

Remembering

1 Write down both statements for the law of reflection. Draw a diagram to show this law.

2 Use the ray model to illustrate diffuse reflection.

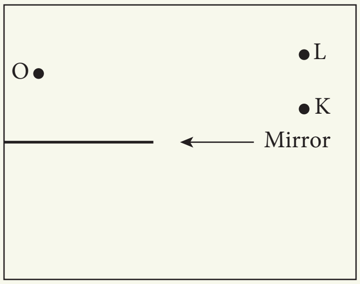
Understanding

3 A ray from a point object strikes a plane mirror at an angle of incidence of 30°. Use a carefully measured diagram to show that the object and the image are equidistant on opposite sides of the mirror.

4 How do we see a virtual image in a plane mirror? Use a ray diagram to assist in your explanation.

Applying

5 Object O is placed in front of a plane mirror as shown in Figure 11.21.



◀ Figure 11.21

a Locate the positions of the images of O, L and K.

b Construct a ray model diagram to show that an observer at L can see the image of O.

c From the point of view of an observer moving from L to K, how does the image of O move?

Analysing

6 The eyes of a 170 cm tall woman are 160 cm above the ground. She stands 0.60 m in front of a plane mirror that is mounted vertically and sees her entire image. What is the shortest mirror that can be used for such a purpose? Illustrate your answer with a diagram.

7 Prove that the image is exactly the same distance behind the mirror as the object is in front of the mirror. (MO = MI in Figure 11.20).

Reflecting

8 Reflect on what you have learnt about the way reflection affects your everyday experiences.

The ray model: refraction

When a ray of light travels from one transparent medium into another, it changes direction. This phenomenon is called refraction. The amount of refraction is mainly related to differences in the electrical properties of each medium. The electromagnetic wave changes speed depending on how well the electromagnetic wave is permitted to move through the medium.

Refraction is responsible for many strange optical effects, such as the apparent bending of a straight stick that is partly in water and partly in air.

▲ Figure 11.22
A straight stick apparently bends or breaks at the interface between air and water.

Snell's law

Refraction refers to both the change of speed and the change in direction of the light. Refraction occurs whenever light passes from one medium into another. We can characterise any medium by its refrangibility. **Refrangibility** is a measure of how much refraction will occur when light moves into a particular material from vacuum.

In all media, light travels at a slower speed than in a vacuum. This is mainly because the electrical permittivity for all other media is greater than for a vacuum. It is also colour dependent. For visible light, the magnetic permeability changes very little and has little effect.

The number used to compare refrangibilities is called the **refractive index**. The value of the refractive index of a vacuum is defined as the value 1.00. Other values express the ratio of the refrangibility of a medium to a vacuum. Relative to a vacuum, all other values are greater than 1.00 for visible light.

When light moves from one material to a second material with a similar refractive index there is very little refraction. This is the case when light moves from a vacuum to air, which has a refractive index close to 1.00. When light moves from one medium to a second medium with a very different refractive index, there is strong refraction. For example, diamond has a refractive index of 2.42 for visible light. Hence, light entering a diamond from air is slowed down a lot and bends a lot.

The values given for refractive indices are generally an average or middle value for visible light. The refractive index is actually different for different wavelengths. This variation of refractive index with colour for a particular medium is important. It gives rainbows their colours and diamonds their sparkle. Diamonds sparkle because the refractive index for blue light in diamond is larger than for red light. Hence blue light bends more than red light, and the different colours are separated.

Table 11.2 shows the relevant angles that can be measured in a refraction experiment. Here a ray strikes a glass block and refracts. The angle of incidence is i, and R is the **angle of refraction**. Both angles are measured relative to the normal.

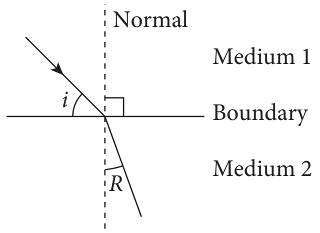


Figure 11.23 ▲ Angles are measured in a refraction experiment.

Table 11.2 Raw data and derived quantities that demonstrate Snell's law

i	R	$\frac{i}{R}$	$\sin i$	$\sin R$	$\frac{\sin i}{\sin R}$
20°	13°	1.54	0.342	0.225	1.52
29°	19°	1.53	0.485	0.326	1.49
33°	21°	1.57	0.545	0.358	1.52
43°	27°	1.59	0.682	0.454	1.50
47°	29°	1.62	0.731	0.485	1.51
55°	33°	1.67	0.819	0.545	1.50
60°	35°	1.71	0.866	0.574	1.51

WOW

High refrangibility

Most materials have refractive indices of between 1.00 and 1.60, but some materials have very high refractive indices. Some that have refractive indices greater than 2.00 are shown in the table. Diamond is a relatively common material with a very high refractive index.

However, silicon, with a refractive index of 4.01, has a refrangibility that is $\frac{4.01}{2.42} = 1.66$ times greater than that of diamond.

Zirconium dioxide	2.16
Strontium titanate	2.41
Diamond (carbon)	2.42
Mercuric sulfide	3.02
Gallium(III) arsenide	3.93
Silicon	4.01

In the refraction experiment for Table 11.2, the incident ray is moved around to larger values, starting from 20°. It is clear that the refracted ray travels along lines for which the refracted angle is greater as the incident angle gets greater. However, the relationship is not linear, and $\frac{i}{R}$ is not constant. Within experimental accuracy, the relationship that does appear to be constant is:

$$\frac{\sin i}{\sin R} = 1.50 = \text{constant}$$

The refractive index for glass relative to air, $n_{\text{glass rel air}}$, is 1.50.

Law of refraction

When a light ray refracts at a boundary between two different transparent media, it makes an angle of incidence (i) with the normal to the boundary in the first medium. The refracted ray makes an angle of refraction (R) with the normal in the second medium.

All experiments conducted along similar lines for refraction at a boundary demonstrate the two laws of refraction.

1 The incident ray, the normal and the refracted ray are coplanar.
2 Snell's law is the quantitative expression of the relationship between the incident and refracted rays:

$$\frac{\sin i}{\sin R} = \text{constant}$$

EXPERIMENT 11.1

SNELL'S LAW

Refraction can occur when a light ray travels from one medium into another. The effect depends on the angle of incidence and the relative difference in the optical properties of the media.

Aim

To determine the refractive indices of different materials

Materials

- semicircular glass block
- semicircular plastic or glass dish
- ruler
- protractor
- pencil
- black, fine point marker
- graph paper

Procedure

1 Draw a line to divide the graph paper.
2 On the semicircular glass block, draw a vertical line at the centre of the curved edge. (This is your object.)
3 Place the straight edge of the semicircular glass block along the line on the graph paper.
4 Trace the outline of the block.
5 Mark the point where the vertical black line meets the graph paper.

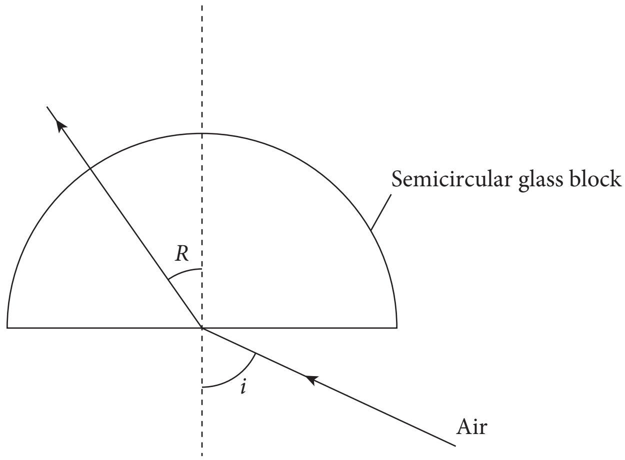


Figure 11.24 ▶
Arrangement for finding the refractive index of different materials

6 Look towards the straight edge and observe the position of the black line.
7 Use the ruler to draw the sight line towards the object.
8 Repeat this for 5 different viewing angles.
9 Remove the glass block.
10 For each observation:
    a draw lines from the object position to the point where the sight line touches the block.
    b construct the normal at the glass block.

Results

1 Record the following data in a properly constructed data table.
    a Raw data:
        i Angle of incidence, i
        ii Angle of refraction, R
    b Derived data:
        i $\frac{i}{R}$
        ii $\sin i$
        iii $\sin R$

Analysis of results

1 Plot the following graphs:
    a $\frac{i}{R}$ versus i
    b $\sin R$ versus $\sin i$

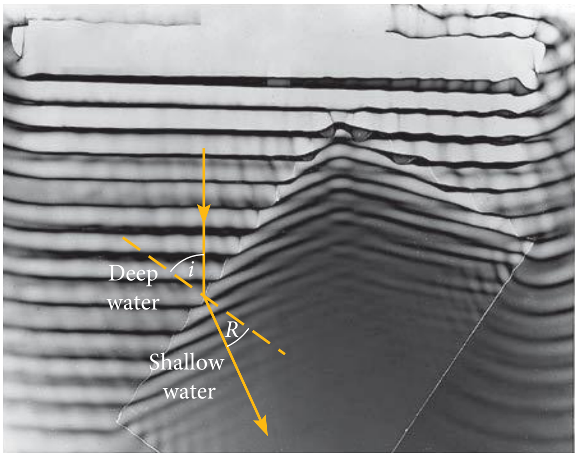
Discussion

1 Is the ratio $\frac{i}{R}$ constant for all values of i?
2 Explain how you can derive the refractive index of glass from the graph of $\sin R$ versus $\sin i$.
3 How was the reversibility of light used in this experiment to find the refractive index of glass?
4 Provide an estimate of the uncertainty in the value of the refractive index.
5 Repeat this experimental procedure and analysis with a variety of liquids in the semicircular dish.

Snell's law for waves

Figure 11.25 shows the wavefronts of waves moving from deep water into shallow water. We can see that the waves are bending towards the normal, from which we deduce that they are slowing down. We can also see that the wavelength is changing when this happens – it is becoming shorter. Figure 11.26 is a schematic diagram showing the wavefronts and interface in Figure 11.26. (Note that the schematic diagram in Figure 11.26 is rotated 90° relative to the picture shown in Figure 11.25.)

Mechanical waves demonstrate refraction. They go slower when they refract towards the normal. They go faster when they refract away from the normal. But Newton had concluded that light went faster when it refracted towards the normal. So, there was a dispute. Did light go slower or faster when it refracted towards the normal? Between 1849 and 1853, Fizeau and Foucault measured the speed of light in air and in water. The speed of light in water was less than the speed of light in air. Since light refracts towards the normal when going from air to water, light must be a mechanical wave of some sort.



▲ **Figure 11.25** Water waves in deeper water refract towards the normal when they pass into shallower water. Their speed in the shallower water is less than in the deeper water.

When light crosses the interface between two media it may slow down or speed up, depending on the difference in the optical properties of the media. This difference is encapsulated in the relative difference between the refractive indices. If the light slows down, then the ray that describes its direction of travel bends towards the normal to the interface. If the light speeds up, then it bends away from the normal. Hence, in Figure 11.24 in which light is shown bending towards the normal when it moves from medium 1 into medium 2, the light must be slowing down as it crosses the interface.

We will use the geometry of Figure 11.26 to show two useful results:

$$\frac{\sin i}{\sin R} = \frac{\lambda_1}{\lambda_2} \quad \text{and} \quad \frac{\sin i}{\sin R} = \frac{v_1}{v_2}$$

In Δ*ACP*:

$$\sin i = \frac{\lambda_1}{AC}$$

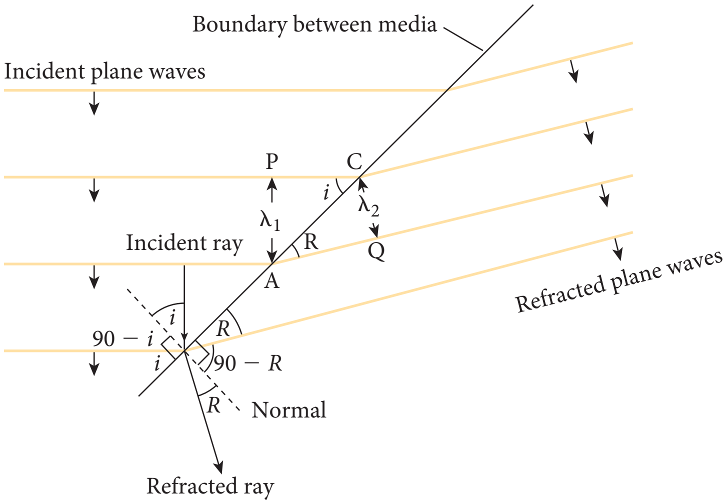
and in Δ*ACQ*:

$$\sin R = \frac{\lambda_2}{AC}$$

Thus:

$$\frac{\sin i}{\sin R} = \frac{\lambda_1/AC}{\lambda_2/AC}$$

$$\Rightarrow \frac{\sin i}{\sin R} = \frac{\lambda_1}{\lambda_2}$$



◀ **Figure 11.26** Schematic of refraction of waves. Incident rays in medium 1 and refracted rays in medium 2 are drawn at right angles to the wavefronts. Wavelengths λ_1 and λ_2 relate to medium 1 and 2 respectively.

This result enables us to show the ratio of speeds. The waves enter and leave the boundary at the same rate because the frequency of the waves does not change. From the equation, $v = f\lambda$, we can easily show that:

$$\frac{f\lambda_1}{f\lambda_2} = \frac{v_1}{v_2}$$

$$\Rightarrow \frac{\lambda_1}{\lambda_2} = \frac{v_1}{v_2}$$

This expression shows that waves slow in a medium in which the wavelength decreases and the refraction is towards the normal. The speed decrease is not a cause of the refraction. Neither is the decrease in wavelength a cause of the speed change. The speed decreases because of the interaction of the waves with the medium. For electromagnetic radiation, this means that the different materials have different electrical permittivities and magnetic permeabilities. It is the interaction of light with these properties of the materials that causes the change of speed.

Absolute and relative refractive indices

The **absolute refractive index** is a measure of the refrangibility of a medium placed in a vacuum and subjected to an incident ray of light. The same material will exhibit different absolute refractive indices depending on the incident light. In order to compare the absolute refrangibilities of different materials it is therefore necessary to use a standard wavelength. This is 589 nm, a precisely known wavelength associated with sodium atoms. Each absolute refractive index is experimentally determined. Refractive index is one of the ways in which materials can be identified. Notice that we often shorten 'absolute refractive index' to 'refractive index' when it is clear what we mean (see Table 11.3).

Table 11.3 Refractive indices of some common materials

Material	Vacuum	Air	Water	Crown glass	Flint glass	Diamond
Refractive index	1.0000	1.0003	1.33	1.52	1.65	2.42

Relative to a vacuum, air has almost the same refractive qualities. In fact, the two media do not differ until the fourth decimal place. Rounded to two decimal places, the two media are effectively the same, which is why air is usually used as a good approximation to a vacuum in cases where very high levels of accuracy are not required.

The **relative refractive index** is the comparative difference in refrangibility between two media with different absolute refractive indices. From Table 11.3, we see that water is 1.33 times and diamond is 2.42 times more refractive than air. If a diamond were to be placed in water, its refrangibility would be reduced – it would only be $\frac{2.42}{1.33} = 1.82$ times as refractive as it is in air: ($n_{\text{diamond rel water}} = 1.82$). This is still highly refractive compared with various types of glass.

If a piece of sand, $n_{\text{sand}} = 1.46$, is placed in oleic acid of a similar colour, $n_{\text{oleic acid}} = 1.46$, it cannot be distinguished optically from the oleic acid because their refractive indices are the same:

$$n_{\text{sand rel oleic acid}} = \frac{1.46}{1.46} = 1.00$$

The relative difference in refractive index between two media does not have to be very much for the effect to be noticed. Hot air has a slightly lower refractive index than cold air, of the order of 0.1%, yet this difference is why we can see a shimmering heat haze above a fire or near the road surface on a hot day. We shall see that quite small differences in the refractive indices of different types of glass enables light to travel very efficiently down optical fibres.

Refraction towards and away from the normal

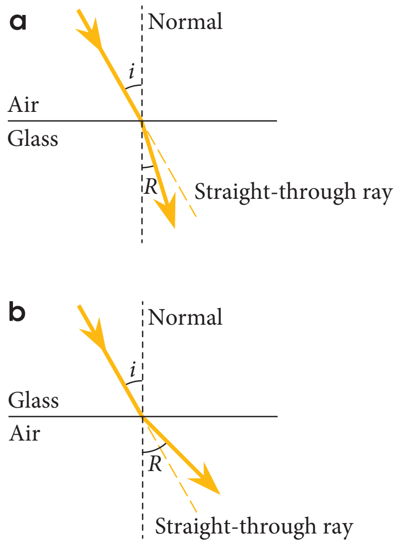
Relative refractive indices can be greater than or less than 1.00. If the relative refractive index is greater than 1.00, then the refracted ray deviates from the straight-through ray towards the normal. Figure 11.27(a) shows a ray refracting towards the normal as it travels from air to glass. The relative refractive index is:

$$n_{\text{glass rel air}} = \frac{1.33}{1.00} = 1.33$$

If the relative refractive index is less than 1.00, then the refracted ray deviates away from the normal.

If the rays are reversed and travel from glass to air (Figure 11.27(b)), the relative refractive index becomes less than 1.00, and refraction away from the normal occurs:

$$n_{\text{air rel glass}} = \frac{1.00}{1.33} = 0.75$$



▲ **Figure 11.27**
a) Refraction at the air–glass boundary is towards the normal, b) but away from the normal when the rays are reversed (glass–air).

In defining refrangibility in terms of relative refractive indices we have used the general form of Snell's law:

$$\frac{\sin i}{\sin R} = \frac{n_2}{n_1}$$

Combining this with the expression of Snell's law for waves, we can summarise Snell's law as follows:

Snell's law: $\frac{\sin i}{\sin R} = \frac{\lambda_i}{\lambda_R} = \frac{v_i}{v_R} = \frac{n_R}{n_i} = \text{constant}$

Total internal reflection

At every boundary between media, reflection always occurs. Mostly, so does refraction. However, for refraction away from the normal, there comes an angle of incidence for which no refraction occurs. At angles of incidence greater than this **critical angle**, the ray is totally reflected back into the medium in which it was travelling when it reached the boundary. At the critical angle of incidence, the refracted angle is 90°.

Thus:

$$\frac{\sin i}{\sin R} = \frac{n_2}{n_1}$$

but, at the critical angle, $i = i_c$ and $R = 90°$.

As long as $n_2 < n_1$, a value of the critical angle, i_c, can be calculated.

WORKED EXAMPLE 11.3

Light travelling in water (n_w = 1.33) strikes the interface with flint glass (n_g = 1.65) at an angle of 36° to the normal.

a What is the angle of refraction in the glass? (4 marks)

b What is the critical angle for light travelling from flint glass into water? (4 marks)

Answers	Logic	
a $\frac{\sin i}{\sin R} = \frac{n_g}{n_w}$	Use the correct formula.	1 mark
$\frac{\sin R}{\sin i} = \frac{n_w}{n_g}$		
$\sin R = \frac{n_w \sin i}{n_g}$	Rearrange the formula to solve for *R*.	1 mark
$\Rightarrow \sin R = \frac{1.33 \sin 36°}{1.65}$	Substitute the correct values into formula.	1 mark
$\Rightarrow R = 28.3°$	Calculate the correct answer.	1 mark
(Note that the light is deviated by 7.7° towards the normal.)		
b The critical angle, i_c, occurs when the refracted angle reaches 90°.		
$\frac{\sin i}{\sin R} = \frac{n_g}{n_w}$	Use the correct formula.	1 mark
$\sin i = \frac{n_g \sin R}{n_w}$		
$\sin i_c = \frac{1.33 \sin 90°}{1.65}$	Substitute the correct values into formula.	1 mark
$\sin i_c = \frac{1.33}{1.65}$	Rearrange to solve for i_c.	1 mark
$\Rightarrow \sin i_c = 0.806$		
$\Rightarrow i_c = 53.7°$	Calculate the correct answer.	1 mark

Try these yourself

1 A ray of light travelling in air enters a surface of a diamond, refractive index 2.42, at an angle of 23°. What is the angle of refraction in the diamond? (4 marks)

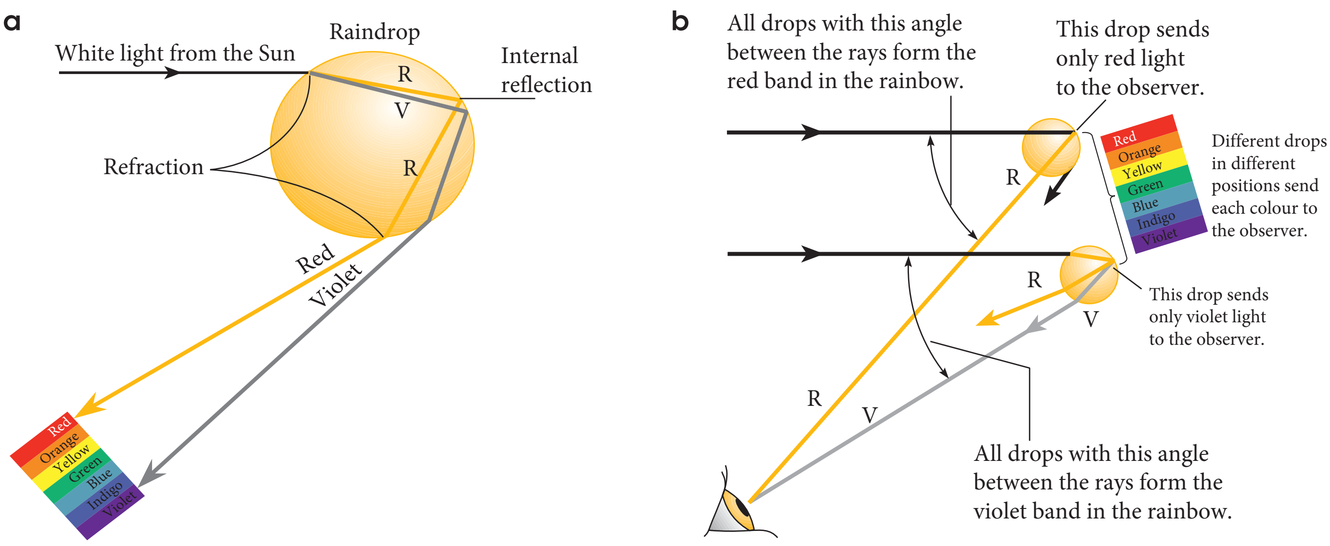
2 A ray of light enters a medium of refractive index 1.36 and refracts as it enters a new medium with $n = 1.29$. The angle of refraction is 25°. What is the angle of incidence for this ray? (4 marks)

3 What is the critical angle for light travelling out of diamond ($n = 2.42$) into water ($n = 1.33$)? (4 marks)

Colour dispersion

Different colours of light refract by different amounts. This effect is called **chromatic dispersion**. Red light refracts least, blue light refracts most: $n_{red} < n_{blue}$. Rainbows are a result of colour dispersion. Colours disperse in every drop. Altogether, the raindrops produce different colours at slightly different angles.

The composition of glass can be changed in order to change its sparkling qualities. The refractive index changes markedly between types of glass, and for different colours within a particular type of glass.



▲ **Figure 11.28** A rainbow is formed by the addition of the dispersed light in all the raindrops.

Table 11.4 Refractive indices for different-coloured light in two types of glass

Colour	Crown glass	Flint glass
Red	1.514	1.638
Yellow	1.520	1.650
Blue	1.527	1.664
Violet	1.533	1.675

WOW

Twinkle, twinkle little star

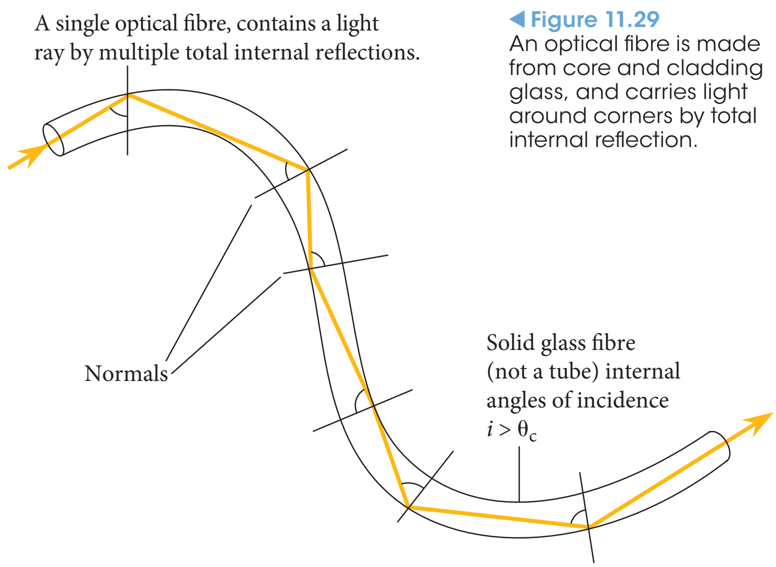
Stars twinkle as a result of refractive index differences in the air. As the light travels through the atmosphere, it is refracted to your eyes. If there were no air, the light would always strike your eye at the same place. However, due to the differences in refractive index along the path, some of these rays wander away from that point. This effect is often very striking, especially for stars low in the sky, because the light comes through the maximum amount of atmosphere. Sirius can be seen flashing from red to blue as the chromatic dispersion is added to the twinkling effect. Similarly, Arcturus flashes red to green as it rises around Easter time in southern Australia.

Optical fibres

An optical fibre is made of a glass **core** that has a refractive index slightly higher than the surrounding glass **cladding**.

In this way, light that spreads to the boundary is mostly constrained to travel down the core by total internal reflection. The energy loss per reflection is about 500 times less than for a highly polished mirror surface. Optical fibres are highly flexible so that the light can be readily carried around corners. Every bend causes an increase in energy loss, but this is still much better than for ordinary mirror surfaces.

Any light source, including the monochromatic sources used for optical fibre communications, has a spread of 'colours'. This means that chromatic dispersion affects the spread of light travelling down the optical fibre.



◀ **Figure 11.29** An optical fibre is made from core and cladding glass, and carries light around corners by total internal reflection.

QUESTION SET 11.4

Remembering

1 Draw a diagram to illustrate the:
  a angle of incidence.
  b angle of refraction.
  c normal.
2 State Snell's law.
3 Define 'absolute refractive index'. Why is it necessary to use a specific wavelength of light in the definition?

Understanding

4 Use a diagram to show how Snell's law is obtained for waves in terms of their:
  a wavelength.
  b speed.
5 Absolute refractive index is really an example of a relative refractive index. Explain.
6 Draw and label an optical fibre to show core, cladding, and total internal reflection at the core–cladding boundary.

Applying

7 Light travelling in air (n_{air} = 1.00) enters a glass block (n_{glass} = 1.49) at an angle of incidence of 30°.
  a What is the angle of refraction in the glass?
  b The glass block is now immersed in oil (n_{oil} = 1.28). Does the angle of refraction get larger or smaller? Support your answer with calculations.
8 A ray of light of wavelength 981 nm travels in air at a speed of $3.00 \times 10^8\,m\,s^{-1}$. It meets a transparent medium of refractive index 1.39 at an angle of 25°.
  a Calculate the frequency of the light in:
    i air.
    ii the transparent medium.
  b Calculate the speed of the light in the transparent medium.
  c What is the angle of refraction as the light passes into the transparent medium?

Analysing

9 An underwater diver looking up at the surface of still water (n_{water} = 1.33) at an angle of 45° can see the horizon. What would the diver see if he looked up at the water surface at an angle of 60°?
10 Red laser light is incident at the core from air and travels in an optical fibre.
  a What is the critical angle at the core–cladding boundary?
  b What is the maximum angle of refraction at the air–core boundary to ensure all the red light is transmitted down the fibre?

  For red light: n_{core} = 1.495, n_{clad} = 1.480.

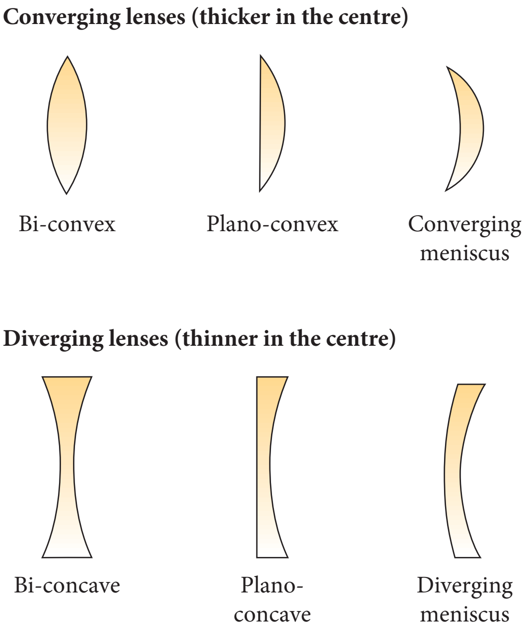
Reflecting

11 What have you learnt about the importance of language and definitions in physics? Support your answer with reference to the definition of refractive index.

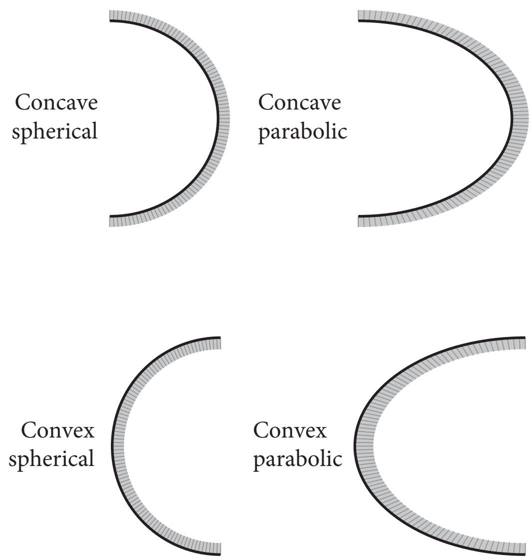
Images formed by refraction and reflection

Lenses are shaped, transparent objects. They may be convex, like the lens in the eye, or concave. A **converging (convex) lens** is thicker at the centre and thinner at the edges. **Diverging (concave) lenses** are thicker at the edges than at the centre. Lenses can produce real or virtual images by refraction.

Curved mirrors can be concave or convex. They can produce real or virtual images by reflection.



▲ **Figure 11.30**
Types of converging and diverging lenses

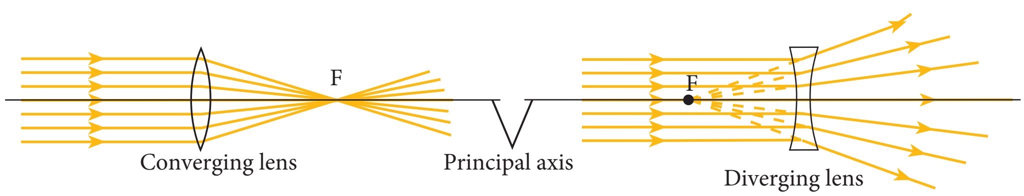


▲ **Figure 11.31**
Types of converging and diverging mirrors

Images formed by refraction in lenses

Figure 11.32 shows how rays of light are refracted in converging (convex) and diverging (concave) lenses. The **principal axis** is drawn perpendicular to the lens and passing through its centre. The converging lens refracts parallel incoming rays towards the principal axis. The rays converge and cross at the **focal point (F)**. The focal point is on the opposite side of the lens to the source of the light. The converging lens forms a real image on the opposite side of the lens to the object. A real image is one for which the light is *actually* coming from the point it appears to be coming from. A screen placed at this point will have an image on it, and a photodetector placed at this point will detect light.

The diverging lens refracts light so that the parallel rays diverge, and do not cross each other on the far side of the lens from the source. However, if we trace the rays backwards from the right-hand side of the lens we see that they appear to originate from a focal point (F) on the same side of the lens as the object. A diverging lens forms a virtual image, which is an image



▲ **Figure 11.32**
Rays that are parallel to the principal axis refract to a real focus (*F*) in a converging lens, and a virtual focus in a diverging lens.

formed at a position where the light rays do not actually converge. A photodetector placed at this point will not detect light, nor will a screen show an image here. This is similar to the way a plane mirror forms a virtual image. The image still exists, and can be seen and photographed. It is just not due to light coming from the image position; rather the light making the image is being collected by the lens to form a real image in the camera.

Our eyes contain a convex lens that converges the rays of light incident on our eye to a focal point. If the lens of the eye forms an image in front of or behind the retina rather than right on the light-sensitive cells of the retina, the result is blurry vision.

When the image is formed in front of the retina, this is usually because the lens is too strongly converging or the eyeball is too long. The result is that the person can focus on objects up close, but not far away. This is called myopia (short- or near-sightedness). When the image is formed behind the retina, usually because the lens is not converging enough, the person can focus on objects far away but not up close. This is a called hyperopia (long-sightedness). When it is due to aging of the eyes and the muscles that control the shape of the lens, it is called presbyopia (literally 'old-age vision').

Ray tracing and the wave model

Light that leaves a point on an object passes through the convex lens and is recombined on the other side to form a real image.

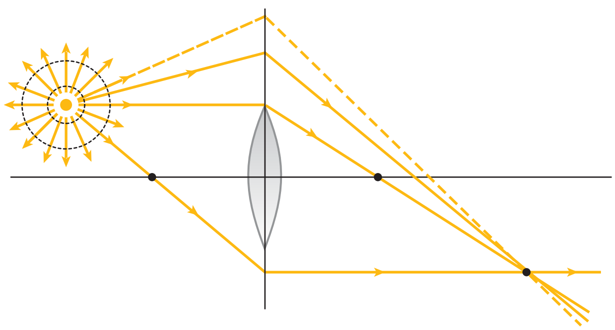


Figure 11.33 ▶ Some of the millions of rays that leave a point on an object. These rays represent the waves emanating from the point.

The waves that leave the point all radiate uniformly into the surrounding space. Some parts of the wavefronts reach the lens and are refracted. The rays shown are representations of the direction of propagation of the wavefronts. There are, therefore, many rays that leave the point and reach the lens. All these rays are recombined beyond the lens to form a real image.

Geometry of the lens image-forming system

Figure 11.34 shows the geometry of a convex lens system.

The principal axis, or axis, is a line that passes through the centre of the lens at right angles to the plane in which the lens stands. The **optical centre** is the point at which these two axes cross. The focal point, or focus, F_2, nearest the object is at the **focal length**, f. The focal point, F_1, is placed symmetrically on the opposite side of the lens.

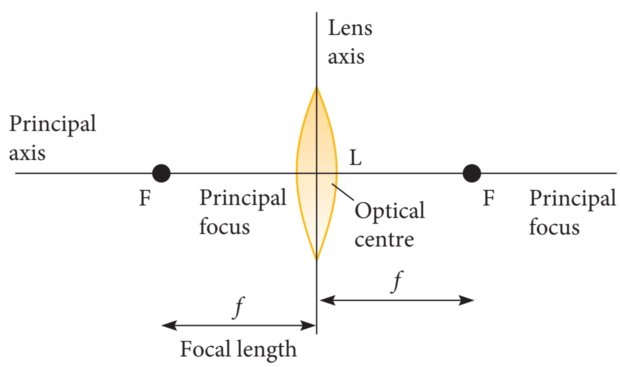


Figure 11.34 ▶ Geometry of the convex lens image-forming system showing lens axis, principal axis, optical centre and two foci

Paraxial assumptions

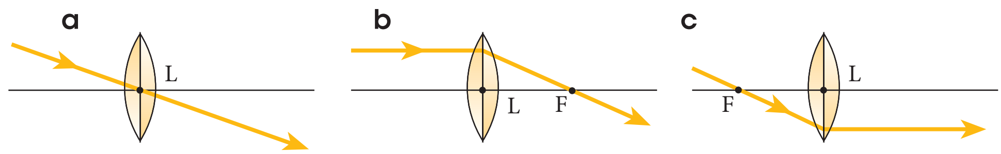
If you want to construct an exact image by ray tracing, you will need to draw the exact lens surfaces and follow the rays exactly into, through and out the other side of the lens. You will also need to take into account the refractive index of the lens. However, it is possible to draw reasonably accurate (first approximation) ray tracing diagrams. To do this, you need the **paraxial assumptions**. (These also apply to curved mirrors, which we consider in a later section.)

1. The rays striking the lens or curved mirror are not too far away from the principal axis.
2. The lens or curved mirror is small and thin so that it can be replaced in the diagram with a straight line. (However, we always draw a small lens or curved mirror around the centre to remind us of what we are doing!)
3. When a ray strikes the straight line that represents a lens or curved mirror, it refracts or reflects respectively as though the line were the lens or curved mirror.

The paraxial assumptions only apply for thin lenses and curved mirrors. Highly accurate image re-construction requires more sophisticated assumptions.

There are millions of rays leaving a point on the object because the wavefront is continuous. Of these, three rays are useful to help trace the rays to the image in a convex lens.

- A ray directed through the centre of the lens travels to the image unrefracted (Figure 11.36(a)).
- A ray parallel to the axis refracts through the lens and passes through the principal focus on the other side (Figure 11.36(b)).
- A ray through the focus nearer the object refracts at the lens and travels parallel to the axis (Figure 11.36(c)).



A ray through the optical centre L is undeviated.

A ray parallel to the principal axis is refracted (by the lens) to pass through F.

A ray arriving through F is refracted parallel to the principal axis.

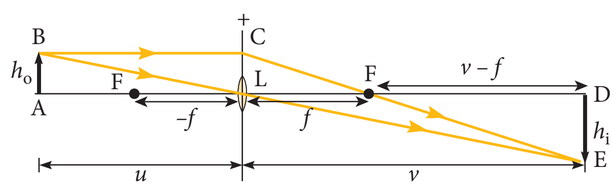
◀ **Figure 11.35** Ray paths for three significant rays can be used to construct the image in a convex lens: a) the ray directed at the optical centre, b) the ray parallel to the principal axis and c) the ray through the focus nearer the object.

The thin lens equation

In general, an image will be inverted or upright with respect to the object, and it will be a different size from the object (enlarged or diminished). The magnification, M, is the ratio of image height, h_i, to object height, h_o.

$$M = \frac{\text{height of image}}{\text{height of object}} = \frac{h_i}{h_o}$$

Figure 11.36 enables us to deduce a mathematical relationship, the **thin lens equation**, which connects the position of the object, u, the position of the image, v, and the focal length, f.



◀ **Figure 11.36** Ray tracing diagram to find an image in a convex lens

The thin lens equation is deduced from similar triangles.
In ΔABL and ΔDEL:

$$\frac{h_i}{h_o} = \frac{v}{u}$$

In ΔEDF and ΔFLC:

$$\frac{h_i}{h_o} = \frac{v}{u}$$

$$\frac{h_i}{h_o} = \frac{v - f}{f} \quad (CL = h_o,\ DF = v - f)$$

$$\Rightarrow \frac{v}{u} = \frac{v - f}{f}$$

$$\Rightarrow vf = uv - uf$$

Divide by uvf (it's just a trick!):

$$\frac{vf}{uvf} = \frac{uv}{uvf} - \frac{uf}{uvf}$$

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

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

Thin lens equation

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

$$M = \frac{-h_i}{+h_o} = -\frac{v}{u}$$

- Real images: M is negative.
- Virtual images: M is positive.

EXPERIMENT 11.2

IMAGES IN A CONVEX LENS

A small light forms a real image on a screen when it shines through a convex lens. The light must be further from the lens than the focal distance. The image may be enlarged, equal to, or diminished relative to the actual light. This depends on the distance of the object from the lens.

Aim

To apply the thin lens equation to data to find the focal length of a convex lens:

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

where u = distance from lens to object, v = distance from lens to image and f = distance from lens to focal point

Materials

- small convex lens and lens holder
- vertical, white screen
- small, bright light source (a single filament globe is best) in a globe holder
- 9-volt DC battery pack

- rubber band
- 2 × metre rulers
- tape
- darkened space

Procedure

1 Tape the rulers end to end along a table.
2 Put the lens in the lens holder and place it where the rulers join.
3 Use the rubber band to attach the globe to the 9-volt battery pack.
4 Adjust the height so the filament of the globe is at the same height as the centre of the lens.
5 Move the globe to one end of the rulers.
6 Place the screen on the opposite side of the lens to the globe.
7 With the globe as far from the lens as possible, move the screen so as to estimate the focal length of the lens. Record this length.
8 Move lens and screen until a clear, focused image appears on the screen. Do this for at least three positions of the object.

Results

In a properly constructed data table:

1 record object and image distances from the lens.
2 show the computed value for the focal length of the lens, to an appropriate number of significant figures.

Analysis of results

1 What was the average focal length for the lens?
2 What is a reasonable estimate of the uncertainties in the distance measurements and in the derived focal length?

Discussion

1 Compare the computed average focal length with your experimental estimate of the focal distance.
2 Explain the basis for the direct experimental estimate of the focal length.
3 Justify quantitative estimates of the uncertainty in the data.

Taking it further

Measure directly the magnification of the object and compare with the magnification computed from object and image distance.

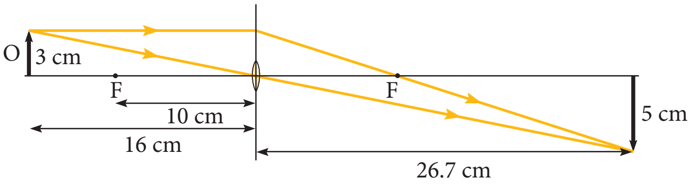
WORKED EXAMPLE 11.4

An object 3.0 cm high is placed 16.0 cm in front of a converging lens of focal length 10.0 cm.

Use an accurately drawn ray tracing diagram to find these properties of the image. (6 marks)

a Its position (1 mark)
b Its nature (1 mark)
c The size of the image (1 mark)
d The magnification (2 marks)

Answers



▲ Figure 11.37
Ray tracing diagram showing a real, inverted image formed by a convex lens.

Logic

Draw the axes correctly, label the foci and mark in the object correctly. Use a consistent scale.	2 marks
Draw two useful rays to and from the mirror.	2 marks
Locate the image correctly. It must be located correctly, both horizontally and vertically.	2 marks

By the accurately drawn ray diagram in Figure 11.37:

a The image is 26.7 cm from lens on opposite side from the object.		1 mark
b The image is real.		1 mark
c Size: 5.0 cm		1 mark
d $M = \frac{-h_i}{h_o}$	Use correct formula.	1 mark
$M = \frac{-5\text{ cm}}{3\text{ cm}} = -\frac{26.7\text{ cm}}{16\text{ cm}} = -1.7$	Substitute correct values from diagram.	1 mark

Try these yourself

1 An object 6.0 cm high is placed in front of a converging lens of focal length 5.0 cm. The object is at double the focal distance, that is, 10.0 cm, from the lens.

Use an accurately drawn ray tracing diagram to find these properties of the image.

- **a** Its position (2 marks)
- **b** Its nature (1 mark)
- **c** The size of the image (1 mark)
- **d** The magnification (2 marks)

2 An object 4.0 cm high is placed in front of a converging lens of focal length 6.0 cm. The object is at a position that is less than double the focal distance, that is, 10.0 cm from the lens.

Use an accurately drawn ray tracing diagram to find these properties of the image.

- **a** Its position (2 marks)
- **b** Its nature (1 mark)
- **c** The size of the image (1 mark)
- **d** The magnification (2 marks)

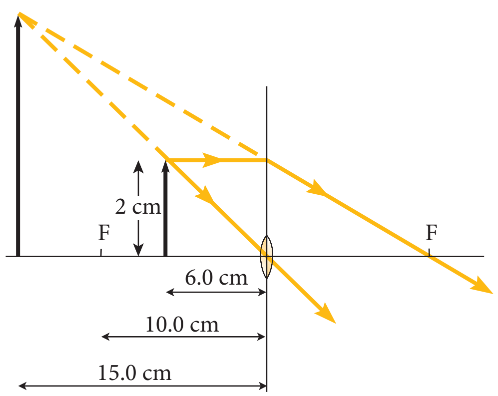
WORKED EXAMPLE 11.5

A 2.0 cm high object is 6.0 cm from a lens of focal length 10.0 cm. (The object is inside the focal length). Use the lens formula to find these properties of the image.

- **a** Its position (5 marks)
- **b** Its nature (1 mark)
- **c** Its magnification (3 marks)
- **d** Its size (2 marks)

Answers **Logic**

a



◀ **Figure 11.38**
Ray tracing diagram showing a virtual, upright image in a convex lens

From the diagram in Figure 11.38, it can be seen that the image is magnified, upright and on the same side of the lens as the object, hence virtual.

By calculation:

$u = 6.0\,\text{cm}, \quad v = ?, \quad f = 10.0\,\text{cm}$

$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$	Use the correct formula.	1 mark
$\frac{1}{v} = \frac{1}{f} - \frac{1}{u}$		
$\frac{1}{v} = \frac{u - f}{fu}$		
$v = \frac{fu}{u - f}$	Rearrange the formula to make *v* the subject.	2 marks
$v = \frac{10\,\text{cm} \times 6\,\text{cm}}{6\,\text{cm} - 10\,\text{cm}}$	Substitute the correct values.	1 mark
$v = -15\,\text{cm}$	Calculate the answer.	1 mark
Image is 15 cm from the lens on the same side as the object.		1 mark
b A negative result means a virtual image.		1 mark
c $M = -\frac{v}{u}$	Use the correct formula.	1 mark
$M = -\frac{-15\,\text{cm}}{6\,\text{cm}}$	Substitute the correct values.	1 mark
$M = +2.5$	Calculate the answer.	1 mark
d $M = \frac{h_i}{h_o}$	Use the correct formula. Substitute the known values and calculate the answer.	2 marks
$\frac{h_i}{h_o} = M$		
$h_i = Mh_o$		
$h_i = 2.5 \times 2.0\,\text{cm}$		
$h_i = 5.0\,\text{cm}$		

Try these yourself

1 An object 2.0 cm high is placed in front of a converging lens of focal length 10.0 cm. The object is placed at half the focal distance; that is, 5.0 cm from the lens. Use an accurately drawn ray tracing diagram to find these properties of the image. (11 marks)

- **a** Its position
- **b** Its nature
- **c** Its magnification
- **d** Its size

2 An object 2.0 cm high is placed in front of a converging lens of focal length 8.0 cm. The object is close to the focal distance away from the lens, that is, at 6.0 cm from the lens. Use an accurately drawn ray tracing diagram to find these properties of the image. (11 marks)

- **a** Its position
- **b** Its nature
- **c** Its magnification
- **d** Its size

Images formed by reflection in curved mirrors

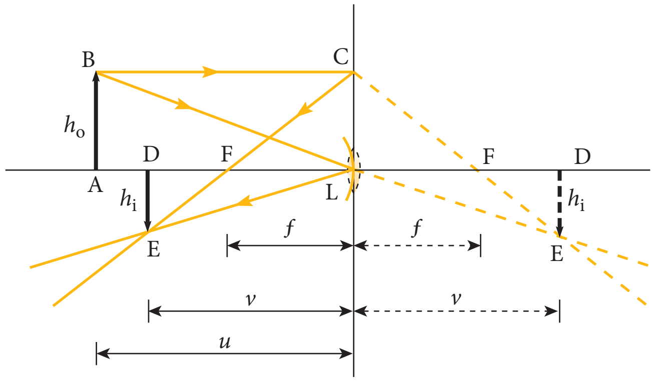


Figure 11.39 ▲
The geometry for a converging mirror is the reflection of the geometry for the converging lens.

When you take the converging lens geometry and reflect it in the plane in which the lens stands, you produce the converging, concave mirror diagram. This diagram has all the same triangles as the convex lens diagram, only this time they are on the same side as the object.

Because all the geometry and sign conventions are the same as for the converging lens, the same formulas apply.

$$M = -\frac{v}{u}$$

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

Curved mirrors simply reflect rays where lenses refract rays. The reflected geometry produces the same relationships between triangles: hence the curved mirror formulas are the same as the thin lens formulas:

$$M = -\frac{v}{u}$$

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

WOW

Reflecting telescopes vs refracting telescopes

Almost all the first telescopes were constructed from lenses. This made them long and heavy. Newton introduced an important change to telescopes when he folded the ray paths back on themselves in a reflecting telescope. Reflecting telescopes became shorter and less cumbersome. Magnification increased. Newton's first reflecting telescope produced images that were about 64 times larger than in a similar length refractor.

Getty Images/English School/Bridgeman Art Library

▲ **Figure 11.40**
Newton's telescope

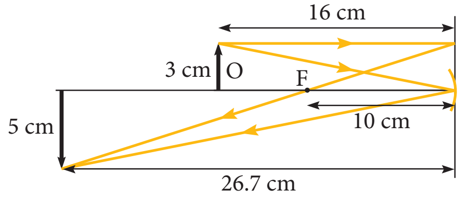
WORKED EXAMPLE 11.6

An object 3.0 cm high is placed 16 cm in front of a concave mirror of focal length 10.0 cm.

a Draw an accurate ray tracing diagram. (6 marks)

b Use the diagram to find these properties of the image.

- **i** Its position (1 mark)
- **ii** Its nature (1 mark)
- **iii** Its size (1 mark)
- **iv** Its magnification (2 marks)

Answers	Logic	
a  ▲ **Figure 11.41** Ray tracing diagram showing a real, inverted image in a concave mirror.	Draw the axes correctly, label the foci and mark in the object correctly. Use a consistent scale. Draw two useful rays to and from the mirror. Locate the image correctly, both horizontally and vertically.	2 marks 2 marks 2 marks
b From the accurately drawn ray diagram in Figure 11.41:		
i The image is 26.7 cm from mirror on the opposite side of the lens axis.		1 mark
ii The image is real.		1 mark
iii Size of image is 5.0 cm.		1 mark
iv $M = \frac{h_i}{h_o}$	Use the correct formula.	1 mark
$M = -\frac{5}{3} = -\frac{26.7}{16} = -1.7$	Substitute the known values from diagram.	1 mark

Try these yourself

1 An object 6.0 cm high is placed in front of a converging (concave) mirror of focal length 5.0 cm. The object is at a position that is less than double the focal distance; that is, 10 cm from the mirror. Use an accurately drawn ray tracing diagram to find these properties of the image. (6 marks)

- **a** Its position (2 marks)
- **b** Its nature (1 mark)
- **c** The size of the image (1 mark)
- **d** The magnification (2 marks)

2 An object 4.0 cm high is placed in front of a concave mirror of focal length 6.0 cm. The object is at a position that is more than double the focal distance; that is, 15.0 cm from the mirror. Use an accurately drawn ray tracing diagram to find these properties of the image. (6 marks)

- **a** Its position (2 marks)
- **b** Its nature (1 mark)
- **c** The size of the image (1 mark)
- **d** The magnification (2 marks)

Case study

Professor Joanne Wood, School of Optometry and Vision Science at Queensland University of Technology

Professor Joanne Wood works in teams that study the effect on driving performance of spectacles and contact lenses designed to correct presbyopia, eye diseases (glaucoma, age-related macular degeneration), and optical blur, which is increased when people do not use their spectacles or have the wrong prescription. They found that the effects of blur are much greater at night, which may be a function of the increased size of the pupil. Other researchers in the School of Optometry are also investigating a link between eye aberrations and myopia, where the eye grows longer, as well as whether or how eye growth is regulated by eye optics.

Figure 11.42 ▲ Professor Joanne Wood

Joanne's group was the first to discover the link between 'biomotion' and safety for pedestrians, bicyclists and road workers. Placing reflective materials on wrists, knees and ankles, makes it 4–6 times easier to recognise people at night. That means fewer people are killed or injured. By law, night workers must wear clothing with reflectors in the biomotion configuration. Now, the group is trying to get recreational walkers, runners and bicyclists to wear biomotion reflectors.

Most kids from Joanne's school did not go on to university. 'I worked really hard to get the best grades that I could', she said, 'because I wanted a career in some sort of science'. She decided on optometry after starting to wear glasses. 'I found the way the eye works fascinating.'

Joanne gained a first class honours science degree in Optometry in Birmingham (UK). Her 'interest in research had been ignited' so, after two enjoyable years in clinical practice, Joanne started a PhD, 'to learn how to become a research scientist.' Subsequently she worked at Oxford University in a postdoctoral position, and then came to QUT. 'I just loved the variety of work. The climate and beaches were very appealing too!' It was lonely, however, until she made friends.

In Australia, Joanne first supervised a small-scale project about how the functional effects of eye diseases and loss of peripheral vision affected driving performance. 'It was', she said, 'a very ambitious project, but it turned out to be very promising and fun. It formed the basis for a series of important studies that are still widely cited in the scientific literature today.'

Vision science requires experts in many fields, such as eye diseases (for example, cataracts, macular degeneration, glaucoma), neuropsychology (stimulus/response, recognition, attention, mental processing), and community wellbeing (road safety, protective behaviours). But interesting puzzles require her to work within multidisciplinary teams with colleagues from psychology, optics, ophthalmology, audiology, engineering and surveying. There's nothing very 'typical' about Joanne's working day. 'We usually work on a number of 'puzzles' at any one time. It's what keeps me so interested and so motivated.'

On the scientific process, Joanne said, 'Conducting experiments can be very exciting and rewarding. But it does involve a high level of rigour to ensure that the results are valid. That's why having an excellent grounding in scientific techniques at high school and university is critical. Generally, we start with a series of research ideas, which are discussed with the research team. This may involve email or Skype contact with international partners. A series of hypotheses or research questions are developed, and the most appropriate experimental design determined. We conduct a series of pilot experiments to check that the design isn't flawed. Refining the experimental design can take some time in order to ensure that it is as rigorous as possible – that's why it's good to have a multidisciplinary team. They provide different perspectives at this stage of development. All research on human participants also has to have ethical approval. We then collect and analyse the data and compare the results with previous literature. This can be very exciting, particularly if the results are novel or very different from those you predicted.'

'Presenting results to your peers and writing them up for publication is a critical stage of the process. You get feedback, which provides really important insights into your results (and generates more ideas too). Developing good scientific writing skills and learning to present results to an experienced audience is a skill that is really important to develop. It's no good having great findings but being unable to communicate them to a wider audience, including your peers and the wider community. But that's part of the excitement and fun of being a scientist. Great researchers are not just excellent scientists they are also good communicators too. It is important to develop skill sets in a wide range of areas as you progress through high school and university, and that affords you the opportunity to make a real difference!'

Questions

1 How can 'biomotion' be observed?
2 Draw and annotate ray optics diagrams to show and explain image formation for someone with:
   a myopia.
   b hyperopia.
3 How do the following affect vision?
   a Cataracts
   b Macular degeneration
   c Glaucoma
4 What are 'aberrations'? How do they affect images in eyes?
5 Why are multidisciplinary teams so useful in vision science? Refer to an example in the article.
6 Construct a flowchart of the scientific process as described by Joanne Wood. Why does she include communication skills as part of the process?

QUESTION SET 11.5

Remembering

1 Draw diagrams to show the defining feature of:
   a convex lenses.
   b concave lenses.
   c convex mirrors.
   d concave mirrors.
2 List the paraxial assumptions for the ray tracing representation of image formation in lenses and curved mirrors. Illustrate these assumptions on a diagram.
3 Write the equations used when calculating image positions and magnification. Sketch and annotate a ray diagram showing real image formation in a convex lens.

Understanding

4 Using Figure 11.33 (see p. 382), describe the way the ray model represents the wave model.

Applying

5 An object 5.0 cm tall is placed 20 cm in front of a concave mirror of focal length 10 cm. Calculate the distance of the image from the mirror.
6 Determine the position, nature and size of the image of an object 5.0 cm tall placed 10.0 cm in front of a concave mirror of focal length 20.0 cm by scale drawing.

Analysing

7 An object 15.0 cm high is placed 35.0 cm from a converging lens of focal length 10.0 cm. Determine the position, nature and size of the image formed by calculation.
8 An object is placed 20.0 cm in front of a converging lens, and an inverted image three times the size of the object is obtained.
   a Show this situation with a geometric scale drawing.
   b Where must the object be placed to get a real image twice its size?

Reflecting

9 Image formation can be analysed by the method of geometric scale drawing or by algebraic methods or by a combination of geometric drawing and algebraic manipulation. Both methods are equivalent representations of the physical situation. Discuss the merits of each method. Which of the three methods do you prefer? Why?
10 How did you respond to the way the image formation geometry for convex lenses was applied to concave mirrors (see p. 388)?

Diffraction

Diffraction occurs when a narrow beam of light passes through a narrow gap, and spreads out into the space beyond. Diffraction is regarded as a wave effect. Thus, sound diffracts around corners. Light diffraction through a single gap is explained by analogy with wave phenomena – the wave model – with which we are familiar.

When looked at closely, the diffraction pattern from a single slit shows 'structure'. It has a large central bright spot, and less intense bright patches on both sides. Between the bright patches are dark patches (Figure 11.43).

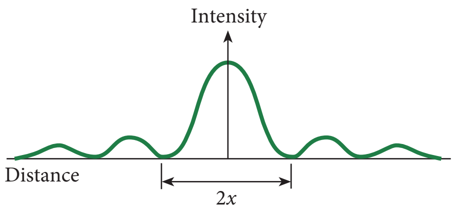


Figure 11.43 ▶ Intensity vs distance from the centre for a single-slit diffraction pattern

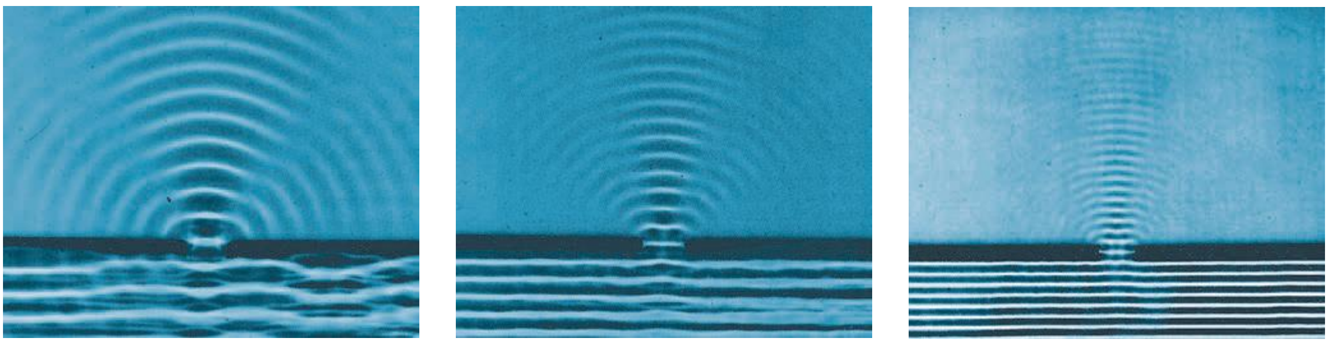
In the wave model, the angular spread of the bright central patch, θ, is explained in terms of the wavelength of the light and the slit width:

$$\theta \propto \frac{\lambda}{w}$$

Diffraction effects become noticeable when wavelength and slit width are comparable. This means that when $\frac{\lambda}{w} > 1, \Rightarrow \lambda > w$, the light simply spreads into most of the area, and the central maximum is quite wide (large θ).

Diffraction effects are more pronounced when the ratio is large:

$$\frac{\lambda}{w} \geq 1$$



▲ Figure 11.44
Diffraction of water waves – a model for light diffraction. Waves spread into the region beyond the gap. The spread of the central maximum decreases as the wavelength becomes similar to, or smaller than the gap width.

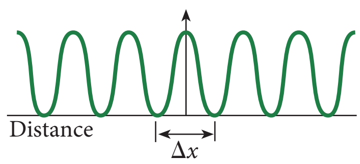
> **Diffraction is the spreading of waves into a region. Angular spread of the central maximum,** $\theta, \propto \frac{\lambda}{w}$.
>
> **Diffraction becomes noticeable when** $\frac{\lambda}{w} \geq 1, \Rightarrow \lambda \geq w$.

Young's double-slit experiment

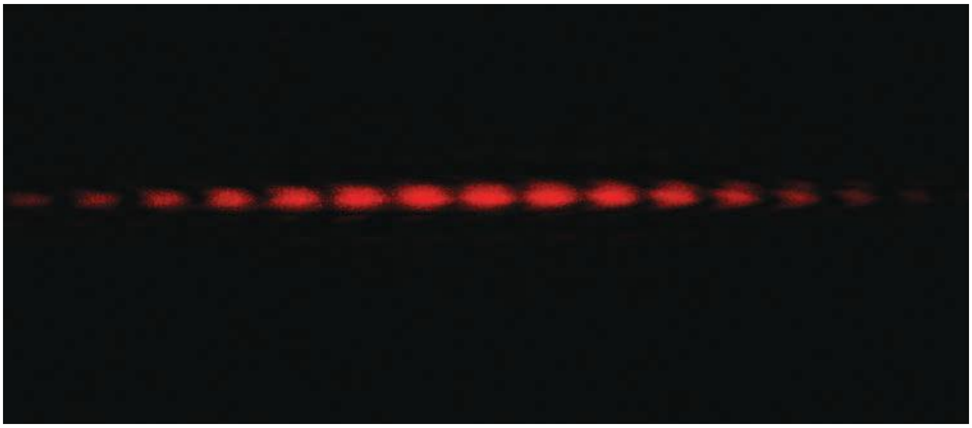
When a narrow beam of light strikes two slits, the slits produce diffraction patterns, which then overlap. A pattern of bright and dark patches is noticeable on a screen some distance away. Unlike a single slit, the central maximum, while still the brightest, is less wide.

This phenomenon was first demonstrated by Thomas Young in 1801. As a result, interest was renewed in the wave model. First developed by Christiaan Huygens (1629–95), a contemporary of Newton's, the wave model had languished in favour of Newton's corpuscular (particle) model for more than a century.

The double-slit phenomenon can be explained as a wave interference effect.



▲ Figure 11.45
Intensity vs distance from centre for a double-slit interference pattern. Δx is the distance between dark bands.



Getty Images/Edward Kinsman

◀ Figure 11.46
A laser beam produces an interference pattern when passed through a double-slit arrangement.

YOUNG'S DOUBLE SLIT
See some everyday examples of Young's double slit.

ACTIVITY 11.1

EVERYDAY EVIDENCE FOR THE WAVE-LIKE BEHAVIOUR OF LIGHT

Aim

To observe diffraction and interference patterns

You will need

Your hand, a laser pointer, a piece of card, a sharp pin, a piece of pantyhose material (good-quality, close-weave dark colours work best). If you have them, a slide with a single slit and another slide with a double slit for observing diffraction and interference can also be used.

What are the risks in doing this activity?	How can you manage these risks to stay safe?
Damage to eyesight can occur from direct exposure to intense light.	Never use this equipment to observe lasers or sunlight directly. Use a low power laser pointer.
Damage to bystanders' eyes from laser light can occur.	Take care not to shine the laser at other people. Reduce all possible reflections.

What to do

1 Hold your first and second fingers very close together – almost but not quite touching. Use your other hand to hold them steady. Now look through the very small gap between your fingers at a light source such as an overhead light or a window. *Do not look at a laser.* Can you see bright and dark lines in the gap between your fingers?

2 Make the smallest hole you can using the tip of the pin to just prick a hole through the card. Shine the laser at the hole and observe the light that has passed through the hole on a wall or screen. The card should be at least a metre from the screen. Can you observe a diffraction pattern? Try shining white light, for example from a torch, through the pinhole. What do you observe now?
3 Repeat your observations using pinholes of various sizes. How does the pattern depend on the size of the hole?
4 Now shine the laser through the piece of panty hose. You will need one person to hold the fabric taut and another person to shine the laser through it. Can you see a pattern on the screen or wall? What happens to the pattern when you stretch the fabric one way? What happens when you stretch it the other way?
5 If you have slides with single and double slits or diffraction gratings, use the laser to observe interference patterns from these slides. You can also use a CD or DVD to observe an interference pattern in light reflected from its surface.

The light from the original source spreads out as waves into the region behind the slits. Each wavefront that strikes the double slit is sampled by the slits. The slits act effectively as new sources of circular waves. A plane wave crest becomes a circular crest at each slit. A plane wave trough becomes a circular trough at each slit. Since the waves from these new sources come from the same original wavefront overlap. Waves that are in phase have peaks and troughs occurring at the same time. Hence a peak is incident on, and leaves from, each slit simultaneously. The troughs coming behind these peaks do the same, and so on. This happens for all wavefronts, even if they were emitted from the original source in a random way. This leads to the formation of an interference pattern that does not change with time.

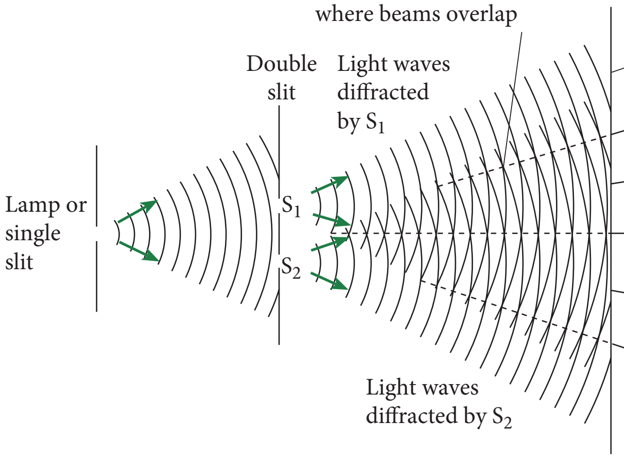


Figure 11.47 ▶ Waves from a source producing waves randomly are incident on a double slit arrangement. Each wavefront is sampled simultaneously at both slits, leading to the formation of a consistent pattern of maxima and minima.

WOW

Thomas Young (1773–1829)

Thomas Young, born in Somerset, England, was a true child prodigy. By 14 he was conversant in Greek, Latin, Hebrew, Arabic, Persian, Turkish and Ethiopian. Later, this would help him decode the Rosetta Stone and pioneer the reading of Egyptian hieroglyphics. Young studied medicine in three countries before practising in London. At medical school he studied the eye lens, focusing and accommodation and the cause of astigmatism. Later he investigated colour sensation, showing that the eye needed only red, green and blue to perceive all the other colours.

In 1801, Young demonstrated the interference effect of light, thus illustrating the wave-like nature of light. This was strongly criticised by supporters of Newton's particle model of light.

Young gave a more precise explanation than Hooke (Hooke's law) of forces applied by materials by introducing the terms 'stress' and 'strain'.

Often, Young failed to pursue his hypotheses, so that, even in his lifetime, other scientists received credit for ideas that he first proposed.

Double slit interference patterns: quantitative analysis

A wave train may be considered as a series of positive crests and negative troughs. If two crests or two troughs overlap, they increase the amplitude. This is called constructive interference. Destructive interference occurs when a crest and a trough overlap.

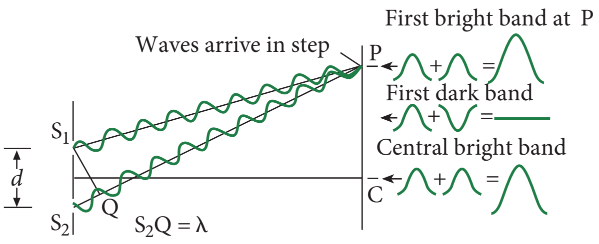


Figure 11.48 Path differences lead to maxima and minima. The formation of the first bright band is shown.

Constructive interference

Everywhere along the perpendicular line between the slits, crests and troughs that have been produced from the same wavefront will overlap. This gives rise to the central maximum. Other maxima occur as a result of constructive overlap between crests and troughs that have been emitted earlier at one slit relative to the other slit. So long as the path difference between these waves is a whole number of wavelengths, there will be constructive interference.

For constructive interference, path difference $= n\lambda$, $n = 1, 2, 3, \ldots$

Destructive interference

In between these maxima there are minima, also called nodes or nodal points, where crests produced earlier at one slit overlap with troughs produced later. In these cases the path difference is an odd number of half wavelengths:

For destructive interference, path difference $= (2n - 1)\frac{\lambda}{2}$, $n = 1, 2, 3, \ldots$

OR

For destructive interference, path difference $= (n - \frac{1}{2})\lambda$, $n = 1, 2, 3, \ldots$

Conditions for interference from two point sources in phase

Constructive interference:

path difference $= n\lambda$, $n = 1, 2, 3, \ldots$

Destructive interference:

path difference $= (2n - 1)\frac{\lambda}{2}$, $n = 1, 2, 3, \ldots$

OR

path difference $= (n - \frac{1}{2})\lambda$, $n = 1, 2, 3, \ldots$

The double slit interference pattern involves the overlap of nodes and antinodes. This is similar to the interference of waves from each end of a string. The string ends function in the same way as the two slits.

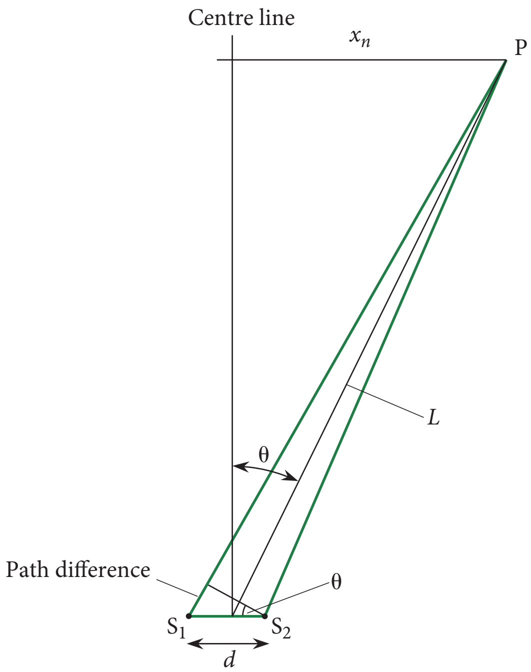


Figure 11.49 ▲ Geometry for Young's double-slit experiment

Determining the wavelength of light from a double-slit interference pattern

A good approximation of the wavelength of light can be experimentally determined by the Young's double-slit experiment. If the screen is a long way from the slits, and this can be easily managed, then the geometry (Figure 11.49) shows:

- the distance across the screen to the *n*th minimum, x_n
- the distance from the centre of the slits to the *n*th minimum, L
- the distance between the slits, d
- the path difference
- the two angles that are equal; that is, the difference between them is negligible, θ.

From the similar triangles:

$$\frac{x_n}{L} = \frac{\left(n - \frac{1}{2}\right)\lambda}{d}$$

$$\Rightarrow \lambda = \frac{dx_n}{\left(n - \frac{1}{2}\right)L}$$

The distance between neighbouring minima, Δx, can be deduced as follows:

$$\frac{x_n}{L} = \frac{\left(n - \frac{1}{2}\right)\lambda}{d}$$

$$\Rightarrow x_n = \frac{\left(n - \frac{1}{2}\right)\lambda L}{d} \text{ and } x_{n+1} = \frac{\left(n + 1 - \frac{1}{2}\right)\lambda L}{d}$$

$$x_{n+1} - x_n = \frac{\left[\left(n + 1 - \frac{1}{2}\right) - \left(n - \frac{1}{2}\right)\right]\lambda L}{d}$$

$$\Rightarrow \Delta x = \frac{\lambda L}{d}$$

This provides another way to measure the wavelength of light.

The spread of the interference pattern on a screen, effectively given by Δx, depends on:

- the wavelength, λ, of the light irradiating the slits. The smaller the wavelength, the smaller the spread of the pattern:

$$\Delta x \propto \lambda$$

- the distance, L, from the slits to the screen. The distance between two minima on the screen, Δx, will double if the distance, L, doubles:

$$\Delta x \propto L$$

- the separation, d, between the slits. As the separation decreases the pattern expands:

$$\Delta x \propto \frac{1}{d}$$

Young's double slit: wavelength of light measurement

Pattern spread is proportional to λ and L and inversely proportional to d

Transposing for wavelength: $\lambda = \frac{d\Delta x}{L}$

THE MOST BEAUTIFUL EXPERIMENT

How was Young's double-slit experiment used in the most beautiful experiment? How was this result strange?

WORKED EXAMPLE 11.7

In an experiment to measure the wavelength of red light, the following results were obtained:

Distance of screen from double slits: 180 cm

Slit separation: 0.090 mm

Distance between 7 dark bands: 7.2 cm (count the start as one dark band)

a What value would the student have obtained for the wavelength? (5 marks)

b What would happen to the spread of the interference pattern if the wavelength of the light were changed to 450 nm? (3 marks)

Answers	Logic	
a $\Delta x = \frac{\lambda L}{d}$	Use the correct formula.	1 mark
$\Rightarrow \lambda = \frac{d\Delta x}{L}$	Rearrange the formula to make λ the subject.	1 mark
$\frac{9 \times 10^{-5} \times \left[\frac{7.2 \times 10^{-2}}{6}\right]}{1.80}$	Substitute the correct values.	2 marks
$= 6.0 \times 10^{-7}$ m = 600 nm	Calculate the answer.	1 mark
b The new wavelength is less than the wavelength in part a.	State the relevant observation.	1 mark
Therefore, as $\Delta x \propto \lambda$, the separation of minima is decreased.	Give the correct deduction from the relevant relationship.	1 mark
Thus, the pattern spreads less into the region.	Draw the correct conclusion.	1 mark

Try these yourself

1 In an experiment to measure the wavelength of a monochromatic source of light, the following results were obtained.

Slit separation: 5.0×10^{-5} m

Distance between 10 dark bands: 12.2 cm

Distance from slits to screen: 1.50 m

a Calculate the wavelength. (5 marks)

b What would happen to the spread of the interference pattern if the wavelength of the light were changed to 600 nm? (3 marks)

2 Calculate the wavelength of light from the following data recorded in a Young's double-slit experiment. (5 marks)

Distance of screen from double slits	4.85 m
Slit separation	0.029 mm
Distance between 80 dark bands	96.1 cm

You will consider this more fully in Unit 4: Quantum physics.

QUESTION SET 11.6

Remembering

1 Draw a diagram to show the intensity of light on a screen when:
  a light diffracts through a single slit.
  b light interferes after travelling through a double-slit arrangement.

2 Write down the path difference relationship and the sequence of values for *n* for:
  a constructive interference.
  b destructive interference.

Understanding

3 For a single-slit diffraction pattern projected onto a screen, what happens to the central maximum if red light is replaced by yellow light?

4 For a Young's double-slit experiment, sketch a diagram to show how the path difference, $S_1P - S_2P$, from slits S_1 and S_2 to point P on a screen some distance away causes constructive interference.

Applying

5 If the path difference to the second dark band away from the central maximum of a Young's double-slit experiment is 750 nm, what is the wavelength associated with the source of light used?

6 A diffraction pattern is formed on a screen when yellow light is incident on a narrow single slit.
  a What happens to the central maximum if the width of the slit is increased?
  b What happens to the pattern if the screen is moved further away from the slit?

Analysing

7 What happens to the spacing between the bright bands of a Young's double slit pattern when:
  a the screen is moved further away from the slits?
  b the source is moved closer to the slits?

8 The wavelength associated with a monochromatic light source was measured in a Young's double slit experiment. The slits were 0.05 mm apart and the distance to the screen was 200 cm. The distance between adjacent dark bands was 16 mm.
  a What was the wavelength of the unknown source?
  b What would be the distance between adjacent dark bands if the source was changed to light of wavelength 512 nm?

Reflecting

9 Describe your experience of observing the regular patterns associated with light going through single and double slits? Was it surprising to learn how to look for these effects?

10 In what ways has your understanding of the interactions of light with matter developed as a result of considering diffraction and interference effects?

CHAPTER SUMMARY

- The interactions of light with matter can be explained using models.
  - Wave model: Light interacts with matter as an electromagnetic wave.
  - Photon model: Light interacts with matter as a particle that exhibits wave properties.
  - Ray model: Light interacts with matter along straight lines.
- Straight-line propagation
  - Speed: $3.0 \times 10^8\,\text{m}\,\text{s}^{-1}$ in a vacuum
- Shadow formation
  - Umbra: deep shadow
  - Penumbra: shadow where some but not all light from an extended source is blocked
- Reflection (diffuse and regular)
  - The incident ray, the normal and the reflected ray are coplanar.
  - The two angles are equal: $\angle i = \angle r$
- Refraction
  - The incident ray, the normal and the refracted ray are coplanar.
  - Snell's law: $\frac{\sin i}{\sin R} = \text{constant}$
- Polarisation
  - Polariser allows only one part of the electromagnetic wave through.
  - Analyser blocks or allows light to pass, depending on its orientation to the polarised light.
- Intensity

$$I = \frac{S}{4\pi r^2}$$

$$\Rightarrow I = \frac{1}{4\pi}\left(\frac{S}{r^2}\right)$$

- Image formation in lenses and curved mirrors

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

$$M = \frac{-h_i}{+h_o} = -\frac{v}{u}$$

- Diffraction

$$\theta = \frac{\lambda}{w}$$

- Interference
  - For constructive interference: path difference $= n\lambda$, $n = 1, 2, 3, \ldots$
  - For destructive interference: path difference $= (2n-1)\frac{\lambda}{2}$ or $(n - \frac{1}{2})\lambda$, $n = 1, 2, 3, \ldots$

$$\Delta x = \frac{\lambda L}{d}$$

CHAPTER GLOSSARY

absolute refractive index a measure of the refrangibility of a medium placed in a vacuum and subjected to an incident ray of light

analyser material that allows or stops polarised electromagnetic radiation

angle of incidence at a boundary, the angle between the normal and the incoming ray

angle of reflection at a boundary, the angle between the normal and the outgoing ray

angle of refraction the angle that the refracted ray makes with the normal line

Brewster angle angle at which light from a surface is polarised

chromatic dispersion refraction occurs differently for different colours in the same material; colours spread

cladding glass surrounding the core of an optical fibre; $n_{clad} < n_{core}$

converging (convex) lens lens thicker in the middle than at the end

coplanar in the same plane

core inner glass of optical fibre

critical angle angle of incidence for which the angle of refraction is 90° (total internal reflection occurs); beyond the critical angle reflection, but not refraction, occurs

diffuse reflection/scattering reflection from a rough surface; rays in a beam reflect in different directions

diverging (concave) lens lens that is thicker at the edges than at the centre

eclipse movement of Earth into the Moon's shadow (solar eclipse) or movement of the Moon into Earth's shadow (lunar eclipse)

electrical permittivity, ε property of a medium that affects the propagation of the electrical component of an electromagnetic wave

electromagnetic wave model light acts like a transverse wave that has electric and magnetic components

focal length distance from lens to focal point

focal point light parallel to the axis of a lens or curved mirror is concentrated at this point

image picture of an object

interference wave overlap

luminous a source that produces light

magnetic permeability, μ property of medium that affects the propagation of the magnetic component of an electromagnetic wave

magnification ratio of image height to object height; $M = \frac{h_i}{h_o}$

non-luminous a source that reflects light

normal at right angles to a surface

optical centre centre of curvature of a lens or mirror

optical fibre transparent light guide making use of total internal reflection at a boundary between two materials of similar refractive index

paraxial assumptions for curved lenses and mirrors: (a) lenses and mirrors must be small and thin, (b) rays must be near and parallel to the principal axis and (c) a straight line approximates the curved surface

partial eclipse when an object is not entirely obscured from an observer by another object

penumbra shadow region where some light penetrates

point source single localised source from which light transmits equally in all directions

polarisation orientation in one direction of the electric part of electromagnetic waves

polariser material that selects the direction of polarisation

principal axis line through both focus and centre of a curved lens or mirror system and perpendicular to the axis of the lens

ray diagram a geometric construction using the law of reflection that is used to find the magnification and position of the image

real image rays pass through the site of the image

refractive index measure of refrangibility; measure of the relative amount of change of direction of waves or light rays when travelling from one medium into another

refrangibility ability of a material to refract light

regular/specular reflection predictable reflection from a very smooth surface; rays in a beam all reflect in the same direction

relative refractive index comparative difference in refrangibility between two media with different absolute refractive indices

thin lens equation for thin lenses and small curved mirrors: $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$

total eclipse when an object is entirely obscured from an observer by another object; total solar eclipse is when the Sun is completely blocked by the Moon

umbra shadow region where no light penetrates

virtual image image of an object where the rays do not pass through the image; the image cannot be projected onto a screen

CHAPTER REVIEW QUESTIONS

Remembering

1 a Draw ray model diagrams to show:
   i shadow formation.
   ii reflection of light from a painted wall.
  b Draw a wave model diagram to show specular reflection.

2 Write the equation for the way intensity of a light source depends on distance from the source. Define the terms used.

3 Write Snell's law for waves. Annotate a diagram to define each quantity.

4 Sketch and annotate the ray diagram showing real image formation in a concave lens.

Understanding

5 Explain how a model of light is really a model of the interaction of light with matter. Refer to polarisation in your answer.

6 Infrared radiation has a wavelength of 980 nm. What is its frequency?

7 a Total internal reflection is a refraction phenomenon. Discuss.
  b The critical angle for light passing from glycerine to air is 42.9°.
   i What is the index of refraction of glycerine?
   ii What is the angle of refraction for light from air passing into glycerine at an angle of incidence of 42.9°?

8 Why can light from a torch make a wall easier to see than a mirror?

9 Rays from an object radiate in all directions from every point on an object. How do ray diagrams represent wave transmission of energy from each point?

10 For a Young's double-slit experiment, sketch a diagram to show how the path difference, $S_1P - S_2P$, from slits S_1 and S_2 to point P on a screen some distance away causes destructive interference.

Applying

11 If it is possible, how can light from a point source pass through:
  a both a polariser and an analyser.
  b an analyser but not a polariser.
  c a polariser but not an analyser.

12 Jana is 160 cm tall and stands 2.0 m in front of a flat mirror mounted vertically on a wall. She is just able to see her entire image. Jana's friend, Rana, is 190 cm tall. The eyes of both girls are 10 cm below the top of their heads.
  a What is the length of Jana's mirror? Illustrate your answer with a diagram.
  b Rana cannot see herself fully in Jana's mirror even if she stands back further from the mirror. Explain why.
  c What length mirror is needed so that both Rana and Jana can see themselves fully?

13 A ray of blue light of wavelength 485 nm travels from air into a crown glass block at an angle of 40.0°. The speed of light in air = $3.000 \times 10^8\,\mathrm{m\,s^{-1}}$ and the refractive index for blue light in crown glass = 1.527.
  a Calculate the angle of refraction as the light passes into the crown glass.
  b For light transmitted into crown glass, find:
   i the frequency of the light.
   ii its wavelength.
   iii the speed of the light in the crown glass.

14 An object 4.0 cm tall is placed 12.0 cm in front of a concave mirror of focal length 8.0 cm. Use ray tracing to find:
  a the height of the image.
  b the distance of the image from the mirror.

15 An object 4.0 cm tall is placed 12.0 cm in front of a convex lens of focal length 8.0 cm. Determine these properties of the image.
   a Position
   b Nature
   c Size
16 The path difference to the third bright band away from the central maximum of a Young's double-slit experiment is 1485 nm. What is the wavelength associated with the source of light used?

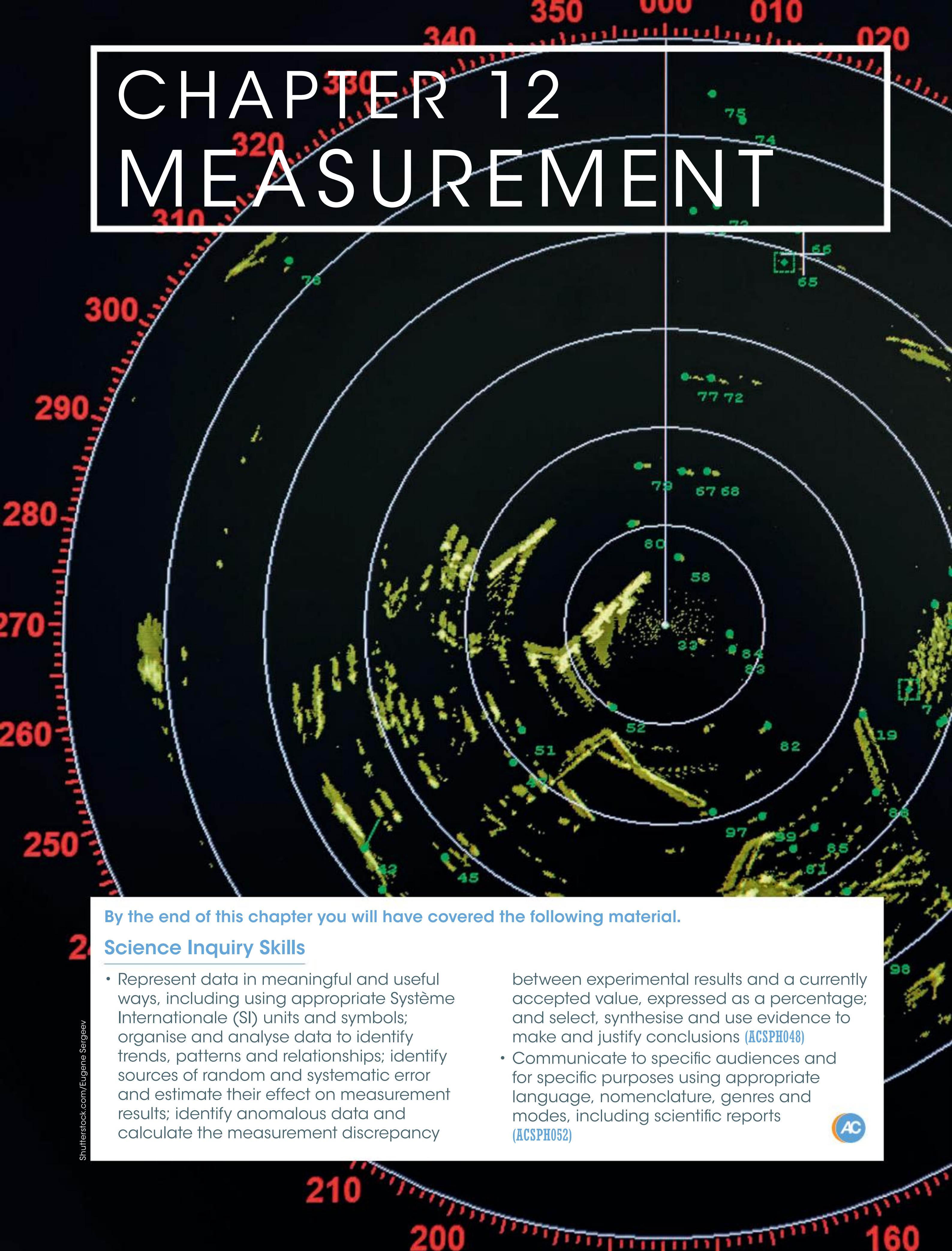
Analysing

17 Is grass really black when illuminated by orange light?
18 Show how changes in technology helped Roemer to demonstrate that light has a finite speed.
19 At 1.5 m from a light source the intensity is 0.13 $W\,m^{-1}\,s^{-1}$. What is the intensity at:
   a 4.5 m?
   b 0.5 m?
20 How is it possible to 'see' a virtual image?
21 Two mirrors that meet at right angles are called a corner mirror. A ray of light from a point 2.5 cm from both mirrors is incident on one mirror at an angle of incidence of 20°. It reflects from both mirrors.
   a What will be the subsequent path of the light ray?
   b How many images will be formed in the two-mirror system?
   c Show that an incoming narrow beam of parallel rays will be reflected from a corner mirror as a narrow beam that is parallel to the incoming beam.
22 At what minimum angle of incidence would a diver need to shine a laser towards the surface of the water (n_{water} = 1.33) so that the light was not transmitted into the air above?
23 A ray of light is shone from air into an optical fibre. All the light transmits down the fibre by total internal reflection. What is the maximum angle of incidence at the air–core boundary for which this can occur, given that n_{core} = 1.500 and n_{clad} = 1.490?
24 The image of an object in a convex mirror is magnified 2.5 times. Use a ray tracing diagram to scale to show how this is possible. Describe the image fully.
25 A Young's double slit experiment was used to find the wavelength of a monochromatic light source. The screen used was 1.80 m from the slits. The distance between the fifth dark band on either side of the central maximum was measured to be 145 mm. The slits used were 0.08 mm apart.
   a Calculate the distance between adjacent bright bands.
   b What is the wavelength of the unknown source?

Reflecting

26 What have you learnt about the following?
   a The description and explanation of images in mirrors and lenses
   b Physics, technology and knowledge building
27 How has your understanding of the relationship between geometry and algebra changed?
28 Do you think diffraction and interference effects prove that light is a wave? Discuss.

CHAPTER 12
MEASUREMENT



Shutterstock.com/Eugene Sergeev

By the end of this chapter you will have covered the following material.

Science Inquiry Skills

- Represent data in meaningful and useful ways, including using appropriate Système Internationale (SI) units and symbols; organise and analyse data to identify trends, patterns and relationships; identify sources of random and systematic error and estimate their effect on measurement results; identify anomalous data and calculate the measurement discrepancy between experimental results and a currently accepted value, expressed as a percentage; and select, synthesise and use evidence to make and justify conclusions (ACSPH048)
- Communicate to specific audiences and for specific purposes using appropriate language, nomenclature, genres and modes, including scientific reports (ACSPH052)

Introduction

Physics, and science more generally, is the study of phenomena and their causes. We ask two questions of phenomena: What happens? Why does it happen?

By careful investigation, we attempt to make sure we know what happens, and with what degree of certainty we can say that the phenomenon happens. That is the purpose of experimentation. Once we are confident that we know what actually does happen, we search for a satisfactory **explanation**. There may be competing explanations, so we try to test the explanations to see whether they stand up in all circumstances. In this way, the explanations enable us to look for further data: data and theory inform each other.

In the case of competing explanations, we try to find **critical tests** of the explanations. A critical test will show that one explanation is superior to the other. A superior explanation is able to explain more of the evidence and has greater power to predict, and then explain, novel results. This should mean that one of the explanations is discarded. In reality, both explanations tend to exist side by side until it becomes obvious that one is superior, or an even more effective explanation takes hold.

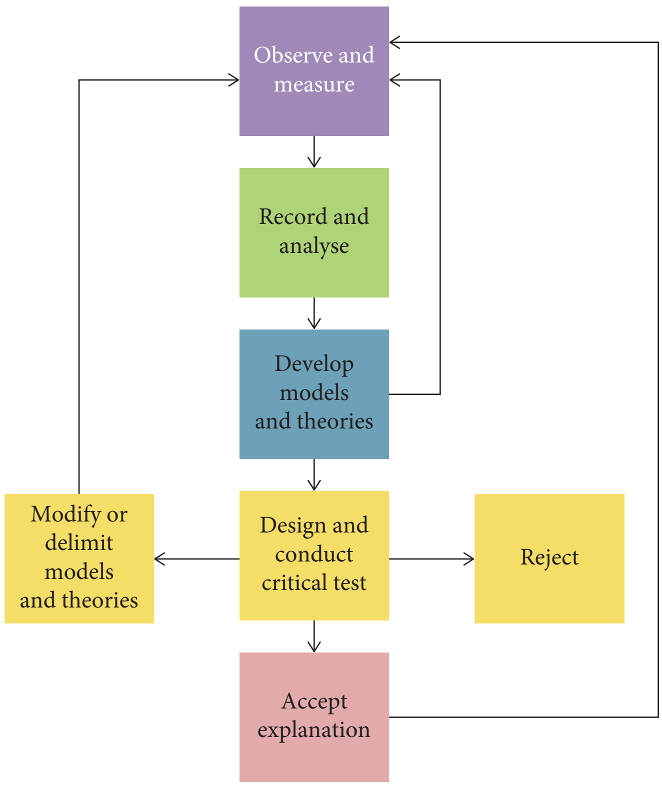


Figure 12.1 ▶ Flowchart representing the enterprise of doing physics

Philosophy and science

Physics was known from classical times as 'Natural philosophy'. It was the study of, and explanation for, the physical world. Aristotle (384 BCE–322 BCE) was one of the most famous and influential of the classical natural philosophers. In the 13th century, Roger Bacon (c.1214–c.1292) integrated new interpretations of Aristotelian ideas into university teaching of theological and philosophical ideas. He argued successfully for reform in university education to include experiential science. He was well-versed in Greek and Muslim thought, especially in optics, which he used to develop a clear description of empirical science as well as its purposes – to gain understanding of phenomena, to explain it and to make predictions.

Aristotle's ideas came under increasingly intense scrutiny during the 15th through 17th centuries. Careful data collection and analysis provided the impetus for the gradual development of a new form of natural philosophy. Finally, the work of Nicolaus Copernicus (1473–1543),

ROGER BACON

Find out more about the incredible breadth of Roger Bacon's scholarship.

Francis Bacon (1561–1626), Galileo Galilei (1564–1642), Isaac Newton (1642–1727) and many others radically changed the philosophical basis of natural philosophy. Physics, the academic discipline we now study, was born.

Karl Popper (1902–94) was a philosopher of science. He argued that science proceeds by the way of **falsifiability**. That is, a theory is formed that starts off with a set of concepts, models, connected ideas and consequences linked to data, just like the solution to a mathematical problem. This construction of a theory fits some data and suggests new experiments. All experiments that fit with the theory and related models serve only to prove what we already know. However, if a theory is to be refined or changed, then we need to search for data from experiments that make the consequences of the theory false (falsifiability). Only then can we improve our understanding.

Falsifiability
Science proceeds by way of critical tests of explanations. A critical test is an experiment that, having been rigorously conducted, shows that one or other of competing theories is false.

Popper's ideas have been challenged by other philosophers of science. They disagree in three ways. First, they argue that falsifiability is too strict a criterion. Scientists gain confidence that they are on the right track when they confirm their theories through experiment.

Second, they point to the way scientists hold firm to useful ideas even when those ideas are under threat. It is not sufficient for new data to outweigh older data. Theories do not fail completely simply because some data falsifies them.

Often, the emerging theory does not seem to be better than the accepted one. This is a common experience for physics students: they do not accept what they are told unless it is better than what they already think. For example, Newton's laws are well accepted. They predict that all masses falling near Earth's surface will accelerate at the same rate. Yet, students who observe different freefalling masses striking the ground simultaneously frequently say, 'I've seen this many times. I've seen it, but I don't believe it.' Students, like scientists, are human. They hang on to dearly held ideas well beyond their use-by date. That slows progress towards better understanding and explanations.

The third way in which falsifiability is used is to constrain or put limits on the applicability of a theory. A theory may be found to be less general than previously claimed. In this way, the theory is saved but reduced in applicability. This happens for Newton's laws, which work well for big, slow things, but not for speeds approaching the speed of light.

Physics and philosophy are always in dialogue. The development of the quantum theory, which described the universe as fundamentally probabilistic, caused consternation to those who favoured a philosophically deterministic physics. It can be argued that Einstein's relativity arose in the 19th century philosophical ferment that resulted in deconstructionism and post-modernism. And let us not forget the possibility that the philosophical underpinnings of mathematics itself may be holding back further developments in science and philosophy.

Models are central to science because scientists use them to describe, explain, relate and predict phenomena. Models can be expressed in a range of ways – via words (with language that is commonly metaphorical), images (actual or imagined), mathematics, computer simulations or physical constructions (including some machines). Models help scientists to frame physical laws and theories, and these laws and theories are also models of the world. Models are not static – as scientific understanding of concepts or physical data or phenomena evolves, so too do the models scientists use to describe, explain, relate and predict these.

Models in science

Data represent physical systems. Data can be organised in tables, graphs, images, diagrams and words. It can be in **quantitative** (numerical) or **qualitative** forms. All of these data sets represent the actual physical system. Relationships between data sets can be represented using mathematical relationships, geometric constructions and diagrams, words, computer programs, physical models, etc. These representations model the situation – they are not the physical reality any more than an architectural drawing is a building.

Models in science demonstrate relationships between measureable quantities. In the process, they are used to explain things. Models or **representations** can show multiple relationships between parameters so that we can predict new observations and relationships. New observations, whether made as a result of experiments carried out to test a model or otherwise, shed light on models. That is, representations are subject to data, and data is described and explained by models. Neither the data nor the model is identical with the situation being observed.

Figure 12.2 ▲ 'My love is like a red, red rose' (Robert Burns). The rose, the girl and the poet's love are not identical with the poet's image or model of love. This photo is also a model; it is neither the girl nor the flowers.

In the case of light, there are competing explanations of data concerning the interactions between light and matter. These explanations are based on a particle model or a wave model. Interestingly, we can show that both these explanations are useful, but for different data sets. There are clear, critical tests where one or the other of the models fails to explain some of the data.

This causes us to reflect on the nature of explanation itself. When we try to explain something, we use our common experience as a starting point. We say that something '*is like*' something else. A picture, an image, a word, a photograph and a computer-generated simulation are examples of models. So are mathematical equations. Theories model reality. They are not themselves reality. Poets use imagery to model their vision of the world.

Lord Kelvin (1824–1907), of kelvin temperature fame, said about models: 'I am never content until I have constructed a mechanical model of the subject I am studying. If I succeed in making one, I understand; otherwise I do not.'

A model is an aid to understanding. But a model is not the same as the thing itself. A model is a model. The thing to which a model refers is the thing to which it refers.

Units and standards

Physicists are required to make very careful and accurate observations and measurements of physical quantities. Good measurements are crucial to good science. It is the responsibility of experimenters to know about measurement and to cover all possible ways in which a measurement could fail to be accurate and precise.

An internationally agreed system of units called the Système Internationale or SI system is used in science. It provides definitions of quantities, lists the most up-to-date values for important quantities, codifies measurement theory and practice, and provides standards for the reporting of measurements. No country uses only SI units. If it did, it would use kiloseconds instead of hours.

BUREAU INTERNATIONAL DES POIDS ET MESURES

Visit the home page of the organisation that manages all international agreements about weights and measures. Look for SI base units, SI derived units and SI prefixes. You could even download and read the brochure that explains all.

Fundamental units

Seven units are defined for the fundamental or basic quantities – length, mass, time, current, temperature, luminous intensity and amount of matter.

The fundamental or basic SI units and their definitions are:

- *Length*: The unit of length is the metre (m), which is defined as 1 650 763.73 wavelengths of the orange-red line of the spectrum of ^{86}Kr (krypton) in a vacuum.
- *Mass*: The unit of mass is the kilogram (kg), which is based on a cylinder of platinum–iridium alloy kept by the International Bureau of Weights and Measures in Paris.
- *Time*: The unit of time is the second (s), which is defined as the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the ^{133}Cs (caesium) atom at 0 K.
- *Electric current*: The unit of current is the ampere (A), and this is defined as the current that, if maintained in two straight parallel conductors of infinite length and negligible cross-section, separated from each other by a distance of 1 metre in a vacuum, will produce a force equal to 2×10^7 newton per metre of length between the conductors.
- *Temperature*: The unit of temperature is the kelvin (K), which is defined as $\frac{1}{273.16}$ of the thermodynamic temperature of the triple point of water.
- *Luminous intensity*: The unit of luminous intensity is the candela (cd), which is defined as the luminous intensity in the perpendicular direction of a surface of $\frac{1}{600\,000}$ square metre of a black body at the freezing temperature of platinum (2042 K) under a pressure of 101 325 pascals.

- *Amount of substance*: The unit of amount of substance is the mole (mol), which is defined as the amount of substance of a system that contains as many elementary entities as there are atoms in 0.012 kilogram of carbon-12. (This number is approximately 6.023×10^{23}.)

These seven definitions are obviously quite impractical for everyday use, but they have been chosen carefully to provide invariable standards from which more practical devices can be manufactured; for example, a ruler or a stopwatch.

Derived units

Derived units are formed by combinations of the fundamental units. A simple example is the unit for area, the square metre (m^2). Other examples are:

- Volume – cubic metre (m^3)
- Speed – metre per second ($m\,s^{-1}$)
- Density – kilogram per cubic metre ($kg\,m^{-3}$)

A number of derived units have been given special names to commemorate notable scientists. These include frequency (hertz, Hz); force (newton, N); work and energy (joule, J); power (watt, W); current (ampere, A).

Prefixes

Consider the measurement of length. The metre is too large a unit with which to measure the thickness of this page. It is too small a unit to measure the distance to the Moon. For this reason, multiple or submultiple units may be formed by adding a prefix to the SI unit. The prefix is combined with the unit name and is written as one word.

millimetre (mm) equal to 10^{-3} m	megawatt (MW) equal to 10^6 W
centimetre (cm) equal to 10^{-2} m	kilogram (kg) equal to 10^3 g
kilometre (km) equal to 10^3 m	gigajoule (GJ) equal to 10^9 J

The preferred prefixes relate to the SI units usually by powers of three. The common prefixes are given in Table 12.1.

Table 12.1 Common prefixes

Multiple	Prefix	Symbol	Multiple	Prefix	Symbol
10^{18}	exa	E	10^{-2}	centi	c
10^{15}	peta	P	10^{-3}	milli	m
10^{12}	tera	T	10^{-6}	micro	μ
10^9	giga	G	10^{-9}	nano	n
10^6	mega	M	10^{-12}	pico	p
10^3	kilo	k	10^{-15}	femto	f

All measurements require at least a value and a unit. Some also require a direction. In more advanced measurements an indication of accuracy is also given. The value is often written in a standard form.

You should get into the habit of writing numerals and units correctly. The unit is separated from the numeral by a space. Each separate part of the unit is also separated by a space. For example, a speed of 5.3 metres per second is written as $5.3\ m\,s^{-1}$. Notice the space between 5.3 and 'm', and the space between 'm' and 's' (ms^{-1}, no space, means 'per millisecond').

Standard or scientific form

Standard or scientific form is widely used in science. It enables very large and very small numbers to be expressed relatively simply. Numbers are written as a number between 1.0 and 10 multiplied by the relevant power of 10. For example:

- The distance from Earth to the Sun is 150 000 000 km. In standard form this is written as 1.5×10^8 km.
- The average distance between two atoms is 0.000 000 000 16 m. This is written as 1.6×10^{-10} m.

In some cases, it is not sensible to write measurements in this form. For example, the measurement 0.8 m is better left as it is than written as 8×10^{-1} m. As a general rule, it is not usual to express the numbers between 0.01 and 1000 in standard form. It is not wrong to do it; it is just not the preferred method.

Dimensions

In the SI system, the unit of speed is the metre per second. This is not the only unit that can be used. Some alternative, equally correct units are mile per hour, kilometre per hour and centimetre per minute. All of these have one thing in common – each is a unit of length per unit of time. The quantity length per time is called the dimension of the property speed.

When writing the dimensions of a physical quantity, we use the symbols [L], [M], [T] and [I] to represent the dimensions of length, mass, time and current, respectively. Hence, we can write: $[\text{speed}] = [\text{length}]\,[\text{time}]^{-1} = [L]\,[T]^{-1}$, where the brackets are read as 'dimensions of'.

All of the quantities used in this book can be expressed in terms of a few fundamental quantities. Examples include $[\text{area}] = [L]^2$, $[\text{volume}] = [L]^3$, $[\text{density}] = [M]\,[L]^{-3}$.

The unit of all derived quantities can be found by substituting the fundamental unit for each dimension in the dimension equation. For example, in the SI system the unit of mass is the kilogram (kg) and the unit of length is the metre (m). The unit of density – dimension $[M]\,[L]^{-3}$ – then becomes $kg\,m^{-3}$ in the SI system, but $lb\,ft^{-3}$ in the British Imperial System.

The examination of the dimensions of an expression can be used to check whether an equation is likely to be correct. The dimensions of both the left-hand side and right-hand side of all equations must be dimensionally correct and equal, so that they balance. If they are not, *then the equation cannot be correct*. For example, the circumference of a circle, $C = 2\pi r$ has dimensions of length, [L] on both sides: [circumference] = [L], $[2\pi r]$ = [radius] = [L] – the constant, 2π has no dimensions. Hence, the equation is dimensionally correct. Note, however, that the equations $C = 2r$ and $C = \pi r$ are also dimensionally correct, but are in fact wrong. The dimension check will only tell us if the equation is wrong – not that it is necessarily correct.

Making and reporting measurements

When we make a measurement we have some idea that what we measure is relevant and appropriate to our interest. We may also expect there to be a **'true value'**, an exact number that represents what we are measuring. This cannot, however, be known with 100% certainty. Whenever a measurement is undertaken two things need to be guaranteed. The **measurand**, the quantity being measured, should be clearly specified. The **measurement result**, the best estimate of the 'true value', should be both **accurate** and **precise**.

Measurand

Mostly, it will be obvious which quantity is to be measured: length on a ruler, mass on a weighing machine, time on a clock, current on an ammeter, potential difference on a voltmeter. In student-designed practical investigations it may be necessary to decide what to measure: vertical angle of a cone, 'squash' of a crumple zone, extension of a spring, density of a fluid, rate of temperature change of a solid. These quantities are generally measured using length, mass and time quantities. Sometimes, the measuring device obscures this. Digital meters often rely on length, mass and time to produce the number on the dial, but this is not nearly as clear as on an analogue meter. For example, a digital radar gun measures the time between successive sound waves and internally converts this to a speed (distance/time). An analogue thermometer has an obvious, linear scale printed on it.

Many students think that because they get a numerical readout on a digital meter it is more accurate than an analogue meter. In general, this belief must be questioned. When comparing the readings from an analogue meter and a digital meter, the analogue meter is often more accurate. For example, if both give a measurement to the second decimal place, the uncertainty for the analogue meter will be a maximum of ± 0.005. For the digital meter, the uncertainty is ± 0.01.

Measurement result

The measurement result is the best estimate of the 'true value'. Except for discrete, countable things, the 'true value' is an ideal that can never be completely and unambiguously known. It is logically impossible to know the exact value of a continuous variable, even for standard values, such as the speed of light.

Measurand is a specified quantity to be measured.

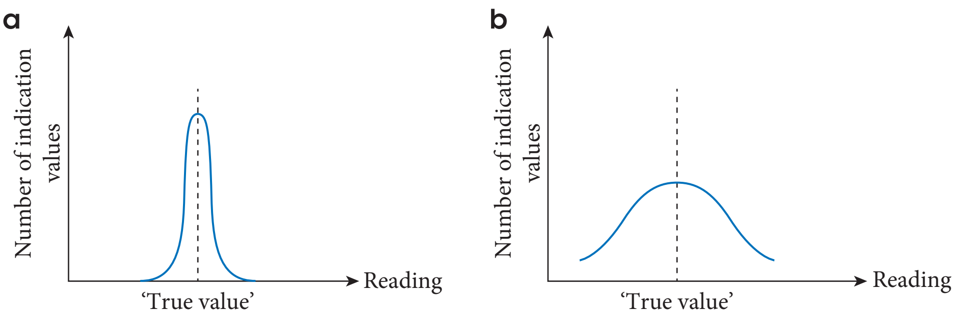
Measurement result is the best estimate of the 'true value' of a measurand.

'True value': for continuous variables this is an unknowable, ideal value that represents the measurand.

It is possible to make better or worse estimates of the 'true value'. This depends on the environment, the quality of the equipment and the skill of the person making the measurement.

Accuracy

An accurate measurement result is one that represents the 'true value' of the measurand as closely as possible. How can this be achieved? Every measurement provides an indication of the 'true value'. If we take repeated measurements, there will always be a spread of these **indication values** or results. However, the mean of these indication values should be a very good estimate of the 'true value'. Notice this is an agreed procedure. We agree that the mean is the most likely or best estimate of the 'true value'. It is not certain that it is the 'true value'. A plot of the number of indication values versus reading shows the spread of results around the supposed 'true value'.

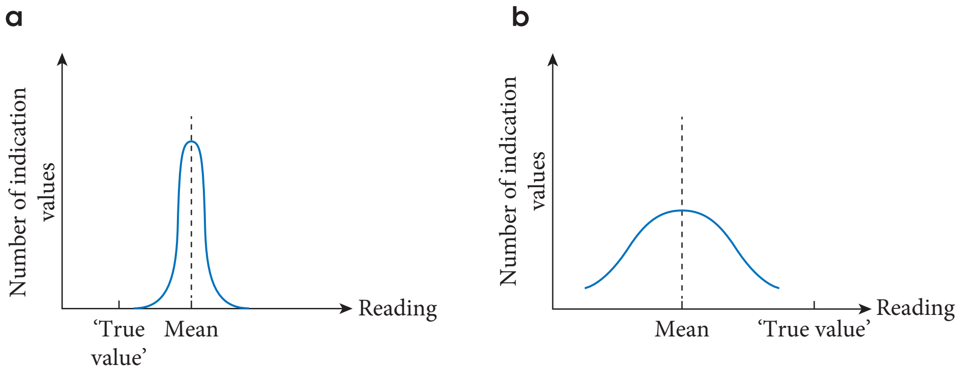


◀ **Figure 12.3**
In a plot of number of indication values versus reading, results may: a) cluster close to the 'true value', or b) be spread significantly around the 'true value'.

The means of Figures 12.3(a) and (b) are the same. The two measurement results are identical. Both report the same mean as the best estimate of the 'true value'. In this sense, both measurement results are accurate.

It is possible to get plots that look similar yet not achieve a proper measurement result for the 'true value'. If the measurand is specified incorrectly, then the measurement result will cluster around another value that is not the 'true value'. This is shown in Figure 12.4.

Figure 12.4 ▶ In a plot of number of indication values versus reading, a) many results cluster close to a value that is not the 'true value'. b) The results may be spread significantly around a value that is not the 'true value'.



The means of Figures 12.3(b) and 12.4(b) are the same. The two measurement results are identical. Both report the same mean of the indication values, but they do not represent the best estimate of the 'true value'. In this sense, both measurement results are inaccurate.

Precision

Precision relates to the skill of the experimenter within the environment and the quality of both the equipment and the measuring techniques used. If the quality of these is high, the measurement result is highly precise; if the quality is low, the result is imprecise. Figures 12.3(a) and 12.4(a) show precise measurement results because the individual indication values cluster closely around the mean. But Figures 12.3(b) and 12.4(b) show imprecise measurement results because the individual indication values spread significantly around the mean.

Accurate: value that represents the 'true value'

Precise: small spread around a mean value

Accurate and precise: small spread around the 'true value', which is taken to be the mean of indication values

Uncertainty

No measurement is exact. There are always effects that contribute to each measurement of a quantity being a bit different. They add to the **uncertainty** with which an indication value or measurement result can be reported. Good experimental observers always ensure that they can estimate the uncertainty, which is the range of values between which they are confident the 'true value' lies. Figures 12.3(a) and (b) both show the dispersion of indication values around the 'true value'. This dispersion is the result of random effects such as small air currents or localised temperature changes. This dispersion of indication values is called **random error**, although this term is starting to be replaced by more precise concepts. The term 'random error' alerts us to the possibility that, if the dispersion of the results is not considered, a mistake will be made in reporting the value correctly.

It is not possible, in principle, to predict the next indication value from the previous measurement. If it were, the effect could be taken into account in making the estimate of the 'true value'. Figures 12.4(a) and (b) show the mean of the indication values, which are offset from the 'true value'. Some of this offset can be accounted for by careful consideration of the situation and the measurement activities. These **systematic errors** can be identified and indication values adjusted for these known, regular effects. For example, when you weigh ingredients for a recipe, the scales may read 200 g before you put the ingredients in the bowl. After weighing the ingredients you would adjust the result by taking 200 g from the scale measurement to work out the actual quantity. This is referred to as calibration. Often calibration requires correction for a **zero error**, as in this example. The term, systematic error, alerts us to the possibility that, if they are not considered, a mistake will be made in reporting the value correctly. When drawing graphs of data, you should never force a line of best fit through the origin, even if you believe it should pass through the origin. A line of best fit that does not pass through the origin when it is expected to do so may be alerting you to a systematic error in your data. The systematic error should not be ignored.

The uncertainty in a measurement is a quantity that makes apparent the quality of the measurement. Some people refer to measurement uncertainty as measurement error. This is unhelpful. An error is a mistake. There should be no mistake in making the measurement. The uncertainty represents a careful estimate of doubt about the value.

Reporting uncertainty

Any measurement should be reported with two numbers: the **best estimate** of the value observed in the measuring system, and an estimate of how much uncertainty there is in this value. For example, the radius of a small ball can be measured with greater or less precision by different measuring devices. Using a ruler, its diameter might be recorded as (2.5 ± 0.1) cm. A different system might enable the diameter to be recorded as (2.513 ± 0.001) cm. Both these results would be accurate if there were no zero error. In general, the measurement with the least uncertainty is the most precise.

Managing uncertainty in practice

Measurements are used to derive new quantities. The best estimate of the measurand is used, for example, to fix a point on a graph or substitute into an equation. The uncertainty comes along as well. We need to manage the effect of this uncertainty on the overall quantities produced from such operations.

Precision of a measurement

We use place value to report the precision of a measurement. For (2.5 ± 0.1) cm, the final place in the value, the 5 in the tenths column for 2.5, is uncertain by 0.1 cm. We are saying, 'The best estimate of the value is 2.5 cm, and we are confident the 'true value' lies between 2.4 cm and 2.6 cm'. Similarly, for the more accurate and precise measurement (2.513 ± 0.001) cm, we are reporting confidence that the 'true value' lies between 2.512 and 2.514. The measurement is uncertain in the thousandths column.

The number of **significant figures** in a measurement can be found by counting the number of reported digits. There are rules for deciding on the number of significant figures in a reported value:

1. Zeros in front of the integer part of a numeral are not significant; for example, 00346 has 3 significant figures.
2. All non-zero figures are significant, for example in 25.4 there are 3 significant figures.
3. All zeros between non-zero digits are significant, for example in 203.4 and 27.6002 the zeros are significant. They have 4 and 6 significant figures respectively.
4. All zeros to the right of a decimal point, which follow a non-zero digit, are significant, for example in 21.000 the zeros are significant (5 significant figures).
5. For numbers less than 1, the zeros before the first non-zero digit are not significant, for example, 0.003 682 has 4 significant figures, starting at the digit 3.

Numbers between zero and 0.01 and numbers between 10 and 1000 are not usually written in standard form. Between 10 and 1000, a zero in the units column is then treated as significant. Thus, 200 and 850 both have 3 significant figures.

Some examples are shown below (the number of significant figures is reported in the bracket).

700 046 (6), 901.040 (6), 5403.2, (5), 350 (3), 67.2, (3), 64.0 (3), 0.409 (3), 0.0038 (2)

Other than 100–1000, a number such as 350 000 has 2 significant figures, because it can be written as 3.5×10^5; however, if written as $3.500\,00 \times 10^5$ it has 6 significant figures.

WORKED EXAMPLE 12.1

How many significant figures are in the following numbers:

a 0.0098 (1 mark)
b 758.0 (1 mark)
c 650 (1 mark)
d 3.0002 (1 mark)
e 21 000 (1 mark)

Answers	Logic	
a 2	Use the rules for significant figures.	1 mark
b 4		1 mark
c 3		1 mark
d 5		1 mark
e 2 (2.1×10^4)		1 mark

Try these yourself

How many significant figures are there in the following numbers?

a 0.0241 (1 mark)
b 6.098 (1 mark)
c 0.0045 (1 mark)
d 0.006 20 (1 mark)
e 650 000 (1 mark)

Relative and percentage uncertainty

Uncertainties can be reported as absolute or relative uncertainties. The examples above are all absolute uncertainties. They always have the same units as the measurement and are usually given to only 1 significant figure.

Uncertainties can also be given as relative (proportional or fractional) or percentage uncertainty. **Percentage uncertainty** is calculated by finding the **relative uncertainty** or **proportional uncertainty**, then converting it to an equivalent fraction out of 100:

$$\text{Relative uncertainty} = \frac{\text{uncertainty}}{\text{value}}$$

$$\text{Percentage uncertainty} = \frac{\text{uncertainty}}{\text{value}} \times \frac{100}{1}\%$$

Relative and percentage uncertainties are usually given to two significant figures. They have no units because they are the ratio of the absolute uncertainty in the measurement. The absolute uncertainty and the measurement have the same units, so when you divide one by the other, the units cancel out.

WORKED EXAMPLE 12.2

Calculate the relative and percentage uncertainties for these values.

a 2.5 ± 0.1 cm (2 marks)

b 2.513 ± 0.001 cm (2 marks)

Answers	Logic	
a Relative uncertainty $= \frac{0.1\text{ cm}}{2.5\text{ cm}} = 0.04$	Use the correct fraction, calculation and number of significant figures.	1 mark
Percentage uncertainty $= \frac{0.1\text{ cm}}{2.5\text{ cm}} \times \frac{100}{1}\% = 4\%$	Use the correct fraction, calculation and number of significant figures.	1 mark
b Relative uncertainty $= \frac{0.001\text{ cm}}{2.513\text{ cm}} = 0.00040$	Use the correct fraction, calculation and number of significant figures.	1 mark
Percentage uncertainty $= \frac{0.001\text{ cm}}{2.513\text{ cm}} \times \frac{100}{1}\% = 0.040\%$	Use the correct fraction and calculation.	1 mark

Try these yourself

1 Calculate the relative and percentage uncertainties for these values.

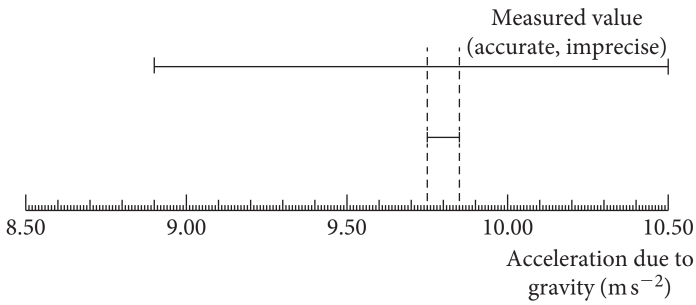
- **a** (2.5 ± 0.1) s (1 mark)
- **b** (0.0524 ± 0.003) μm (1 mark)
- **c** (250.0 ± 0.8) kg (1 mark)

2 In radiation counting, the uncertainty in N counts is $\pm\sqrt{N}$. Calculate the percentage uncertainties for these counts.

- **a** 100 counts (1 mark)
- **b** 1000 counts (1 mark)
- **c** 10 000 counts (1 mark)

Proportional error

Every measurement, including highly precise measurements, should be reported with an uncertainty value. When two measurements of the same quantity are compared, the range of uncertainties must be considered. For example, the accepted value for the acceleration due to gravity near Earth's surface might be reported at a location to be (9.80 ± 0.02) m s^{-2}. This means that the value could be anywhere between 9.78 m s^{-2} and 9.82 m s^{-2}. In a laboratory experiment, a student measured the acceleration due to gravity to be (9.7 ± 0.8) m s^{-2}. This means that the student's value could be anywhere between 8.9 m s^{-2} and 10.5 m s^{-2}. The student's result is clearly less precise than the accepted value. However, the student has managed to measure a value that fits within the range of uncertainty of the accepted value. With a less precise measurement, the student has nevertheless produced a confirming instance of the same value as the accepted value. The accepted value and the student's value both encompass a region in which the 'true value' could actually lie. This is within the overlap region, between 9.78 m s^{-2} and 9.82 m s^{-2}. Hence the two values agree.



◀ **Figure 12.5** The uncertainty of a student-measured value overlaps that of the accepted value. The two values agree.

Some measurements are extremely precise, well beyond the precision generally possible in a school laboratory. These accepted values can be considered *as though* they are 'true values'. For example, the charge on an electron is $1.602\,176\,565 \times 10^{-19}$ C. The uncertainty is $0.000\,000\,035 \times 10^{-19}$ C. The relative uncertainty is an incredibly precise 2.2×10^{-8}.

In these cases, the concept of **proportional error** can be defined. It is the difference between a measurement result and an accepted value, expressed as a fraction of the accepted value:

$$\text{Proportional error} = \frac{|\text{measured value} - \text{accepted value}|}{\text{accepted value}}$$

Proportional error is a way for students to compare their measurement result with an accepted value.

Percentage error is the ratio of the magnitude of the proportional difference between an accepted value and a measurement result, expressed as a percentage:

$$\%\ \text{error} = \frac{|\text{measured value} - \text{accepted value}|}{\text{accepted value}} \times \frac{100}{1}\%$$

Proportional error and percentage error are useful for helping students compare their accuracy with far more precisely measured quantities; however, neither is defined in the international measurement standards from the Bureau International des Poids et Mesures (BIPM).

PHYSICAL CONSTANTS

Search for particular constants, such as the electronic charge, speed of light and mass of a proton.

WORKED EXAMPLE 12.3

A manufacturer gives the wavelength of a light as $(671 \pm 5) \times 10^{-9}$ m. A student measured the wavelength to be $(660 \pm 9) \times 10^{-9}$ m.

a Does the student's measurement result fit within the accepted value measurement result? Sketch a graph of range of indication values versus reading to show any differences between the measurement results. (5 marks)

b Calculate the percentage error in the student's measurement result. (2 marks)

Answers	Logic	
a The two measurement results overlap in the region $(666–669) \times 10^{-9}$ m.	Identify the overlap.	1 mark
	Label the horizontal axis and mark the correct overlap values.	4 marks

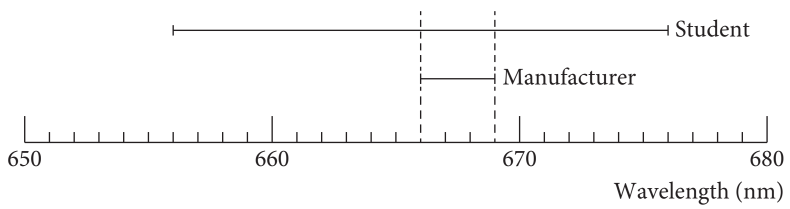


Figure 12.6 Overlap of measurement results for manufacturer's accepted value and student measurement result

b

$$\%\ \text{error} = \frac{|\text{measured value} - \text{accepted value}|}{\text{accepted value}} \times \frac{100}{1}\%$$

Substitute the correct values into the formula. — 1 mark

$$\Rightarrow \%\ \text{error} = \frac{|660\,\text{m} - 671\,\text{m}| \times 10^{-9}}{671 \times 10^{-9}\ \text{m}} \times \frac{100}{1}\%$$

$$\Rightarrow \%\ \text{error} = \frac{11}{671} \times \frac{100}{1}\%$$

$$\Rightarrow \%\ \text{error} = 1.6\%$$

Calculate the answer. — 1 mark

As both measurements incorporate the region in which the 'true value' may be meaningfully said to exist, this figure is of little relevance. This is why the BIPM does not define percentage error.

Try these yourself

The colour code on a resistor claims the resistance to be $1000 \pm 50\,\Omega$. A student who uses a specially designed circuit to determine the resistance finds the value to be $1025 \pm 15\,\Omega$.

a Does the student's measurement result agree with the accepted value of the measurement result? (2 marks)

b Calculate the percentage error in the student's measurement result. Comment on its meaningfulness. (2 marks)

Effect of uncertainty in derived quantities

Measured **raw data** are used to find **derived data**. Any uncertainty in the raw data must be taken into account in reporting derived data. In many instances data are given without uncertainty values. In these cases, the last decimal place represents the first uncertain figure. The number of significant figures in each figure is the number of digits, taking into account the rules for significant figures.

Adding and subtracting raw data

When adding and subtracting numbers with uncertainties, it is useful to take advantage of place value. Perform the operation with the best estimate values. Add the individual uncertainties to find the uncertainty in the sum or difference. Claim the best estimate only as far as the sum of uncertainties will allow; that is, the place value of the first digit of the sum of the uncertainties. For example, consider the following sum:

$$(17.23 \pm 0.02) + (5.1 \pm 0.4)$$

Add the best estimates of the value: $17.23 + 5.1 = 22.33$
Add the uncertainties for each value: $0.02 + 0.4 = 0.42$

The answer could be given as 22.33 ± 0.42, but the result becomes uncertain at the first decimal place in the uncertainty, 4 in 0.42. Hence, the best estimate, with uncertainty, of the sum is 22.3 ± 0.4.

When there is no uncertainty given in a data point, the *last decimal* place in each number is regarded as the first uncertain figure. Then the result must have no more decimal places than the number with the least number of decimal places. The addition or subtraction is performed with the numbers as given, then rounded up from 5 or down from 4. For example:

$$\begin{array}{r} 32.2187 \\ +\ 126.3 \\ +\ 3.132 \\ \hline 161.6507 \end{array}$$

This is rounded up (0.65 to 0.7) to 161.7 because the number 126.3 is given to only one decimal place.

$$\begin{array}{r} 3052.3 \\ -\ 235 \\ \hline 2817.3 \end{array}$$

This is rounded down (0.3 to 0) to 2817 because the number 235 is given to the nearest whole number.

Multiplying and dividing raw data

When multiplying and dividing numbers with uncertainties, perform the operation using the best estimate values. Add the individual relative or percentage uncertainties to find the relative or percentage uncertainty in the product or quotient. To find a value for the uncertainty, the percentage or proportional uncertainty is used. For example, consider the following product:

$$(50.6 \pm 0.8) \times (123.63 \pm 0.91)$$

Multiply the best estimates: 50.6×123.63

$= 6255.678$

Find the percentage uncertainties for each number:

$$= \frac{0.8}{50.6} \times \frac{100}{1}\% \quad \text{and} \quad = \frac{0.91}{123.63} \times \frac{100}{1}\%$$

$$= 1.581\% \qquad\qquad = 0.736\%$$

Add the percentage uncertainties:

$$= 1.58\% + 0.736\%$$

$$= 2.371\%$$

$$= 2.3\% \text{ by convention}$$

The product can be given as (6255.68 ± 2.3)%, but this hides a problem. The product has far more significant figures than the individual raw data numbers from which it was calculated. That is, the derived quantity is claimed to be more accurate than the raw data from which it was derived! That cannot be.

Let us look at the actual value of the uncertainty:

$$2.3\% \text{ of } 6255.68 = 143.88$$

The result could now be reported, inaccurately, as (6255.68 ± 143.88). The result, 6255.68, is shown to be uncertain in the hundreds column. Using standard form to express the significant figures, the result should now be written, correctly, as:

$$(6.2 \pm 0.1) \times 10^3$$

As long as the raw data measures what is intended, this is an accurate and precise statement of the derived quantity. Notice, however, that the derived quantity has 2 significant figures. This is less than the number of significant figures in the number with the least significant figures, namely 50.6 (3 significant figures). The number of significant figures in a derived quantity is always equal to or less than the number of significant figures in the least precise piece of raw data. In experiments, physicists try to ensure that uncertainties in the raw data do not accumulate to the point where derived quantities become meaningless. This takes high-level thinking, planning and effort.

When there is no uncertainty given in the data, calculations are performed using the unrounded data. The result is then rounded to a figure that has the same number of significant figures as the data with the *least number of significant figures*.

Two examples follow.

i	45.71	(4 significant figures)
	× 34.1	(3 significant figures)
	1558.711	(7 significant figures on calculator)

This is rounded to 1.56×10^3 (3 significant figures) because 34.1 has the least number of significant figures (i.e. 3).

ii $\frac{5465.48}{2.4}$ (a 6 significant figure number divided by a 2 significant figure number)

= 2277.283333 (10 significant figures on calculator)

= 2.3×10^3 (2 significant figures)

This is rounded to 2.3×10^3 because 2.4 has the least number of significant figures (i.e. 2).

Numerical methods for calculating uncertainties involving functions

Sometimes, measurements are used in functions such as sin, cos and tan. An uncertainty value can be found numerically, rather than by applying the rules outlined above. Consider the way the function works before deciding how to proceed.

WORKED EXAMPLE 12.4

A student conducts an experiment to find the refractive index of an organic biochemical. Angles of incidence, *i*, and angles of refraction, *R*, are measured several times and an average, with uncertainty, is found for each. The refractive index is then calculated from the equation:

$$n = \frac{\sin i}{\sin R}$$

Use the following data to compute the refractive index, with uncertainty:

$i = (43.6 \pm 0.5)°$, $R = (32.1 \pm 0.5)°$ (6 marks)

Answer	Logic	
Calculate the refractive index from the best value estimate: $n = \frac{\sin(43.6°)}{\sin(32.1°)} = 1.298$	Substitute the known values and calculate the answer.	2 marks
Decide what is the worst case possible. As the sine function is increasing in the range of possible values, calculate: $n = \frac{\sin(43.6 + 0.5°)}{\sin(32.1 - 0.5°)} = \frac{\sin(44.1°)}{\sin(31.6°)} = 1.328$	Substitute the known values and calculate the answer.	2 marks
The difference between these two values is: 1.328 − 1.298 = 0.030. This shows the second decimal place is uncertain, so the answer must not go beyond the second decimal place.	Calculate the difference.	1 mark
Thus, after rounding, the refractive index, *n*, of the liquid is 1.30 ± 0.03.	Give the answer and its uncertainty.	1 mark

CHAPTER SUMMARY

Model

- Models are used to describe, explain, relate and predict phenomena.
- Models can be represented by words, images, mathematics, or physical constructions.
- Models help scientists to frame physical laws and theories.
- Laws and theories are models of the world.
- Models change: Data and ideas affect models.
- Falsifiability is the process that uses a critical test to determine whether a model needs to be replaced or constrained.

Units and standards

- SI units are an internationally agreed consistent set comprising seven fundamental units: length (m), mass (kg), time (s), electric current (A), temperature (K), luminous intensity (cd) and amount of a substance (mol).
- All units for other quantities are derived from these units.
- Standard or scientific form: Numbers less than 0.01 and greater than 1000 are written as a number between 1 and 10 multiplied by 10 raised to an integral power.
- Dimensional analysis can be used to:
  - check if an equation is possibly correct.
  - determine units in different systems.

Measurement

- Purpose: To find a value for a quantity of interest (measurand)
- Value is never exact. It is affected by system, procedure, operator skill and environment.
- Experimenter responsibility: All systematic errors should be quantitatively specified and measurements adjusted accordingly. The effect of random errors should be reported in the value reported for the measurand. No errors, that is, no mistakes, including mistakes about systematic and random errors, should be allowed to affect results.
- Mean of indication values is the best estimate of the 'true value'.
- Uncertainty is the measure of the dispersion of indication values around the mean; hence, it is an estimate of the quality of an indication value or measurement result.

Managing uncertainty in practice

- Significant figures: Place value is used to show the significance of a measure. The number of digits in a measurement is an indication of its accuracy.
- Percentage uncertainty is the relative uncertainty expressed as a percentage:

$$\text{Percentage uncertainty} = \frac{\text{uncertainty}}{\text{value}} \times \frac{100}{1}\%$$

- Percentage error is a quantity produced for learning purposes, but which is not defined by the BIPM:

$$\%\ \text{error} = \frac{|\text{measured value} - \text{accepted value}|}{\text{accepted value}} \times \frac{100}{1}\%$$

- Raw data are directly measured quantities, including uncertainty. Derived data is computed from raw data. The effect of uncertainty in raw data is accumulated in derived quantities.

CHAPTER GLOSSARY

accurate the degree to which a measurement result approaches the 'true value'

best estimate value chosen to represent the indication value or measurement result of a measurand

critical test an experiment that, having been rigorously conducted, shows that one or other of competing theories is false

derived data data that is deduced from raw data by mathematical manipulation, such as graphs, algebraic equations and geometric constructions

explanation generalised account of why a body of data occurs

falsifiability principle used to determine the experimental data that would disprove a model, law or theory; data from a critical test

indication value a single result of a measurement; the indication value gives a hint as to the 'true value'

measurand quantity being measured

measurement result best estimate of a 'true value'; numerical value based on judgements about one or more attempts to measure the 'true value'

model a representation of a system or phenomenon that explains the system or phenomenon. A model may be mathematical equations, a computer simulation, a physical object, words or other form

percentage error proportional error expressed as a percentage; not defined by BIPM

percentage uncertainty proportional uncertainty, expressed as a percentage

precise the degree to which individual measurements cluster around the mean

proportional error difference between a measurement result and an accepted value, expressed as a fraction of the accepted value; not defined by BIPM

proportional uncertainty relative uncertainty

qualitative non-numerical data; descriptive information

quantitative numerical data; specific amount

random error a variation that affects a measurement in a random way so that the measurement is as likely to change in any one direction as in any other

raw data original data taken directly from a measurement system

relative uncertainty ratio of uncertainty to value

representation model of reality

significant figure digit reported in a measurement result; the number of significant figures is the number of meaningful digits in a measurement result

systematic error an error that results in a consistent, predictable offset from the 'true value'; for example, a zero error

true value the exact value of a measurand; the 'true value' is an ideal that can never be known with certainty

uncertainty estimate of the range of values within which the 'true value' of a measurement or derived quantity lies

zero error scale is not zero when measurements are taken; also called a calibration error

Shutterstock.com/Kozorez Vladislav

CHAPTER 13 SCIENTIFIC INVESTIGATIONS

By the end of this chapter you will have covered the following material.

Science Inquiry Skills

- Identify, research and construct questions for investigation; propose hypotheses; and predict possible outcomes (ACSPH045)
- Design investigations, including the procedure to be followed, the materials required, and the type and amount of primary and/or secondary data to be collected; conduct risk assessments; and consider research ethics (ACSPH046)
- Conduct investigations, including the manipulation of devices to measure motion and the direction of light rays, safely, competently and methodically for the collection of valid and reliable data (ACSPH047)
- Represent data in meaningful and useful ways, including using appropriate SI units and symbols; organise and analyse data to identify trends, patterns and relationships; identify sources of random and systematic error and estimate their effect on measurement results; identify anomalous data and calculate the measurement discrepancy between the experimental results and a currently accepted value, expressed as a percentage; and select, synthesise and use evidence to make and justify conclusions (ACSPH048)
- Interpret a range of scientific and media texts, and evaluate processes, claims and conclusions by considering the quality of available evidence; and use reasoning to construct scientific arguments (ACSPH049)
- Select, use and interpret appropriate mathematical representations, including linear and non-linear graphs and algebraic relationships representing physical systems, to solve problems and make predictions (ACSPH051)
- Communicate to specific audiences and for specific purposes using appropriate language, nomenclature, genres and modes, including scientific reports (ACSPH052)

Introduction

Performing investigations is your chance to experience what doing science is really like. Science is about finding things out through observation and experiment, which is what doing investigations is all about. This is why investigations are central to science, *and* why they are so much fun.

Newspix/Norbert Von Der Heidt

Figure 13.1 ▲ Student working on an investigation into rocket propulsion

Sometimes an important advance in science begins with a casual observation or a lucky accident. This was the case when Davisson and Germer first observed electron diffraction after an accident with a vacuum system. This eventually led to advances in quantum theory and to devices such as electron microscopes. However these developments could not have followed if Davisson, Germer and many others had not carried out further investigations. This sort of lucky accident may begin a new field of research, but it then proceeds by carefully planned investigation.

Scientific investigations can take years to complete and may involve collaboration among many scientists. They may require access to special equipment in Australia or overseas. They may cost a lot of money, sometimes millions of dollars, to complete. Hence scientists invest time in *planning* investigations before they begin. When scientists apply for grants to carry out investigations they need to show that they have carefully planned what they will do and how any money provided will be spent. Good planning is crucial to the success of the investigation.

They then make careful *measurements and observations* and record their *results*. They *keep records* of all their experiments. This is a legal requirement. Typically experimental results need to be kept for 5–7 years. There are also requirements on how and where data is stored.

Once data is collected it needs to be *analysed*. There are various ways this is done, but in the physical sciences (physics, chemistry and geology) it almost always involves constructing graphs. Once a relationship is established graphically, an algebraic relationship can be derived.

Finally, the results of the investigation must be *communicated*. Usually this involves publishing a scientific paper either in a journal or conference proceedings. It often includes presenting the results in talks or posters at conferences. If the result is funded by a grant then a research report must be submitted. If the results are really exciting, then the scientists may write a media release. However the results are communicated, this step must happen for the investigation to be completed.

When *you* perform scientific investigations you will also need to plan carefully. You will find out about the topic by reading about what other people have done. You will collect data from your own experiments and secondary sources. You will then analyse that data and draw conclusions about what it means. Finally, you will communicate your findings. Each of these steps is outlined below.

Planning your investigation

There are many things to consider when planning an investigation. You need to think about how much time you will have inside and outside class. You will also need to think about what space and equipment you will need and where you will go if you want to make measurements or observations outside.

You may be working in a group or on your own. Most scientists work in groups. If you can choose who you work with, think about it carefully. It is not always best to work with friends. Think about working with people who have skills that are different from your own.

Finally, and probably the first thing that most students think about, is the topic of the research. You will need to come up with a **research question** or **hypothesis**.

Choosing a research question

Obviously, it is a good idea to investigate something that you find interesting. If you are working in a group try to find something that is interesting to everyone in the group.

A good way to start is by 'brainstorming' for ideas. This works whether you are working on your own or in a group. Write down as many ideas as you can think of. Don't be critical at this stage. Get everyone in the group to contribute and accept all contributions uncritically. Write every idea down.

After you have run out of ideas, it is time to start being critical. Decide which questions or ideas are the most interesting. Think about which of these it is actually possible to investigate given the time and equipment available. Make a shortlist, but keep the long list too for the moment. Once you have your shortlist it is time to start refining your ideas.

Alamy/Jim West

▲ **Figure 13.2**
Brainstorm as many ideas as you can in your group.

Researching and refining your question

The next step is to find out what is already known about the ideas on your list. Use the Internet, your text books and the library to find out. Make sure you *keep a record* of the information that you find as well as *the sources*. You should start a **logbook** at this stage. You can write in references, or attach printouts to your logbook. This can save you a lot of time later on! Many research students forget to do this when they first start reading about their topic and then have to search all over again.

> Good record keeping is important in scientific research, and it begins at this stage of the investigation.

Be critical of what you read. Do not assume that everything you read online or even in books is true. Try to find **reliable** sources of information. Textbooks, websites from universities and government research agencies are usually very reliable. Publications and web pages from professional associations, such as the Australian Institute of Physics and equivalent international organisations are also good sources. Blogs and homepages of other students are not usually reliable, although they are useful to give you ideas. Websites that are trying to sell you something should also be treated sceptically. Talk to your teacher about sources of information as well. They will be able to tell you if a website is reliable, and suggest sites that they know are suitable.

Shutterstock.com/Ermolaev Alexander

▲ **Figure 13.3**
Start researching your topic and make sure you keep a record of all your references.

You may find examples of similar investigations to the one you are thinking of. It is a good idea to look at these, so you can learn from the experience of other researchers. However, in general, it is better not to try to replicate someone else's investigation exactly. If you do decide to replicate someone else's investigation then you need to acknowledge and carefully **reference** their work. (See the section on referencing on page 441). If you do not do so, it is **plagiarism**. This is a very serious form of academic misconduct. Talk to your teacher about how original your research needs to be, and how closely it can be based on someone else's work. It is much better to do this at the start than to be accused of cheating later on!

Finally, talk to your teacher about your ideas. They will be able to tell you whether your ideas are likely to be possible given the equipment available. They may have had students with similar ideas in the past and can make suggestions.

After you have researched your questions and ideas, you will hopefully be able to narrow the shortlist down to the one question that you want to tackle. If none of the questions or ideas look possible (or still interesting), then you need to go back to the long list.

Proposing a research question or hypothesis

AUSTRALIAN INSTITUTE OF PHYSICS

This site contains a lot of useful information on physics-related matters.

AMERICAN INSTITUTE OF PHYSICS

This is a useful resource for keeping up with physics news.

INSTITUTE OF PHYSICS

This is another useful resource on physics-related matters.

Once you have decided on what you will investigate you need to turn it into a research question or a hypothesis.

A research question is one that can be answered by performing experiments or making observations. A hypothesis is a prediction of the results of an experiment, which can be tested by performing experiments or making observations.

You may also be able to do a design, build and test project. These are described later. Make sure you check what sort of investigation or project you are supposed to be doing with your teacher.

Research questions

A research question may be of the form 'How does the height attained by a bottle rocket depend upon the volume of water in the rocket?'. The aim of your research is then to answer the question.

You need to frame the question carefully. It needs to be specific enough that it guides the design of the investigation. A specific question rather than a vague one will make the design of your investigation much easier. Asking 'what volume of water gives the maximum height for a water rocket?' tells you what you will be varying and what you will be measuring. It also gives a criterion for judging whether you have answered the question.

Asking 'How can we make a water rocket fly the best?' is not a good question. This question does not say what will be varied, nor does it tell you when you have answered the question. 'Best' is a vague term. What you mean by 'best' may not be what someone else means.

A good research question identifies the **variables** that will be investigated. Usually you will have one dependent variable and one independent or **controlled variable**. For a lengthy investigation you may have two or more independent variables. Variables are discussed in more detail later.

Finally, a good research question should be answerable with the time and equipment available.

Alamy/Marmaduke St John

Figure 13.4 ▲
You need to develop the question you are researching very carefully. These students are investigating the launch angle at which their water rocket will achieve maximum distance.

Hypotheses

A hypothesis is a tentative explanation or prediction not yet confirmed by experiment, such as 'The height attained by a water rocket will increase with the amount of water contained in the rocket'. Your hypothesis should give a prediction that you can test, ideally quantitatively.

A hypothesis is usually based on some existing model or **theory**. It is a prediction of what will happen in a specific situation based on that model. For example, kinematics describes the trajectory of any projectile. A hypothesis based on the kinematics model predicts the range of a specific projectile launched at a given angle and speed.

A hypothesis should give you a prediction that you can test by performing an experiment. This means it should at least be **falsifiable**. A good hypothesis should be able to be disproved. However, you will *not* generally be able to claim that you have proved your hypothesis.

If your experiments agree with predictions based on your hypothesis, then you can claim that they support your hypothesis. This *increases your confidence* in your model, but *it does not prove that it is true*. Hence an aim for an experiment should not start 'To prove …', as it is not possible to actually prove a hypothesis, only to disprove it.

If your experimental results disagree with your hypothesis, then you may have disproved it. This is *not* a bad thing! Often the most interesting discoveries in science start when a hypothesis based on an existing model is disproved. This means that the model it was based upon either is not a good model, or does not apply to the particular situation. You could then try to work out why the model does not apply, or try to formulate a better model. What to do when your hypothesis is not supported is discussed further in the analysis section.

In summary, a good research question is a question that is specific and can be answered by performing experiments and making measurements. A good hypothesis is a statement that predicts the results of an experiment and can be tested using measurements.

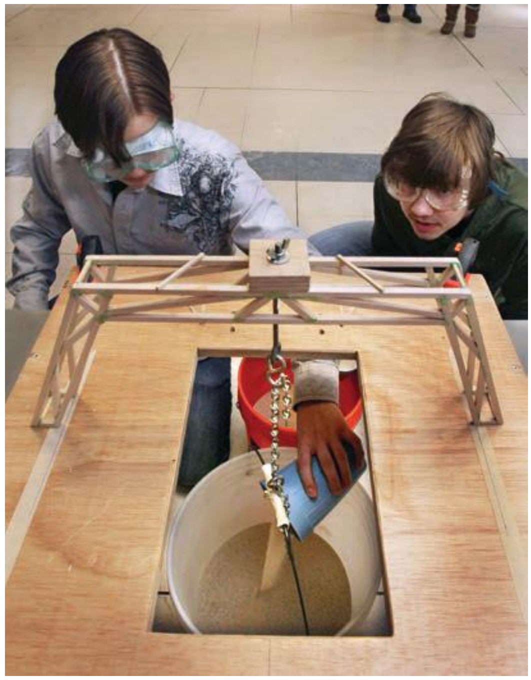
Even if your question or hypothesis meets these criteria, do not be surprised if you change or modify it during the course of your investigation. In scientific research, the question you set out to answer is often only a starting point for more questions.

Design briefs

Sometimes the aim of the investigation may be to design, build and test something. In this case, rather than a research question or a hypothesis, you will need a **design brief**. The design brief will specify what you intend to build, and some criteria by which to judge whether you have succeeded. You need a clear aim, and quantitative criteria by which to judge the final product.

The criteria will usually specify minimum performance characteristics of the end product. For example, if it is a water rocket, it may need some minimum range. A model bridge may need to span some distance and carry some minimum weight without breaking. The criteria may also include limits on what may be used or on the specifications of the product. For example, there may be a maximum weight or cost.

Just like a good research question or hypothesis, a well-written design brief tells you what to do, and how to know if you have succeeded.


The Image Works/©Syracuse Newspapers/D Lassman

▲ Figure 13.5
Testing how much weight a model bridge can carry

Designing your investigation

Once you have a specific research question or hypothesis you need to design your investigation. It is fun to start making measurements or observations immediately, but it is also important to spend time learning how to use the equipment, and experimenting to find the best way to set up your investigation. You may also discover that you need different or more equipment. This may save you time later on.

It is also important not to get distracted playing and forget the purpose of your investigation. At the end of the process, you need good data that answer your question or test your hypothesis. Having a plan allows you to ensure that you make the measurements that you need. The longer the investigation, the more important it is that you have a clear plan. There are several things to consider.

- What data will you need to collect?
- What materials and equipment will you need?
- When and where will you collect the data?
- If you are working in a group, who will collect the data?
- Who will be responsible for record keeping?
- How will the data be analysed?

The data that you collect will always include **secondary data**, and will usually include **primary data**. Secondary data is data that has been collected by someone else.

You will already have collected some secondary data when you investigated your research topic to formulate your question or hypothesis. You will probably want to collect more secondary data. If your topic is not one for which you can collect primary data, then you will need to rely on secondary data. Remember that when you collect secondary data it is important to use reliable, reputable sources.

Primary data is data that you collect yourself. You can collect data by performing experiments or making observations in the field. You should be able to measure distances, times, temperatures, forces including weights, potential differences and currents. You will have had practice at measuring some or all of these things already. You need to decide which variables you will measure and which variables you will control. Consider which variables you can control, and which you cannot.

If you are working on a design-build-test project then your primary data includes the performance tests of your product.

Consider how you will analyse the data. Will you need access to specific software such as a graphing or statistics package? If so, make sure that you know how to use it. If you are using software to draw graphs then you need to know how to produce a scatter graph and fit a **line of best fit** and add **uncertainty bars**. Note that a line of best fit is *not* the same as joining the dots. You should *never join the dots*, even though this is often the default setting in spreadsheet software. You should consult a reference guide, the 'help' menu for your software, or ask your teacher. Graphs are discussed in more detail in the analysis section on page 433.

Keep a record of your planning. This should go in your logbook. Writing down what you plan to do, and why, will help you stay focused during the investigation. If you are working in a group, then a record of what each person agrees to do during the investigation can be very important.

Variables and measurements

Anything that can vary in an experiment is a variable. An independent or controlled variable is one whose value you can control. For example, if you are doing an experiment to measure the voltage–current characteristics of an unknown circuit, then you would control the potential difference input into the circuit and measure the current in the circuit as a result. In this case the potential difference (or voltage) is the independent variable. The current, which is what varies as a result of the independent variable changing, is the dependent variable.

In the question 'What volume of water gives the maximum height for a water rocket?', the volume of water is the independent variable. The dependent variable is the height attained. A second possible independent variable, not mentioned in this question, is the air pressure inside the bottle. The air pressure should be kept constant, so it is not a variable in the investigation. If it was a long investigation, the air pressure could be a second controlled variable. If you decide to have two independent variables then it is important to keep one constant while you vary the other, if at all possible. Then you take multiple sets of measurements, keeping one variable at a fixed value for each set of data while you vary the other.

When variables have a numerical value, you make **quantitative measurements**. You measure that numerical value in the appropriate units. For example, you may measure a current of 15 mA or a height of 15 m.

Continuous variables may take any possible value, usually within some range. Length, time and current are continuous. In the water rocket example, the volume of water is a continuous variable, as it may take any value up to the volume of the bottle. A variable that may take only fixed values is called a **discrete variable**. Often these are whole numbers of things that cannot be broken into smaller parts, such as electrons or students. In the water rocket example, if the rocket has stabilising fins then the number of fins is a discrete variable.

Your measuring equipment will sometimes restrict you to only measuring discrete values. This is always the case with **digital** equipment. A set of digital scales that measures in grams gives you discrete values. It does not, however, mean that mass and weight are discrete variables. The weight of water in a bottle rocket is a continuous variable, but digital scales will only give you discrete measurements of the mass.

In some investigations you may use **qualitative measurements** or data. For example, a chemical reaction may lead to a colour change. You would usually describe the colour in words, such as 'pink' or 'green', rather than using a number. Sometimes you use a combination of qualitative and quantitative data. For example, you may describe the flight of a water rocket as reaching some maximum height in metres (quantitative) but following a spiralling path (qualitative).

Once you have decided on the variables you will be measuring you will be able to identify the equipment and other resources you will need.

Identifying the resources required

If you are going to collect primary data, make a list of all the equipment that you need. Consider how precise the measurements will need to be. If your hypothesis predicts a temperature change of 0.1°C, but you can only measure to a precision of 0.5°C, then you will not be able to test your hypothesis. You may need to think carefully about how you measure some things. For example, in a water rocket investigation, measuring the height attained can be very difficult. You may not be able to measure it directly.

If you are doing a design-build-test project then make sure you list all the materials and tools you will need. It is a good idea to allow for some extra materials in case mistakes are made during construction. For example, if you are testing a water rocket, the rocket could get stuck in a tree or lost. Consider who will supply the materials and how much they might cost. Scientists and engineers generally have tight budgets that they have to work within.

The equipment you plan to use must be safe. Will you need special protective equipment, such as lab coats, safety glasses or ear protectors? For a rocket project, you might need some temporary fencing or witches hats to mark off the launch area. There is a section on risk assessment on page 426. Make sure that you include any safety equipment needed in your equipment list.

Consider where you will perform your experiments or observations. Can you use normal classroom space, or do you need to be outside? If you are outside, what provisions can be made for ensuring that you can work without interference? Will you need to consider the convenience or safety of others? Talk to your teacher about what space is available.

When you have your list, talk to your teacher about what space and equipment is available. You might find that you need to modify your question or hypothesis at this stage.

Planning the experimental procedure

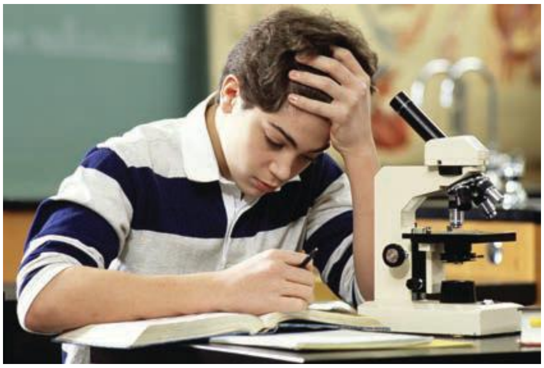
The most common problem that students have when doing research is time management. It is important to plan to have enough time to perform the experiments, *and* to analyse them, *and* to report on them. You also need to allow time to learn how to use the equipment if you have not used it before.

If you are doing a design-build-test project then allow plenty of time for all three stages. Often in this sort of project the design needs to be modified, sometimes several times. Building may not go as smoothly as you expect, and problems are often discovered during testing. If you allow plenty of time and start work early, you will be able to go back and modify your design if needed.

In any investigation you will need to collect reliable and precise data. You cannot do this if you do not know how to use the equipment. Always ask if you are unsure. Reading the user manual is also a good idea. It will usually specify the precision of the device, and let you know of any potential safety risks.

Whenever possible you should make repeat measurements, so allow time for this. This allows you to check that your measurements are **valid**. Valid results require that each independent variable gives similar results each time. If the results are similar each time, then your results are likely to be valid. If a result is not **reproducible**, it is probably not a valid result. A result is reproducible if you make exactly the same measurement more than once and get the same result, within the limits of experimental uncertainty. If a result is not reproducible, then a variable other than the one you are controlling is affecting its value. If this is the case, you need to determine what this other variable is, and control it if possible.

Think about how you can minimise uncertainties. Minimising uncertainty is not just about using the most precise equipment you can find, it is also about clever experimental technique. Very precise measurements are possible using simple equipment. For example, in 1862 Léon Foucault measured the speed of light with an uncertainty of 0.2%, without a computer, data logger or even



Getty Images/Gabe Palmer

Figure 13.6 ▲ Sometimes experiments just don't work.

a digital stopwatch. Remember that it is a poor workman who blames his tools! See the section on uncertainties on page 429, and also in Chapter 12.

Sometimes experiments simply don't work or can't be done for some reason such as equipment failure or bad weather. A water rocket is hard to launch in high winds, and not pleasant to use in heavy rain. Try to think of all the things that could go wrong. If possible, come up with backup plans. Allowing plenty of time helps with this, as does starting your experiments as soon as possible.

Make sure you allow time for analysis. Ideally, do as much analysis as you can while you collect results. If you plot graphs as you take measurements, then you will be able to identify **outliers** early. An outlier is a data point that does not fit the pattern of the rest of the data. If you identify an outlier while you still have access to equipment and space, you can check the measurement and make sure that you didn't make a mistake.

After you have analysed your results, you need to write your report or communicate your findings in some other form. You need to plan ahead how this will be done. If you are working in a group, who will write which part of the report and when? Who will proofread it? Who will be responsible for making sure all the parts fit together?

You may find a timeline useful. A timeline helps keep you on track, and reminds everyone of their responsibilities. If you are working in a group get everyone to agree on it.

You can use the following table as a template.

Date and place	What will be done	By whom	Outcomes

Risk assessment

You may be required to complete a risk assessment before you begin your investigation. Even if this is not a requirement, it is a good idea to think about it. You need to think about three things.

1. *What are the possible risks* to you, to other people, to the environment or property?
2. *How likely is it* that there will be an injury or damage?
3. If there is an injury or damage to property or environment, *how serious are the consequences* likely to be?

A 'risk matrix', such as Table 13.1, can be used to assess the severity of a risk associated with an investigation. The consequences are listed across the top from negligible to catastrophic. Negligible may be getting clothes dirty or a very minor injury such as a scratch. Marginal might be a bruise from falling off a bike, or a broken branch in a tree. Severe could be a more substantial injury or a broken window. Catastrophic would be a death or the release of a toxin into the environment. In general, you need to ensure that your investigation is low risk. You can use a risk matrix either for individual identified risks, or for the investigation overall. If there are multiple experiments, then you would use a risk matrix for each one.

Table 13.1 Risk matrix for assessing for severity of risk

Consequences→ Likelihood↓	Negligible	Marginal	Severe	Catastrophic
Rare	Low risk	Low risk	Moderate risk	High risk
Unlikely	Low risk	Low risk	High risk	Extreme risk
Possible	Low risk	Moderate risk	Extreme risk	Extreme risk
Likely	Moderate risk	High risk	Extreme risk	Extreme risk
Certain	Moderate risk	High risk	Extreme risk	Extreme risk

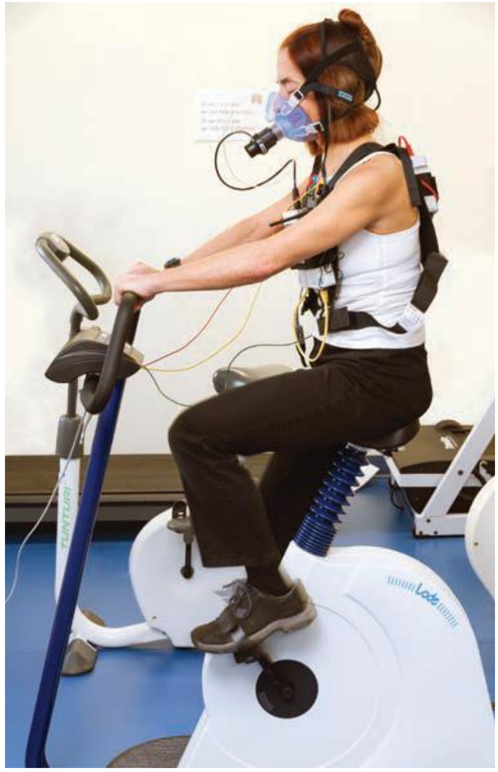
Once you have considered what the possible risks are, you need to think about what you will do about them. What will you do to minimise them, and what will you do to deal with the consequences if something does happen? This may be as simple as, 'Keep everyone at least 10 m away from the water rocket launch site.' You can use a risk assessment table like the one shown.

What are the risks in doing this experiment?	How can you manage these risks to stay safe?
Water from the rocket may be spilled on the ground and someone might slip.	Clean up all spills immediately. Keep the area around the rocket clear.

Ethics

Ethics in research can be controversial. More than one scientist has lost their job for unethical research behaviour. Being ethical in your research has two aspects. The first is about being honest as a scientist. This means recording data accurately, and not ignoring, hiding or changing any data that doesn't support your hypothesis. It means acknowledging and referencing sources of information including books, websites, articles and people who have helped you. It means not using other people's ideas or data without their knowledge and permission. Put simply, it is showing integrity or 'doing the right thing'. A good rule is that if you wouldn't want someone to know what you are doing, you probably shouldn't be doing it. It is no different from behaving ethically in any other area of your life.

The other aspect to ethics is treating animals, other people and the environment with care and respect. If your investigation will be using humans or animals, then you need to make sure you do not harm them, either physically or psychologically. If you are working with animals, then you need to make a strong case for any investigation that harms or could potentially harm them. When scientists want to use humans or animals in their research, they need to be able to show that the benefits to the environment, other animals or humans significantly outweigh the negative effects on the animals or humans used. The National Health and Medical Research Council (NHMRC) has guidelines on the ethical use of humans and animals in experimentation.



Science Photo Library/Antonia Reeve

▲ **Figure 13.7**
When working with humans you need to make sure that they are not harmed in any way.

Collecting your data

Once you have planned what you are going to do, collected your equipment and set it up, it is time to start collecting data. This is usually the fun part of any investigation. Don't forget you have a question to answer or a hypothesis to test! To do this, you need to make sure that you think carefully about what you do and keep good records. Good logbook records are just as important for design-build-test projects. Make sure you carefully record the process, including all the changes to the design that you make when you come to actually build it.

NHMRC

You can find NHMRC human and animal ethics guidelines here.

Record keeping – your logbook

You will need to keep a record of what you do during your investigation. You do this in a logbook. Even if you are using data loggers to collect your results and doing your analysis electronically, you should still keep a hard-copy logbook.

Scientists keep a logbook for each project that they work on. It is a record of what they did, why they did it, and what they found out. A logbook is a legal document for a working scientist. If someone's work is called into question, then the logbook acts as important evidence. Every entry in a scientist's logbook is dated, records are kept in indelible form (pen, *not pencil*), and entries may even be signed. Scientists' logbooks include details of experiments such as methods and results. They include comments and ideas, thoughts about the experiments and

iStockphoto/SteveStone

Figure 13.8 ▲
Make sure you keep an accurate record of what you do *as you do it.*

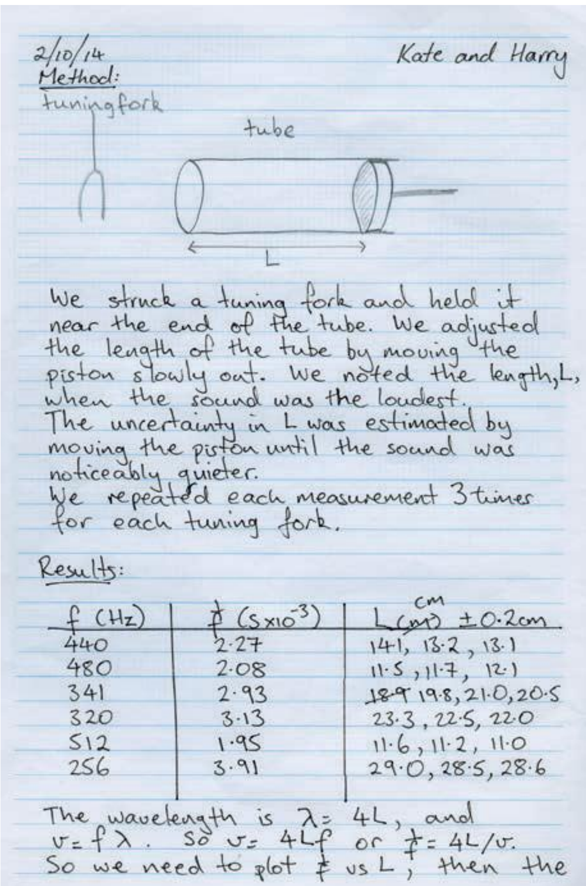
analysis. They frequently include printouts of data, photocopies of relevant information, photos and other items. The logbook is the primary source of information when a scientist writes up their work for publication. Logbooks are even provided as evidence in court cases sometimes; for example, in patent disputes or when a researcher is accused of falsifying data or stealing someone else's results.

Some scientists keep their research records electronically, but most experimental scientists still keep a hardcopy logbook. There are several advantages to a hardcopy logbook over an electronic one. First, electronic records are easy to make changes to, and it is hard to track what was changed, when and by whom. Second, if you are working in a group, it can be hard to keep track of who has the most recent version of the file(s). Third, files can be easily deleted or corrupted. It takes much more care and discipline to maintain a good electronic logbook than a good hardcopy. Remember that the purpose of a logbook is to record and maintain evidence of what you did. Electronic evidence is not as reliable as a signed hardcopy document.

You should talk to your teacher about what form of logbook records they require you to keep.

If you are working in a group then you will need to decide whether to keep one logbook for the entire group, or one each. If you will all be working in the same places at the same times, then one for the whole group is best. If you will be in different places (e.g. doing field observations), then you will need one each. Your teacher may also require each of you to keep your own logbook for assessment, or for authentication purposes.

Your logbook is a detailed record of *what you did* and *what you found out* during your investigation. Make an entry in the logbook *every* time you work on your investigation. At the start of each session you should record the date and the names of all the people with whom you are working at the time.



2/10/14 Kate and Harry

Method:

We struck a tuning fork and held it near the end of the tube. We adjusted the length of the tube by moving the piston slowly out. We noted the length, L, when the sound was the loudest.
The uncertainty in L was estimated by moving the piston until the sound was noticeably quieter.
We repeated each measurement 3 times for each tuning fork.

Results:

f (Hz)	T (s×10^{-3})	L (cm) ±0.2cm
440	2.27	14.1, 13.2, 13.1
480	2.08	11.5, 11.7, 12.1
341	2.93	~~18.9~~ 19.8, 21.0, 20.5
320	3.13	23.3, 22.5, 22.0
512	1.95	11.6, 11.2, 11.0
256	3.91	29.0, 28.5, 28.6

The wavelength is $\lambda = 4L$, and $v = f\lambda$. So $v = 4Lf$ or $T = 4L/v$.
So we need to plot T vs L, then the

◀ Figure 13.9
A page from a student's logbook

Write down what you do as you do it. It is easy to forget what you did if you do not write it down immediately. An accurate record is important if you need to repeat any measurements or if you get unexpected results.

Write down the names, model and serial numbers of any equipment used.

Include large, clear diagrams of any experimental set-up. Label all the parts or pieces of equipment. You can also include photos of experiments. Include large and clearly labelled circuit diagrams of all circuits that you use. Diagrams in your logbook do not need to be neat, but you must be able to understand the diagrams later on.

If you are doing a design-build-test project, then include diagrams of the build process. Flowcharts are very useful for this. Take lots of photos during the building and testing stages. Print these out and attach them to your logbook, and make a note of where the files are kept. These will be useful when you write your report later.

Record the results of *all* measurements *immediately and directly into your logbook, in pen*. *Never* record data onto bits of scrap paper instead of your logbook! Results must be recorded in indelible form. This means using a pen. Never write your results in pencil. Never use white-out or scribble over anything in your logbook. If you want to cross something out, just put a line through it. It is also a good idea to make a note explaining why it was crossed out.

A good logbook contains:
- notes taken during the planning of your investigation
- a record of when, where and how you carried out each experiment
- diagrams showing the experimental set-ups, circuit diagrams, etc.
- all your raw results
- all your derived results, analysis and graphs
- all the ideas you had while planning, carrying out experiments and analysing data
- printouts, file names and locations of any data not written directly into the logbook.

It is not a neat record, but it is a *complete* record.

Performing experiments

If you have planned carefully and learned how to use the equipment, then hopefully your experiments will go smoothly.

In a design-build-test project, this is the testing stage.

As stated above, *always* record results immediately, with the correct units and with their uncertainty. The raw data should *always* be recorded directly into the logbook unless it is recorded using data loggers connected to a computer. In this case a printout of the data should be attached to the logbook, and the file name and location recorded.

Make sure that you measure and record everything you will need for your analysis. For example, if you are investigating water rockets, you could measure the weight of the rocket, the air pressure used and the wind strength. It is much better to measure something and then discover that you didn't need to, than to start your analysis and realise that you didn't measure something that you do need.

Use SI units. This means metres (m) for lengths, seconds (s) for time and kilograms (kg) for mass. If you are measuring current, use amperes (A) and use volts (V) for potential differences. The uncertainties on raw data will be in the same units as the measurements. Always record these along with the measurements as you go. Do not try to add them in later.

See Chapter 12 for a discussion of SI units.

If you are going to be collecting multiple data points, then it is a good idea to draw a table to record them in. Label the columns in the table with the name and units of the variables. Do not put the units in the table cells. If you know that the uncertainty in all your measurements is the same, then you can record this in the heading cell at the top of the column as well. Otherwise, each data entry should have its uncertainty recorded in the cell with it. Remember that the uncertainty in the raw result should have the same unit as the result.

It is a good idea to start your analysis while you are collecting your data. If you spot an outlier and you are still making measurements then you have the opportunity to repeat that measurement. If you made a mistake, then put a line through the mistake, write in the new data, and make a comment in your logbook. Do not scribble out or remove mistakes, they may turn out to be useful.

If you have not made a mistake, then plotting and analysing as you go allows you to spot something interesting early on. You then have a choice between revising your hypothesis or question to follow this new discovery, or continuing with your plan. Many research projects start with one question and end up answering a completely different one. These are often the most fun, because they involve something new and exciting.

Estimating uncertainties

When you perform experiments there are typically several sources of uncertainty in your data.

Sources of uncertainty that you need to consider are the:
- limit of reading of measuring devices
- precision of measuring devices
- variation of the measurand.

Limit of reading

For all devices there is an uncertainty due to the limit of reading of the device. The limit of reading is different for **analogue** and digital devices.

Analogue devices include swinging needle multimeters, liquid in glass thermometers and clocks with hands. Analogue devices have continuous scales. For an analogue device, the **limit of reading**, sometimes called the **resolution**, is half the smallest division on the scale. We take it as half the smallest division because you will generally be able to see which division mark the indicator (needle, fluid level, etc.), is closest to. You may be able to estimate the measurement to one-fifth or even one-tenth of the smallest division if the spacing between divisions is large; however, the limit of reading uncertainty is still half of the smallest division. So, for a liquid in a glass thermometer with a scale marked in degrees Celsius, the limit of reading is 0.5°C.

Digital devices such as digital multimeters, clocks and thermometers have a scale that gives you a number. It is limited to a specific number of figures, typically three or four, so it is a discrete scale. A digital device has a limit of reading uncertainty of a whole division. So a digital thermometer that reads to whole degrees has an uncertainty of 1°C. For a digital device the limit of reading is *always* a whole division, not a half, because you do not know whether it rounds up or down, or at what point it rounds.

The resolution or limit of reading is the *minimum* uncertainty in any measurement. Usually the uncertainty is greater than this minimum.

Figure 13.10 ▶
a) This digital scale has a limit of reading of 0.001 s. b) This analogue scale has a limit of reading of 0.1 s.

a

Shutterstock.com/Jan van der Hoeven

b

Shutterstock.com/Ehrman Photographic

Precision of measuring device

The measuring device used will have a **precision**, usually given in the user manual. For example, a multimeter may have a precision of 0.5% on a voltage scale. This means if you measure a potential difference of 12.55 V on this scale, the uncertainty due to the precision of the meter is $0.005 \times 12.55\text{ V} = 0.06\text{ V}$. This is greater than the limit of reading uncertainty, which is 0.01 V in this case.

Many students think that digital devices are more precise than analogue devices. This is often not the case. A digital device may be easier for you to read, but this does not mean it is more precise. The uncertainty due to the limited precision of the device is generally greater than the limit of reading.

a

Courtesy of FLIR Commercial Systems

b

Function	Range	Resolution	Accuracy	
DC voltage	400 mV	0.1 mV	±(0.3% reading + 2 digits)	
	4 V	0.001 V	±(0.5% reading + 2 digits)	
	40 V	0.01 V		
	400 V	0.1 V		
	1000 V	1 V	±(0.8% reading + 3 digits)	
AC voltage			50 to 400 Hz	400 Hz to 1 kHz
	400 mV	0.1 mV	±(1.5% reading + 15 digits)	±(2.5% reading + 15 digits)
	4 V	0.001 V	±(1.5% reading + 6 digits)	±(2.5% reading + 8 digits)
	40 V	0.01 V		
	400 V	0.1 V		
	750 V	1 V	±(1.8% reading + 6 digits)	±(3% reading + 8 digits)
Frequency	5.000 Hz	0.001 Hz	±(1.5% reading + 5 digits)	
	50.00 Hz	0.01 Hz		
	500.0 Hz	0.1 Hz	±(1.2% reading + 2 digits)	
	5.000 kHz	0.001 kHz		
	50.00 kHz	0.01 kHz		
	500.0 kHz	0.1 kHz		
	5.000 MHz	0.001 MHz	±(1.5% reading + 4 digits)	
	10.00 MHz	0.01 MHz		
	Sensitivity: 0.8 V rms min. @20% to 80% duty cycle and <100 kHz; 5 V rms min. @20% to 80% duty cycle and >100 kHz			
Duty cycle	0.1 to 99.9%	0.1%	±(1.2% reading + 2 digits)	
	Pulse width: 100 μs – 100 ms, Frequency: 5 Hz to 150 kHz			

Note: Accuracy specifications consist of two elements:
- (% reading) – This is the accuracy of the measurement circuit.
- (+ digits) – This is the accuracy of the analog to digital converter.

Courtesy of FLIR Commercial Systems

▲ **Figure 13.11**
a) A typical small digital multimeter; b) A page from the user manual giving the precision on various scales

Variation of the measurand

The measurand itself may vary. For example, the flight of a water rocket is strongly dependent on initial conditions, wind and other factors. Even keeping launch conditions as close to identical as possible, it is unlikely that in repeat experiments you will be able to get a rocket to attain the same height within the limit of reading or equipment precision. Making repeat measurements allows you to estimate the size of the variation.

Sometimes you will be able to see how the measurand varies during a measurement by watching a needle move or the readings change on a digital device. Watch and record the maximum and minimum values. The difference between these is the range:

$$\text{Range} = \text{maximum value} - \text{minimum value}$$

The value of the measurand is the average value, or the centre of the range:

$$\text{Measurand} = \text{minimum value} + \frac{1}{2}(\text{range})$$
$$= \text{minimum value} + \frac{1}{2}(\text{maximum value} - \text{minimum value})$$

The uncertainty in the measurement is half the range:

$$\text{Uncertainty} = \frac{1}{2}(\text{range}) = \frac{1}{2}(\text{maximum value} - \text{minimum value})$$

For example, if you are using an analogue multimeter and you observe that the needle fluctuates between 12.2 V and 12.6 V then your measurement should be recorded as (12.4 ± 0.2) V. Note that the measurement and uncertainty are together in the brackets, indicating that the unit applies to both the measurement and its uncertainty.

For further discussion of uncertainties see Chapter 12.

When you take repeat measurements, the best estimate of the measurand is the average value. If you have taken fewer than 10 measurements then the best estimate of the uncertainty is half the range. If you have more than 10 measurements, the best estimate of the uncertainty is the standard deviation, given by:

GUM

Read more about uncertainties in the GUM (Guide to expressions of Uncertainty in Measurement).

$$\text{Standard deviation} = \left(\frac{\sum(x_i - x)^2}{n - 1}\right)^{\frac{1}{2}}$$

where x_i is an individual value of the measurand, x is the average value of the measurand and n is the total number of measurements. The sum is over all values of x_i. Most calculators have built-in statistical functions such as standard deviation. Spreadsheet software such as Excel also calculates functions such as standard deviation. Remember that repeat measurements means repeating under the same conditions. It is not the same as collecting lots of data points under different conditions.

Random and systematic errors

These sources of uncertainty all give rise to random errors. That means that repeated measurements will be randomly spread about the 'true value', and centred on that value.

You may also have systematic errors in your data. These typically occur when there is a calibration error, such as a zero error, in a measuring device. Always check that your equipment reads zero when you expect it to. For example, the volts and amps scales on a multimeter should read zero when the leads are not connected to anything. A weighing scale should read zero if no force is applied. If it shows some other value, then all your measurements will be out by this amount.

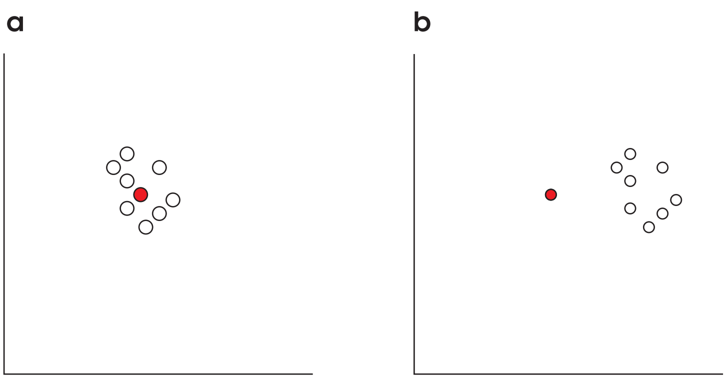


Figure 13.12 a) Results are clustered about the true value when the errors are random. b) Results are clustered about some other value as a result of systematic errors.

Analysing your data

Once you have collected your data you will need to analyse it. Record your analysis in your logbook. If this is done on a computer, then record the file name and location and attach a printout of the analysis into your book. Many scientists have logbooks that are bulging with printouts.

The first step is organising your data. This will usually involve tabulating it. Plotting graphs is a useful way to begin the analysis of your data. Graphs are a very useful way of representing data so that trends and relationships can be identified. There are many different sorts of graphs that can be used to organise and display data. These are described on the following pages.

You will usually need to do some calculations with your data to be able to answer your question or test your hypothesis. Remember to keep units on all quantities, so that any derived values have the correct units. You will also need to calculate uncertainties on any derived quantities.

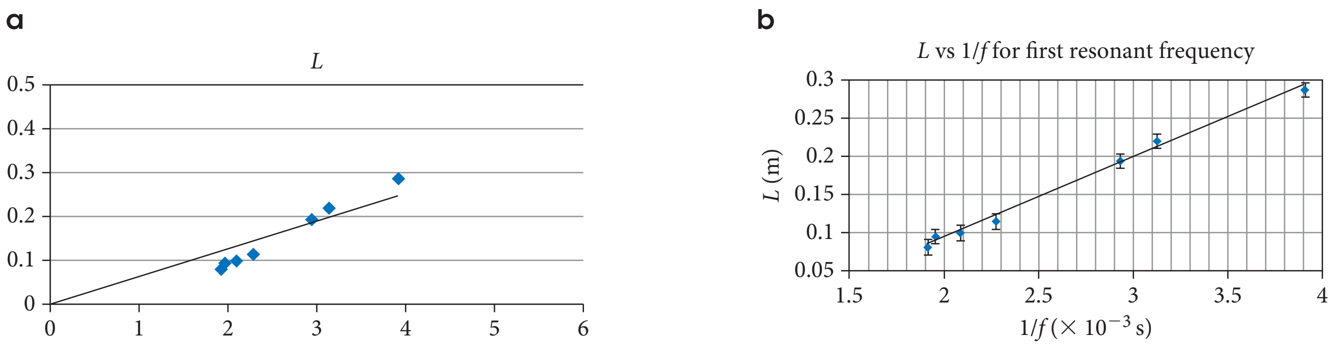
Organising your data

If you have more than a few data points then it is a good idea to display them in a table. You may have several tables, for different experiments. You may also need to do some analysis of the data. For example, imagine you have measured the temperature of different materials as a function of time after heating them. In this case you might have five tables of raw data, each with measurements of temperature as a function of time for a different material. A table summarising the data in some way will be useful. For example, a table showing the time taken for each material to drop from 55°C to 45°C may be useful. Alternatively, a table showing the change in temperature over a given time period may be more useful. You cannot have too many tables or graphs in your logbook. You can decide later, when you write your report, which are the most useful for communicating your results.

Identifying trends, patterns and relationships

You may be able to see a pattern simply by looking at a list of numbers in a table. However, the most reliable way to identify a pattern in data or a relationship between variables is to plot a graph. If you have a hypothesised equation then use it to generate a fit on a graph of your data, as described below. Do not substitute your data into your hypothesised equation and try to show that it fits.

A graph should be large and clear. The axes should be labelled with the names of the variables and their units. Choose a scale so that your data takes up most of the plot area. This will often mean that the origin is not shown in your graph. Usually there is no reason why it should be.

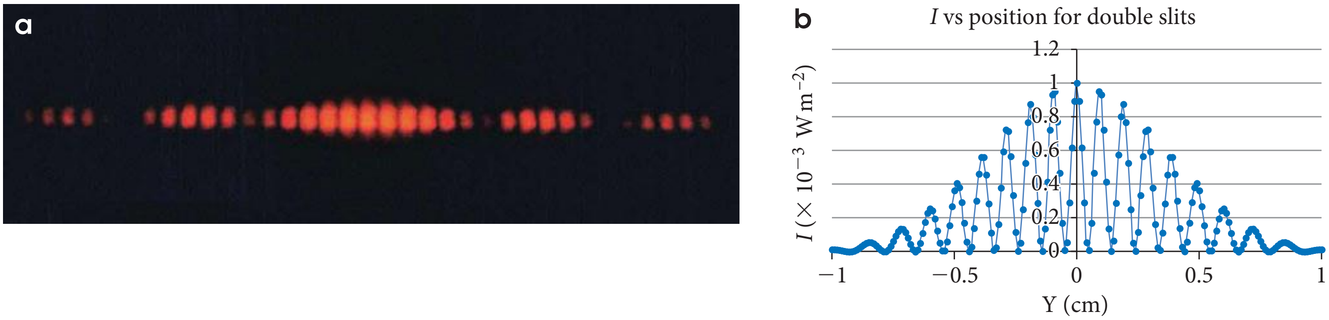


▲ **Figure 13.13**
a) A poor example of a graph; b) A good example of a graph of the same data. How many problems can you identify on the graph in part a?

When you are looking for a relationship between variables, plot a **scatter graph**. This is a graph showing your data as points. *Do not join them up as in a dot-to-dot picture*. Usually the independent variable is plotted on the *x* axis and the dependent variable goes on the *y* axis, unless there is a good reason to do otherwise. In the water rocket example, height would be on the *y* axis plotted against water volume on the *x* axis.

To determine a relationship you need to have enough data points and the range of your data points should be as large as possible. A minimum of six data points is generally considered adequate if the relationship is expected to be linear, but always collect as many as you reasonably can, given the available time.

For non-linear relationships you need more data points than this. Try to collect more data in regions where you expect rapid variation. Imagine you are measuring an interference pattern from two slits, as shown in Figure 13.14. You may need more than a hundred data points to clearly see the sinusoidal pattern due to the two slit interference.



▲ Figure 13.14
a) Interference pattern from two slits; b) Plot of intensity as a function of position for this experiment

DATA POINTS

Some helpful advice on deciding the number of data points

A good graph to start with is simply a graph of the raw data. You will usually be able to tell by looking whether the graph is linear. If it is, then fit a straight line using a graphing package. You can then use a linear regression tool to check how good the straight line fit is. This will give you an R^2 number, which is a measure of 'goodness of fit'. The closer R^2 is to 1 (or –1), the better the fit. If it is not *very* close to 1, then the relationship is not linear. Alternatively, you can calculate the uncertainty in the gradient by using lines of maximum and minimum gradient. If the uncertainty is large, then the relationship may not be linear.

If it is a linear relationship, then finding the equation for the line of best fit may be useful. *Never* force a line of best fit through the origin. Often the intercept gives you useful information. It may even indicate a systematic error, such as a zero error in calibration of your equipment.

When you plot your raw data you may find that one or two points are outliers. These are points that do not fit the pattern of the rest of the data. These points may be mistakes; for example, they may have been incorrectly recorded or a mistake was made during measurement. They may also be telling you something important. For example, if they occur at extreme values of the independent variable then it might be that the behaviour of the system is linear in a certain range only. This is the case for materials under stress. You may choose to ignore outliers when fitting a line to your data, but you should be able to justify why.

When you extend a line of best fit beyond your measured points this is called **extrapolation**. Any data that you read off a graph outside the range of your data points is extrapolated, and should be viewed with caution. You cannot say for sure that the system continues to behave in the same way beyond the bounds of your data. For example, imagine you measure the spring constant of a spring by applying weights and measuring its extension. If you plot extension as a function of weight, you should get a straight line. You could in theory extrapolate your line of best fit to any weight. But you know that in practice if you continued adding weights you would eventually break the spring.

Reading points, other than data points, from a line of best fit within the region in which you have data is called **interpolation**. You cannot be sure that this is exactly what you would find if you measured that point. However, if your line of best fit really represents the behaviour of the system, then you can use interpolated points in your analysis.

Relationships between variables are often not linear. If you plot your raw data, for example height of rocket trajectory as a function of pressure, and it is a curve, then *do not draw a straight line through it*. In this case you need to think a little harder. If your hypothesis predicts the shape of the curve, then try fitting a theoretical curve to your data. If it fits well, then your hypothesis is supported.

If possible, you should **linearise** your data based on your hypothesis. Remember that linear graphs have equations of the form $y = mx + c$. Here y is the variable plotted on the vertical axis, usually the dependent variable. The independent variable x is the variable plotted on the horizontal axis. The gradient is $m = \Delta y/\Delta x$. The constant c is the y intercept.

For example, if your hypothesis is that $h = \frac{1}{2}gt^2$, try plotting your data as a function of t^2. Here h is the initial height of a falling object, g is the acceleration due to gravity and t is the time taken for it to fall:

$$\begin{array}{cccc} h = & \left(\frac{1}{2}g\right) & t^2 & + 0 \\ \uparrow & \uparrow & \uparrow & \uparrow \\ y = & m & x & + c \end{array}$$

Hence a plot of h vs t^2 should be a straight line with gradient $\frac{1}{2}g$ and a y intercept of zero. So if you plot h vs t^2 and get a straight line of gradient $\frac{1}{2}g$ with a y intercept of zero, then your hypothesis is supported.

A second example is radioactive decay. Your hypothesis could be that the activity of your sample at some time t is $A = A_0e^{-kt}$. If you measure A as a function of time, t, then a plot of ln (A) vs t will have a gradient of $-k$ and an intercept of A_0. You can see that this is the case if you take the natural logarithm (ln) of both sides:

$$\ln(A) = \ln(A_0e^{-kt}) = \ln(A_0) + (-kt)$$

which is again of the straight line form:

$$\begin{array}{cccc} \ln(A) = & -k & t & + \ln(A_0) \\ \uparrow & \uparrow & \uparrow & \uparrow \\ y = & m & x & + c \end{array}$$

It is better to linearise your data rather than to try fitting a curve to non-linear data. Often a curve for an exponential relationship can look very much like a curve for a power law. Linearising your data allows you to distinguish between the two.

Log–log graphs are useful for power laws. A log–log graph will give you a straight line if there is a power law relationship between the variables. For example, if the relationship is of the form $y = ax^n$, then if we take logarithms of both sides we get $\log(y) = \log(a) + n\log(x)$. A plot of log ($y$) vs log ($x$) then has gradient n and intercept log (a).

Remember that the argument of a logarithm is dimensionless. Hence we ignore the units of a, x and n during this process. It is important to remember to put them back at the end and make sure your final expression is dimensionally correct.

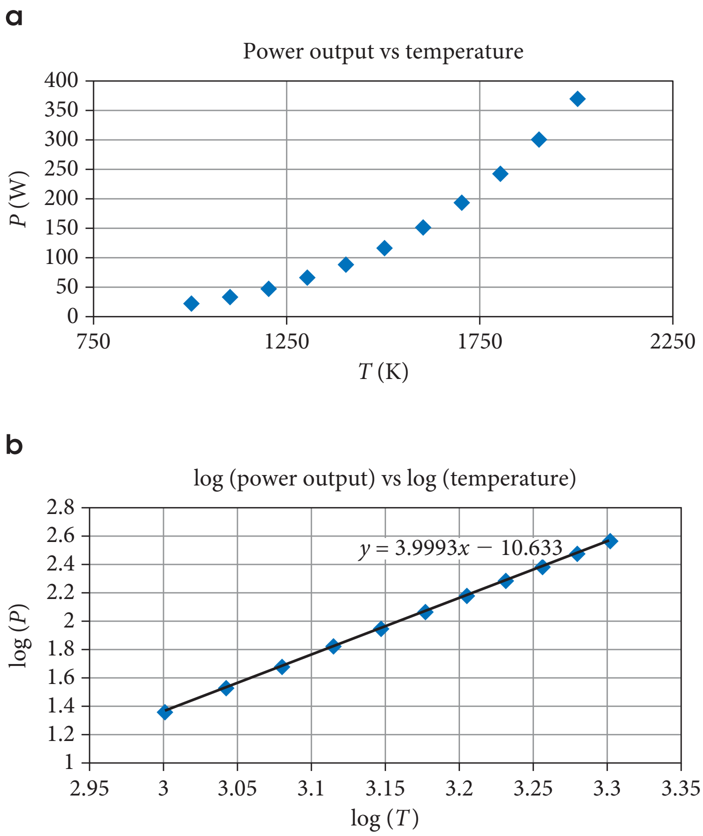
Consider the data shown in Table 13.2.

Table 13.2

Temperature (K)	Power (W)
1000	23
1100	34
1200	48
1300	67
1400	89
1500	117
1600	152
1700	194
1800	243
1900	301
2000	370

Plots of T vs P and log T vs log P are shown in Figure 13.15.

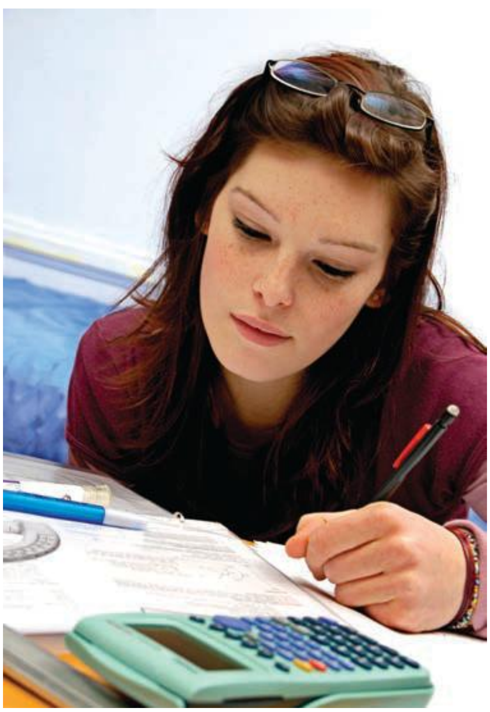
Figure 13.15 ▶
a) Direct plot of the data in Table 13.2; b) Log–log graph of the same data, with trendline displayed



We can see from the equation for the line of best fit that the gradient is 3.9993 and the intercept is −10.633. Hence we deduce that the relationship between P and T is $P = (2.7 \times 10^{-11}\ \text{W K}^{-4})\ T^{4}$. This is a powerful technique, and one well worth practising.

Sometimes you do not know what relationship to expect between variables, or you may have tried to fit your data and it has not worked. With spreadsheet software it is very quick to generate log–linear, log–log graphs and other plots, so it is worth trying a few.

Figure 13.16 ▼
You will usually need to analyse your raw data in some way.



Science Photo Library/AJ Photo

Sometimes the relationship between variables will be more complicated than a linear, exponential or power law. In this case, a graph is still useful but the most you might be able to say is that one variable increases with another, or that there is a peak at a particular position. This is likely to be the case with the water rocket example we have been using. A graph is still a useful way of identifying trends and patterns, even if you are not able to extract a mathematical relationship from the graph.

Performing calculations with your data

You will usually have to do some calculations with your data as part of your analysis. When you recorded your data you wrote down the units for all your measurements. You may need to convert these to SI units (e.g. cm to m). Include the units with all numbers as you do your calculations. In this way you will make sure you have the correct units on all derived data. It also allows you to check that any equations you are using are dimensionally correct. For example, if you are using $h = \frac{1}{2}gt^2$ to estimate a maximum height obtained, then you should be using t in s and g in m s^{-2}. If you measured time in ms, then you need to convert to s before calculating h. If you make a mistake and try to use $h = \frac{1}{2}gt$ instead, then if you have included units in your calculation you should notice that your answer for height is in m s^{-1}. For example, if you measure $t = 3.0$ s and use $g = 9.8$ m s^{-2}, then using the correct equation:

$$h = \frac{1}{2}gt^2 = \frac{1}{2}(9.8 \text{ m s}^{-2})(3.0 \text{ s})^2 = \frac{1}{2}(9.8 \text{ m s}^{-2})(9.0 \text{ s}^2) = 44 \text{ m}$$

If you use the incorrect equation you get:

$$h = \frac{1}{2}gt = \frac{1}{2}(9.8 \text{ m s}^{-2})(3.0 \text{ s}) = \frac{1}{2}(9.8 \text{ m s}^{-2})(3.0 \text{ s}) = 15 \text{ m s}^{-1}$$

The incorrect unit for h should alert you to an error, even if the difference in numerical value does not.

It is good practice *in general*, not just in investigations, to include units at each step in all your calculations.

Your raw data should be recorded with uncertainties. All your derived results should also have uncertainties. Chapter 12 shows how to calculate uncertainties on derived values. You can also refer to books on experimental methods and uncertainties in measurement for a more detailed description of how to treat uncertainties.

Interpreting your results

Once you have analysed your results you need to interpret them. This means being able to either answer your research question or state whether your results support your hypothesis. In a design-build-test project this is where you determine whether, or to what extent, your final product meets the requirements of the design brief.

You need to take into account the uncertainties in your results when you decide whether they support your hypothesis. For example, suppose you have hypothesised that the maximum range of a water rocket occurs at a launch angle of 45°. Your results show that the maximum range occurs at an angle of 47°. You may think that this result does not support your hypothesis. To say whether the result agrees with the prediction, you need to consider the uncertainty. If the uncertainty is 1°, then the results disagree with the hypothesis. If the uncertainty is 2° or more, then the results do agree and the hypothesis is supported.

You need to know the uncertainty to be able to interpret the results.

If your hypothesis is not supported

It is not enough to simply say 'our hypothesis is wrong'. If the hypothesis is wrong, *what* is wrong with it?

It may be that you have used a model that is too simple. For example, if you have based your hypothesis on the kinematics model and ignored the effect of air resistance, the range is likely to be shorter and the maximum height lower than you predicted. For many projectiles, including water rockets, air resistance is not negligible. If you find that this is the case, then you may conclude that your situation is better described by a model that includes air resistance.

Before you decide that the model is at fault, however, it is a good idea to check carefully that you have not made any mistakes or ignored any variables.

Think carefully about any factors that you did not take into account but which might have affected your experiment.

Go through your method, results and analysis. Check that your equipment was correctly calibrated, and that you were using it correctly. Check that data is recorded in the correct units, and that units are correctly carried through all calculations during analysis. Check your analysis carefully. If you are working in a group, get another person to repeat the calculations.

It is never good enough to conclude that 'the experiment didn't work'. Either a mistake was made or the model used was not appropriate for the situation. It is your job to work out which.

If you have done a design-build-test project and your final product does not meet the design brief then you should explain why. Try to determine whether the basic design was flawed, or if there was a problem with materials or something else. If you can, offer suggestions for a better design or process.

Communicating your results

If research is not reported on, then no-one else can learn from it. An investigation is not complete until the results have been communicated. Most commonly a report is written.

Writing reports

A report is a formal and carefully structured account of your research. It is based on the data and analysis in your logbook. However the report is a *summary*. It contains only a small fraction of what appears in the logbook. Your logbook contains all your ideas, rough working and raw data. The report typically contains none of this.

A report consists of several distinct sections, each with a particular purpose.

Abstract
- Introduction
- Method
- Results and analysis
- Discussion
- Conclusion
- Acknowledgements
- References
- Appendices

Reports are always written in the past tense, because they describe what you have done.

The abstract

The abstract is a very short summary of the entire report. It is the most important part, because often it is the only part that people read. Typically an abstract is between 50 and 200 words long. It appears at the start of the report, but is always the last thing that you write. Try writing just one sentence to summarises each part of your report.

Introduction

The introduction tells the reader why you did the investigation and what your research question or hypothesis is. This is the place to explain why this research is interesting.

The introduction also provides any background information needed to be able to understand the rest of the report. This is the place to summarise any existing theories and models. You need to do this to justify your hypothesis. You should also summarise any similar investigations. All of this should be correctly referenced, as described in the section on referencing on page 441.

Method

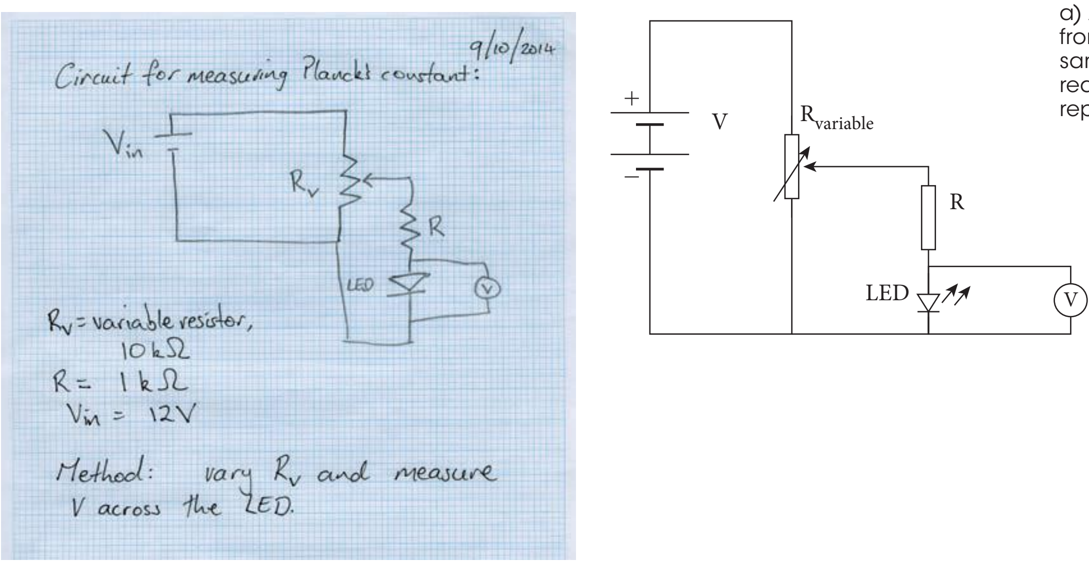
The method describes what you did. It is not a recipe for someone else to follow.

The method summarises what you measured and how you measured it. It also explains, briefly, why you chose a particular method or technique.

Write your method using sentences, not dot points. Remember that these need to be written in the past tense – *it is not a recipe*. You are not commanding anyone to do anything. You are telling people what you did. For example, you would write 'we measured the height' not 'measure the height'.

Include any diagrams, such as circuit diagrams, needed to make your method clear. The diagrams in your logbook will usually be rough sketches (see Figure 13.17). The diagrams in your report should be very neat and carefully labelled. Flowcharts can be useful to describe any procedures in which a series of steps was followed. Each diagram should have a figure number and you should refer to it in the text of your report. Position the diagram close to where it is referred to in the text. You should take the time to learn how to position figures neatly using your word processor software.

a b



◀ **Figure 13.17**
a) A circuit diagram from a logbook; b) The same circuit diagram redrawn in a formal report

Results and analysis

The results section is a *summary* of your results. It is usually combined with the analysis section, although they may be kept separate.

Avoid including tables of raw data in your report unless they compare the results of a few different experiments. For example, you would include a table showing the maximum height attained for a few different designs of water rocket. You would not include a table of raw data showing height attained for a water rocket for many different volumes of water. Wherever possible use a graph instead of a table.

If a table has more than a few rows of data, it is better to represent that data in some other way. Usually this will be a graph.

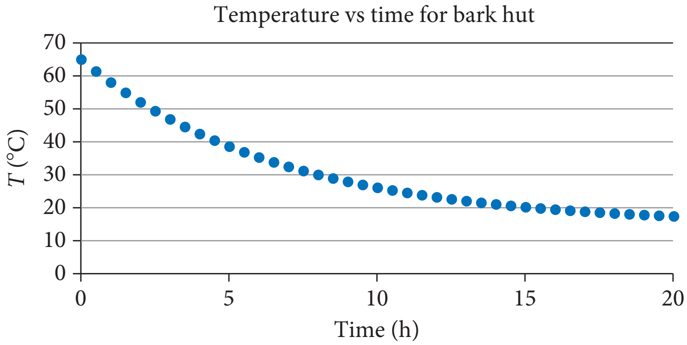
Think about what sort of graph is appropriate. If you want to show a relationship between two variables then use a scatter plot. Display your data as points with uncertainty bars and clearly label any lines you have fitted to the data. Always make sure you label your axes, including units. Choose an appropriate scale so that the data takes up most of the plot area.

Column and bar charts are useful for comparing two data sets, such as average height attained for different types of water rocket. *Do not* use a column or bar chart to try to show a mathematical relationship between variables.

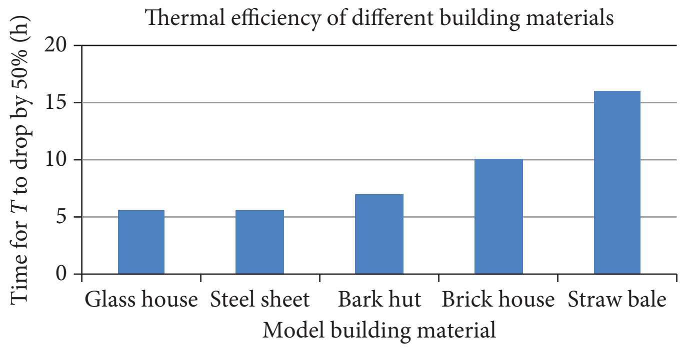
Figure 13.18 gives examples of the two types of graphs.

Figure 13.18 ▶ a) A scatter plot demonstrating a mathematical relationship; b) A column graph comparing results from different experiments

a



b



Any data and derived results should be given in correct SI units with their uncertainties. If you performed calculations then show the equations you used. You might want to show one example calculation, but do not show more than one if the procedure used is repeated.

Discussion

The discussion should summarise *what your results mean*. If you began with a research question, give the answer to the question here. If you began with a hypothesis, state whether your results support your hypothesis or not. If not, explain why. You may not be able to say why, other than that the model was not suitable for the situation being investigated.

If there are any implications of your work, such as how to build or do something in a better way, put them here.

The discussion is also the place to briefly describe any difficulties that you had and make suggestions for improving the process. Remember that you should never say 'the experiment didn't work' if you didn't get the results you expected. You might choose to make some comments on possible further work that could be done.

Conclusion

The conclusion is a *very* brief summary of the results and their implications. Say what you found out and what it means. A conclusion should only be a few sentences long.

Acknowledgements

You should thank anyone who helped you in your investigation. This includes people who supplied equipment or funding, as well as people who gave you good ideas or helped with the analysis. In science, as in other aspects of your life, it is polite to say thank you; however, this is not a necessary section of a report.

References

A reference list details the source for each piece of information used, and is linked to that information in the report.

A reference list details the sources of all information that were actually used to write the report. Wherever a piece of information or quotation is used in your report it must be referenced *at that point*. This is typically done either by placing a number in brackets at the point [2], or the author and year of publication (Smith, 2014). The reference list is then either provided in a footnote at the end of the page, or a single complete list at the end of the report. Referencing must be done in a consistent style. Check with your teacher what style they prefer. There are several good online guides to referencing.

A reference list is *not* the same as a bibliography. A bibliography is a list of sources that are useful to understanding the research. They may or may not have actually been used by the report authors. You should have a bibliography in your logbook from the planning stage of your investigation. The references will be a subset of these sources.

REFERENCING GUIDE

This guide is designed to help you with referencing your sources for assignments.

REFERENCING I-TUTORIAL

This tutorial will help you understand referencing and show you how to avoid plagiarism.

Appendices

Appendices may be used to provide additional information such as raw data that is not necessary to understanding the report but which might be of interest to some readers. Your teacher might require you to provide raw data in an appendix. Reports do not always have appendices.

Design reports

If you have done a design-build-test project, then the report will be of a slightly different form. It will follow the form of an engineering report rather than a scientific report.

Design reports usually contain the following sections.
- Design brief
- Introduction
- Design
- Testing
- Analysis
- Recommendations
- Conclusions
- Acknowledgements
- References
- Appendices

Many of the sections have the same function as in a scientific report. However, the analysis will deal mainly with the performance of the product. Design reports sometimes also include a recommendations section. The conclusion of the report states whether the final product meets the requirements of the design brief. There are some guidelines for and examples of these reports in the weblinks.

REPORT WRITING EXAMPLE 1

The brochure outlines the key features of a design report.

REPORT WRITING EXAMPLE 2

This online resource guides you through the sections of a typical report.

Other ways of communicating your results

You may want to present the results of your investigation in some other way. Scientists communicate their work in many ways. Sometimes a poster is presented or a seminar is given. An article may be written or a website produced. Scientists usually use more than one means, and sometimes several of them, to communicate about a really interesting investigation.

REPORT WRITING EXAMPLE 3

This online resource will help you write a case study.

▲ **Figure 13.19** A post-doctorate student presenting her poster at a poster session at a conference

Look at examples of articles in the scientific and the popular media, on websites, posters and so on. This will give you an idea of the different styles used in the different modes. Think about the purpose. Is it to inform, to persuade or both? What sort of language is used?

Think about your audience and use appropriate language and style. A poster is not usually as formal as a report. A website may be more or less formal, depending on your audience.

Posters and websites use a lot of images. Images are usually more appealing than words and numbers, but they need to be relevant. Make sure they communicate the information you want them to.

WEBSITE ACCESSIBILITY

The Royal Society for the Blind has information on making websites accessible.

Make sure you keep readability and accessibility in mind if you are creating a poster or website. Posters should use large clear fonts and not have too much text. They should be readable from a few metres away. Fonts also need to be large enough and clear on websites and digital images should have tags. You can follow the weblink for more information on accessibility and web-page design.

However you communicate your work, make sure you know what the message is and who the audience is. Once you have established that, you will be able to let other people know about the interesting things you have discovered in your investigation.

CHAPTER GLOSSARY

analogue a device or scale that gives a continuous measurement; the scale is continuous and may show any value in a range

continuous variable a variable that is able to take any value, sometimes within a fixed range; for example, a rainbow is a continuous spectrum

controlled variable the variable that is controlled by the experimenter, so that its values are chosen; also called the independent variable

design brief the document that specifies the requirements for a design, including performance of the final product

digital able to measure only a limited number of possible values, usually within a fixed range

discrete variable a variable that is able to take only specific values, not continuous; for example, a line spectrum is a discrete spectrum

extrapolation extension beyond the measured range of data to read or construct new data that has not been measured

falsifiable able to be disproved

hypothesis a tentative prediction, usually based on an existing model or theory; also a tentative explanation of an observation based on an existing model or theory

interpolation to read or construct a new data point that has not been measured but is within the range of measured data

limit of reading the minimum uncertainty in a measurement due to the precision with which the scale can be read

line of best fit the line that most accurately fits the data, usually calculated using linear regression

linearise to make linear; to convert into a form that can be described by a straight line

logbook the record of an experiment or investigation kept by the scientist performing the experiments; it is a legal record of the experiments and their results

outlier a data point that does not fit the pattern shown by other measured data points

plagiarism presenting someone else's work, including their words or ideas, as your own

precision the variation in repeated measurements, or the uncertainty of a measuring device

primary data data that you have measured or collected yourself

qualitative measurements measurements with descriptive or non-numerical results

quantitative measurements measurements with numerical values

reference the source of a specific piece of information or quotation; to state the source of information

reliable highly likely to be true; a trustworthy source of information or reproducible data

reproducible giving the same result, within uncertainty, when repeated measurements are made

research question the specific question that a particular experiment or investigation is designed to answer.

resolution the limit of reading of a measuring device

scatter graph a graph or plot showing data points, without a line joining the points, and used to demonstrate or determine a mathematical relationship between variables. The axes are defined by the variables

secondary data data or information that has been collected by someone else

theory a collection of models and concepts that explain specific systems or phenomena. Scientific theories allow predictions to be made and hence are falsifiable

uncertainty bars bars drawn above and below and/or to left and right of a data point on a graph to indicate the size of the uncertainty in that point

valid results that are affected only by a single independent variable and hence are reproducible

variable something that can change or be changed, as distinct from a constant, which does not

APPENDICES

Appendix 1: Advice for studying physics and reading the textbook

Overview

Each chapter has a logical structure. This can be seen in the way the headings are written. The largest headings are used to start a complete section. For a logical segment of the text, such as a complete chapter, do the following, in order, before you do a close reading of the text:

- Read all headings and sub-headings.
- Look at all diagrams and read their captions.
- Read the chapter and section reviews.
- Ask: What is this segment or chapter about? What am I going to be learning about?

Close reading

There is no substitute for careful and specific close reading of all relevant parts of the textbook. This is a student's direct responsibility.

A close reading of the text, plus the notes taken in class, will form the basis for the notes that students generate in order to study, understand and remember.

[High-performing students frequently also read around the subject, for example, by using science-related feeds from the Internet and/or subscribing to a science-type magazine.]

Note taking

All notes should be:

- written in your own words.
- in point form and each point should contain one idea only. Points should not be longer than one line.

(Some people like to use a highlighter on their textbook to emphasise what is important – this makes them too dependent on the author; but it can be a worthwhile, intermediate step towards full understanding.)

Questions

- Keep a list of questions about things you need to check up on.
- BRING THE LIST TO CLASS and use it.
- Seek assistance by using the questions as a basis for your learning.

Seeking help

- Always try to specify what you need help about.
- Ask a precise question.

Reading ahead – always!

- The best learning takes place when you are prepared in advance and have some idea about what is coming up.
- Seek advice from your teacher about the next lesson or series of lessons.

Appendix 2: SI and non-SI units

International System of Units (SI)

The international body that decides the appropriate units to be used for the various physical quantities is the *Conférence Générale des Poids et Mesures* (CGPM). The system of units approved by the CGPM and now widely used by the scientific community throughout the world is known as Système International d'Unites (abbreviated SI).

In your experimental work you should use SI units (or their multiples or submultiples).

The SI consists of seven base units and two supplementary units. All other derived units are based on these nine fundamental units.

The base and supplementary units, together with the derived units with special names that might be relevant to your experimental work are listed in Tables A2.1 and A2.2.

Table A2.1 SI base units

Physical quantity	Name of unit	Abbreviation
Length	metre	m
Mass	kilogram	kg
Time	second	s
Electric current	ampere	A
Thermodynamic temperature	kelvin	K
Luminous intensity	candela	cd
Amount of substance	mole	mol

As you become familiar with each new unit you should make a practice of correctly using its abbreviated form.

The internationally recognised prefixes for the SI units together with their abbreviations are given in Table A2.3.

Table A2.2 SI supplementary units derived from SI base units

Name	Symbol	Quantity	Equivalents	SI base unit equivalents
hertz	Hz	frequency	1/s	s^{-1}
radian	rad	angle	m/m	dimensionless
steradian	sr	solid angle	m^2/m^2	dimensionless
newton	N	force, weight	$kg\ m/s^2$	$kg\ m\ s^{-2}$
pascal	Pa	pressure, stress	N/m^2	$kg\ m^{-1}\ s^{-2}$
joule	J	energy, work, heat	N m	$kg\ m^2\ s^{-2}$
watt	W	power, radiant flux	J/s	$kg\ m^2\ s^{-3}$
coulomb	C	quantity of electric charge	A s	A s
volt	V	electromotive force, electrical potential difference, electric potential voltage	J/C	$kg\ m^2\ s^{-3}\ A^{-1}$
farad	F	electrical capacitance	C/V s/Ω	$kg^{-1}\ m^{-2}\ s^4\ A^2$
ohm	Ω	electrical resistance, impedance, reactance	V/A	$kg\ m^2\ s^{-3}\ A^{-2}$
siemens	S	electrical conductance	1/Ω A/V	$kg^{-1}\ m^{-2}\ s^3\ A^2$
weber	Wb	magnetic flux	J/A	$kg\ m^2\ s^{-2}\ A^{-1}$
tesla	T	magnetic field strength, magnetic flux density	$V\ s/m^2$ Wb/m^2 N/(A m)	$kg\ s^{-2}\ A^{-1}$
degree Celsius	°C	temperature relative to 273.15 K	K – 273.15	K – 273.15
lumen	lm	luminous flux	cd sr	cd
lux	lx	illuminance	lm/m^2	m^{-2} cd
becquerel	Bq	radioactivity (decays per unit time)	1/s	s^{-1}
gray	Gy	absorbed dose (of ionising radiation)	J/kg	$m^2\ s^{-2}$
sievert	Sv	equivalent dose (of ionising radiation)	J/kg	$m^2\ s^{-2}$

Table A2.3 Prefixes for SI units

Prefix	Abbreviation	Value
exa	E	10^{18}
peta	P	10^{15}
tera	T	10^{12}
giga	G	10^9
mega	M	10^6
kilo	k	10^3
hecto	h	10^2
deka	da	10

Prefix	Abbreviation	Value
deci	d	10^{-1}
centi	c	10^{-2}
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}
atto	a	10^{-18}

Non-SI units

A number of non-SI units are still in use in scientific literature for a variety of reasons. Some of these are being phased out, but others are likely to remain in use. The more common non-SI units that you might come across are listed in Table A2.4.

Table A2.4 Non-SI units

Physical quantity	Unit	Abbreviation	Conversion to SI units
time	minute hour day year	min h d y	60 s 3.6×10^3 s 8.64×10^4 s 3.156×10^7 s
mass	unified mass unit tonne	u t	1.661×10^{-27} kg 1000 kg
angle	degree	°	dimensionless
energy	electron-volt kilowatt hour	eV kW h	1.602×10^{-19} J 3.60×10^3 J
pressure	millimetre of mercury	mmHg	133.3 Pa
charge	elementary or electronic charge	e	1.602×10^{-19} C
source activity	curie	Ci	3.7×10^{10} Bq
radiation absorbed dose equivalent dose	 rad rem	 rad rem	 0.01 Gy 0.01 Sv

Using SI units

There are certain conventions now adopted widely in scientific literature when SI units are being used. Some of the more important ones are given below.

1 When recording a measurement, write the unit in full or use the recommended abbreviation (e.g. 25 metre or 25 m). Using abbreviations save space and time. Notice the space between the numeral and the unit.

2 SI units named after scientists:

   a If the full word is used, it starts with a lower case letter (e.g. 10 newton, 7 joule, 10^5 pascal, 50 hertz).

   b If the abbreviation is used, it is (or at least commences with) a capital letter (e.g. 10 N, 7 J, 10^5 Pa, 50 Hz).

   c Measurements are written as products. '3 kg' means 'the product of 3 and the mass known as a kilogram', just as '3*x*' in maths means the product of 3 and *x*. Therefore 's' is not added to units (e.g. 5 kg or 5 kilogram, not 5 kgs or 5 kilograms).

   d A full-stop is not placed after the abbreviation of a unit, unless it is at the end of a sentence.

   e When units are combined as a quotient (e.g. metre per second), a solidus (/) or negative index may be used. So m/s or m s^{-1} are both acceptable, though the latter is used more widely. Never use more than one solidus in a unit as in m/s/s for acceleration, which should be m/s^2 or m s^{-2}. It is ambiguous, just as writing 36/6/3 in maths is ambiguous. (This could mean 2 or 18.)

Converting between units

Treat the unit as a multiplier. Use the prefixes in Table A2.3 also as multipliers. For example, 4 kg is the same as:

$$4 \times k \times g = 4 \times 1000 \times g = 4000\,g$$

There are 4000 gram in 4 kg.

1500 cm is the same as $1500 \times (10^{-2}\text{ m}) = 15$ m.

CONVERTING BETWEEN UNITS

Learn more about converting between units at this weblink.

Appendix 3: Some important physical quantities

From time to time you will need to find a value of a physical property from a reputable source. These might include finding:

- the value of a physical constant, such as Newton's universal gravitational constant or the electric constant.
- a physical property, such as boiling point or refractive index, which is characteristic of a particular material.
- a conversion factor such as micrometres to metres, electron-volt to joule, unified mass unit to kilogram.

NIST PHYSICAL REFERENCE DATA

The National Institute of Standards and Technology (NIST) provides a wide range of data, including Standard reference data (SRF). For example, click on 'Other Data' to enter the NIST Gateway.

All physical quantities, including physical constants, are measured to very precise levels of accuracy.

NIST PHYSICAL ELEMENT LABORATORY

This website provides atomic and nuclear data for every element.

Some important physical quantities, including some physical constants, are listed alphabetically in Table A3.1. They are given to four significant figures. The uncertainty in most of these figures is better than six-figure accuracy. They are taken from sources such as the National Institute of Science and Technology (NIST). NIST is a specialist organisation dedicated to metrology (study of measurement).

Table A3.1 Physical constants, physical measures and conversion factors

Physical constants	
Avogadro constant, N_A	6.022×10^{23} mol^{-1}
Coulomb law constant, $\frac{1}{4\pi\varepsilon_0}$	8.988×10^{9} N m^{2} C^{-2}
Universal gravitation constant, G	6.674×10^{-11} m^{3} kg^{-1} s^{-2}
Permittivity of free space electric constant, ε_0	8.854×10^{-12} F m^{-1}
Permeability of free space magnetic constant, μ_0	$4\pi \times 10^{-7}$ H m^{-1} = 12.57×10^{-7} H m^{-1}
Planck constant, h	6.626×10^{-34} J s = 4.136×10^{-15} eV s
Speed of electromagnetic radiation in free space, c	2.998×10^{8} m s^{-1}
Physical measures	
Mass of electron	9.109×10^{-31} kg = 5.486×10^{4} u
Mass of proton	1.6726×10^{-27} kg
Mass of neutron	1.6749×10^{-27} kg
Rydberg constant (for hydrogen), R_H	1.097×10^{7} m^{-1}
Gravitational field strength at Earth's surface), g	(9.80 ± 0.3) N kg^{-1}
Acceleration due to gravity at Earth's surface, g	(9.80 ± 0.3) m s^{-2}
Mass of Earth	5.976×10^{24} kg
Mass of Moon	7.348×10^{22} kg
Mass of Sun	1.989×10^{30} kg
Period of rotation of Earth	8.616×10^{4} s
Radius of Earth (equatorial)	6.378×10^{6} m
Radius of Earth (mean)	6.371×10^{6} m
Radius of Earth's orbit about Sun (mean)	1.496×10^{11} m
Radius of Moon's orbit around Earth (mean)	3.844×10^{8} m

Radius of Sun	6.960×10^{8} m
Solar constant (mean)	1.370×10^{3} W m^{-2}
Density of water (pressure and temperature dependent)	9.982×10^{3} kg m^{-3}
Air density (pressure and temperature dependent)	1.292×10^{3} kg m^{-3}
Air pressure (temperature dependent)	1.013×10^{5} Pa
Speed of sound in air at 0°C	331.4 m s^{-1}
Conversion factors	
Absolute zero, 0 K	−273.15°C
Unified mass unit, u	1.661×10^{-27} kg = 931.5 MeV
Electron-volt, eV	1.602×10^{-19} J
Elementary electron charge, *e*	1.602×10^{-19} C
Coulomb	6.242×10^{18} elementary charges

Appendix 4: Analysis of data

Graphical analysis

The purpose of experiments in physics is to find regularities in the physical world. If one thing varies regularly when another is changed, then it may be possible to find the reason behind the relationship. Therefore, physicists try to exclude all variables that might interfere with their investigations. They look for one variable that causes another one variable to change. Sufficient data is collected to draw graphs in order to visualise the relationship. They compare the shape of the graph against known graph shapes in order to determine the form of the relationship.

Look at Figure A4.1. What information can you obtain from this graph of daily solar intensity in Alice Springs? What was the annual average? When could you expect the greatest benefit from solar panels? Could the annual variation be turned into an algebraic equation?

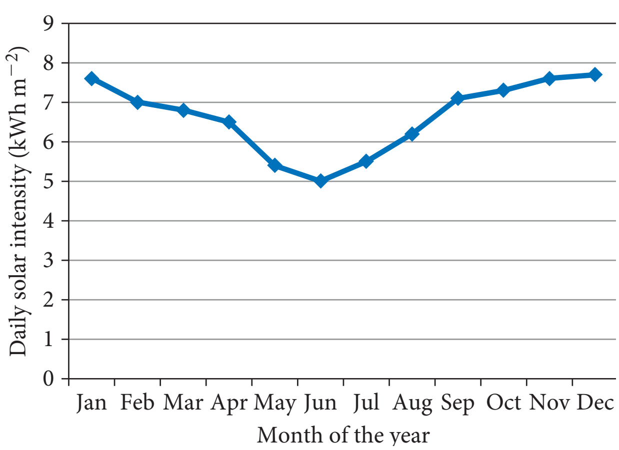


Figure A4.1 ▲
Daily solar radiation per area in Alice Springs for one year

Variables

A variable is some quantity that changes or varies. Some changes cause other changes to occur, For example a force acting on an object causes the object to accelerate. The acceleration of the object does not cause the force to be applied to it. The variable that causes some other to change is called the independent variable. The variable that changes is the dependent variable.

Linear graphs

The most important graph in physics is the straight line or linear graph. In its general form the equation is:

$$y = mx + c$$

The shape of this graph is shown in Figure A4.2.

y is the dependent variable.

x is the independent variable.

m is the gradient of the line.

c is the intercept on the y axis.

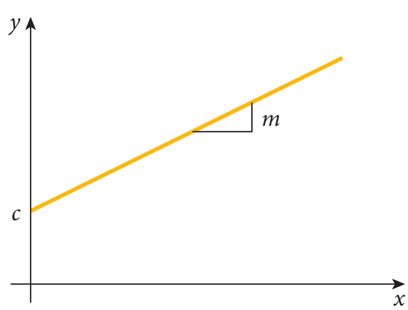
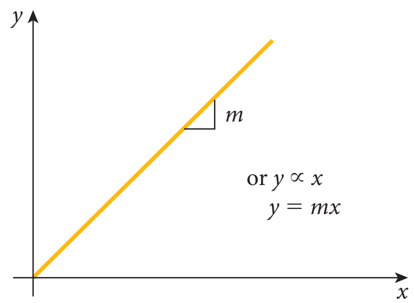


Figure A4.2 ▲
Graph of $y = mx + c$

It is usual to plot the independent variable (the one that is controlled) on the horizontal axis (x axis) and the dependent variable (the one that is not controlled) on the vertical axis (y axis). This will become clearer with practice. It is also possible, and indeed common, to plot the dependent variable against the square or inverse values of the independent variable, for example, y vs x^2, y vs $\frac{1}{x}$, etc.

A special case is the straight-line graph that passes through the origin. The variables are then described as being directly proportional to each other. You need to be careful *not* to assume that (0, 0) at the origin is necessarily a data point.

For direct proportionality or direct variation, $y = mx$ or $y = kx$, the value of the independent variable, y, doubles, triples etc. whenever the value of the independent variable, x, doubles, triples etc.



▲ **Figure A4.3**
Graph of $y = mx$

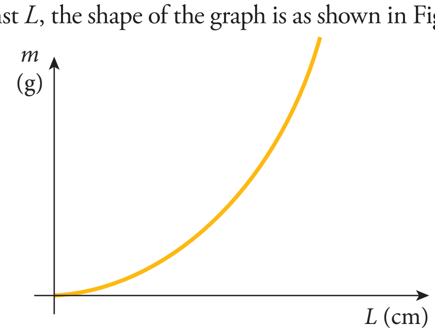
For example, say that you had three cubes of aluminium. Cube A has sides 1.0 cm in length. Cube B has sides that are twice as long (2.0 cm) and cube C has sides three times as long (3.0 cm). You find the mass of each cube and put the data in a data table. What is the relationship between mass and volume?

Measurements made on the cubes gave the results given in Table A4.1.

Table A4.1

Cube	Length of side (L) (cm)	Mass (m) (g)	L^3 (calculated) (cm^3)
A	1.0	2.7	1.0
B	2.0	21.6	8.0
C	3.0	72.9	27.0

Changes in the length cause the mass to change. If we plot m against L, the shape of the graph is as shown in Figure A4.4.



▲ **Figure A4.4**
Graph of m against L

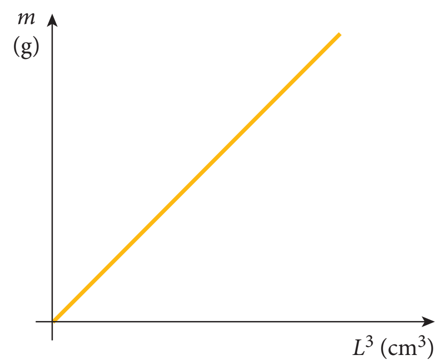
A little experience with graph shapes and variations in the numbers in data tables suggests that the variation is a cubic. We might expect, therefore, that $m = kL^3$. If we plot m against L^3, we see that the shape of the graph is a straight line. We find the gradient, which enables the equation to be written:

$$\text{gradient, } k = \frac{(72.9 - 2.7)\,\text{g}}{(27.0 - 1.0)\,\text{cm}^3}$$

$$k = 2.7\,\frac{\text{g}}{\text{cm}^3}$$

$$k = 2.7\ \text{g cm}^{-3}$$

Thus, $m = 2.7\,L^3$ (see Figure A4.5).



▲ **Figure A4.5**
Graph of m against L^3

In most cases, the gradient has a physical meaning. In this case, the gradient is the density, in g cm^{-3}, of aluminium.

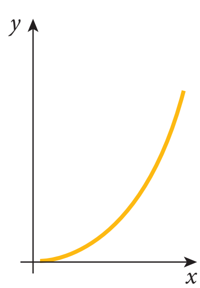
A similar relationship holds for other graphs. That is, if a straight line is obtained, then the quantity plotted on the vertical axis (y axis) is proportional to the quantity plotted on the horizontal axis (x axis).

If we plot the graph of y against x^2, and find the graph is a straight line, the equation becomes:

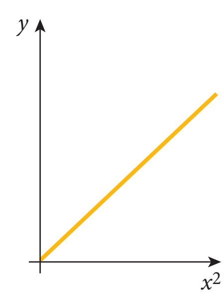
$$y = kx^2$$

where k is the gradient of the graph.

The graphs are shown in Figures A4.6 and A4.7.



▲ **Figure A4.6**



▲ **Figure A4.7**

If we plot the graph of y against the inverse of x, that is $\frac{1}{x}$, and the graph is a straight line, the equation becomes:

$$y = k\left(\frac{1}{x}\right) \text{ or } y = \frac{k}{x}.$$

The graphs are as shown in Figures A4.8 and A4.9.

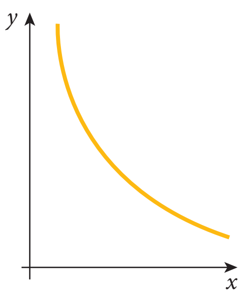


Figure A4.8 ▲

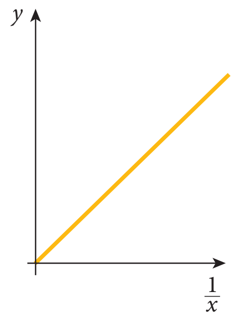


Figure A4.9 ▲

In this case, y and x are said to be inversely proportional.

This graphical method is used to analyse experimental data. Measured values are plotted on a graph. The shape of the graph suggests a possible relationship between the quantities. A further graph may need to be plotted to confirm the relationship and obtain the equation.

Table A4.2 shows experimental results, from the measurement of pressure and volume of a fixed mass of gas, and it can be analysed to find a relationship between P and V.

Table A4.2 Experimental results from the measurement of pressure and volume of a fixed mass of gas

V (m^3)	P (Pa)	$1/V$ (m^{-3})
10.00	81.0	0.100
5.00	159.0	0.200
2.00	397	0.500
1.00	792	1.00

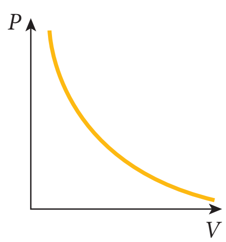


Figure A4.10 ▲
Graph of P against V

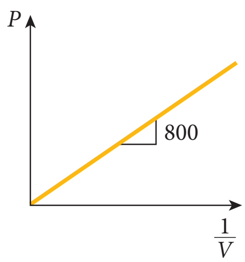


Figure A4.11 ▲
Graph of P against $\frac{1}{V}$

Plot P against V. The graph is of the form shown in Figure A4.10.

This suggests that the relation is either inverse $\left(\frac{1}{V}\right)$ or inverse square $\left(\frac{1}{V^2}\right)$.

It could even be to a higher order, but this will not be dealt with here. The next step is to plot the graph of P against $1/V$, as shown in Figure A4.11.

Because this graph is a straight line, the equation can now be given exactly, and is in this case:

$$P = \text{gradient} \times \frac{1}{V} \quad \text{or} \quad P = \frac{800}{V}$$

Recognition of graph shapes is extremely useful. When plotting data, you can quickly see whether you have the expected relationship or can plan further analysis (see Figure A4.12).

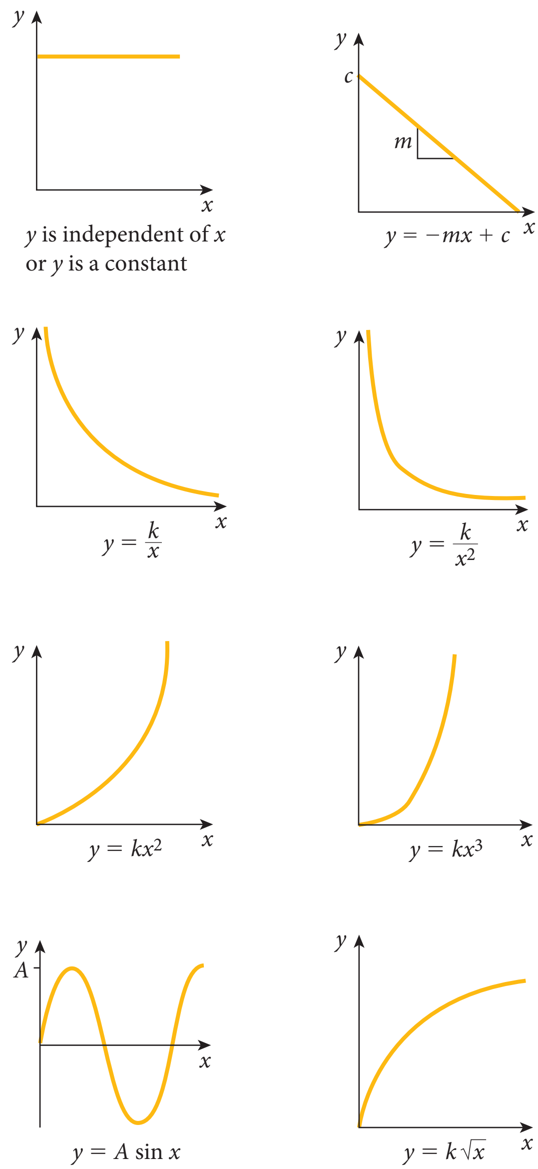


Figure A4.12 ▲
Graphs of important relationships

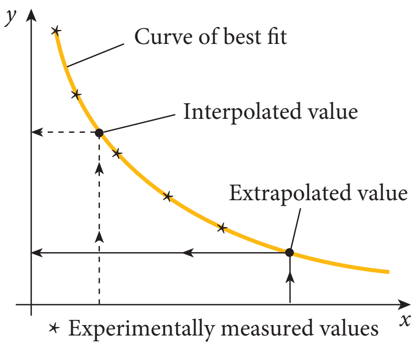
The gradient and the area under a graph may have physical meaning. The units of the gradient are the units of the quotient of the units of the vertical axis and units of the horizontal axis. The units of the area are the product of the units on the axes.

Interpolation and extrapolation

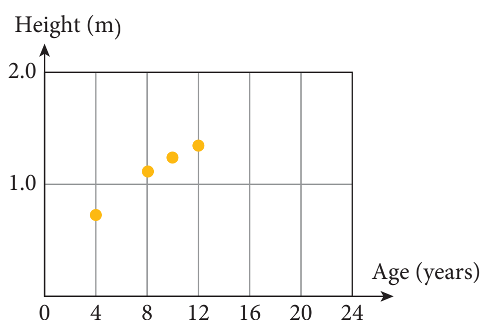
Interpolation is the estimation of a quantity between data points. Interpolation is only valid when there is good reason to believe that the graph is a smooth curve between the data points.

Extrapolation is the estimation of quantities beyond the domain and range of the data. This is less reliable than interpolation because the exact shape of the graph is unknown beyond the measured values.

In general, it is wise to be sceptical of the first and last data points on the graph. This means that, for simple relationships, you will need a minimum of three data points between the first and last points. The data points should also be spread out relatively evenly to ensure the data field is adequately sampled. There is little point in collecting three points near the first and a last point a long way away.



▲ Figure A4.13
Interpolated values are between measured values. Extrapolated values are outside the range of measured values.



▲ Figure A4.14
From this graph of the height of a student against age, predict height at birth, at 6 years of age and at 24 years of age. Comment on the accuracy of your predictions.

Numerical analysis

Another way to test for proportionality is to find the constant directly from the data coordinates. That is, you use the numbers directly to find whether the suspected relationship applies. If you refer back to Table A4.2 of P against V, you can see that, as P gets bigger and V gets smaller. This suggests an inverse relation. To check this numerically, we can use the following:

If: $P = \frac{c}{V}$, where c is a constant

then: $PV = c$

From Table A4.2 on page 450, the product PV is equal to 810, 795, 794 and 792. This is close enough to be considered a constant within the limits of experimental uncertainty. The equation becomes $PV = 798$ (the mean of the above values). This is close to 800, which is the value found graphically.

Graphical and numerical methods must both be evaluated against the data and the experimental uncertainty. Neither is superior.

If we had found that PV was not a constant, then we might have tried:

$$P = \frac{c}{V^2}$$

That is, $PV^2 = c$.

For linear relationships, $A = cB$, where c is a constant. Parabolic, square relationships, have $A = cB^2$, so: $A/B^2 = c$ etc.

Numerically measured values from the table can be substituted to check that the value of c is constant.

Power laws

Consider a series of cubes of sides 1.0, 2.0, 3.0 and 4.0 cm. If we calculate the perimeter of one side, the area of one side, the total surface area, the volume and the mass in terms of the 1.0 cm cube, we can find a relationship between the values.

The values are given in tabular form in Table A4.3.

1 Perimeter (P): There is a direct relationship between the perimeter and the length. If one is doubled, then the other is doubled, and so on.

$$P \propto l$$

2 Area (A): There is a square relationship between the area and the length of one side. If the length is doubled, then the area is increased fourfold (2^2), tripling the length increases the area ninefold (3^2), etc.

Table A4.3 Relationship between values and length of side

Length of side (cm)	Perimeter (cm)	Area of side (cm^2)	Total surface area (cm^2)	Volume (cm^3)	Mass units
1.0	4.0	1.0	6.0	1.0	1.0
2.0	8.0	4.0	24.0	8.0	8.0
3.0	12.0	9.0	54.0	27.0	27.0
4.0	16.0	16.0	96.0	64.0	64.0

This can be seen to be true for both side and surface area, even though the values are different.

$$A \propto l^2$$

3 Volume (V): There is a cubic relationship between the volume and the length of one side. If length is doubled, then the volume is increased eightfold (2^3); if it is increased by three, the volume is increased 27-fold (3^3), etc.

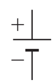
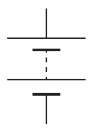
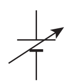
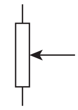
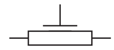
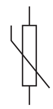
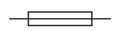
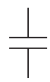
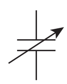
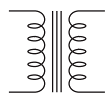
$$V \propto l^3$$

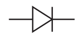
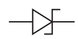
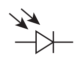
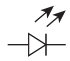
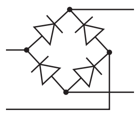
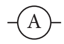
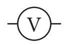
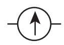
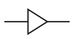
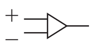
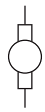
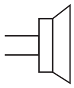
4 Mass (m): If we assume that the density remains constant, then the mass changes as the volume changes. That is, a cubic relationship exists between side length and mass. If the density changes, the value cannot be found unless other facts are known.

CURVE FITTING

For further information on curve fitting and data analysis, go to this website.

Appendix 5: Electric and electronic symbols

Component group	Component	Symbol
Sources of *emf*	Cell	
	Battery, DC power supply	
	Variable DC power supply	
	AC power supply	
Resistance	Resistor	
	Rheostat, resistor with sliding contact, potentiometer, voltage divider	 or 
	Variable resistor	
	Light-dependent resistor (LDR)	
	Filament globe	
	Thermistor	
	Fuse	 or 
Capacitance	(Non-polarised) capacitor	
	Variable capacitor	
	Polarised capacitor, electrolytic capacitor	
Transformer	Iron-cored transformer (one secondary winding)	

Component group	Component	Symbol
Diodes	Junction diode	
	Zener diode	
	Photodiode	
	Light-emitting diode (LED)	
	Four-diode bridge	
Meters	Ammeter	
	Voltmeter	
	galvanometer	
	cathode ray oscilloscope (CRO)	
Amplifiers	Voltage amplifier	
	Operational amplifier (op amp)	
Transducers	Motor	
	Microphone	
	Loudspeaker	
External connections	Earth	
Circuit connections	Non-connected leads	
	Connected leads	 or 

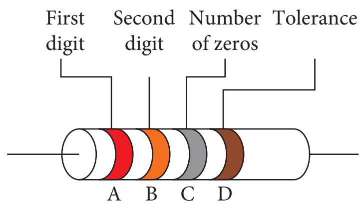
Appendix 6: Resistance codes

A resistor is a physical object or circuit element that has resistance. The resistance of an actual, physical resistor is indicated in one of two ways:

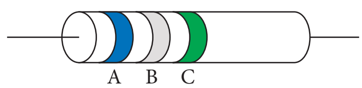
- A colour code: a set of stripes on the resistor to indicate the resistance and the tolerance or uncertainty in the value
- A number-letter code: numbers with a letter to mark the decimal point

Resistance colour code

The resistor is marked with three or four colour bands painted on the resistor near one end as shown in Figures A6.1 and A6.2.



▲ **Figure A6.1**
A resistor marked with four coloured bands



▲ **Figure A6.2**
A resistor marked with three coloured bands

The colour of band A nearest the end is the first digit in the resistance value. The colour of band B represents the second digit. The colour of band C gives the number of zeros to follow these two digits (Table A6.1).

Table A6.1 Colour code for digits

Digit	Colour
0	Black
1	Brown
2	Red
3	Orange
4	Yellow
5	Green
6	Blue
7	Violet
8	Grey
9	White

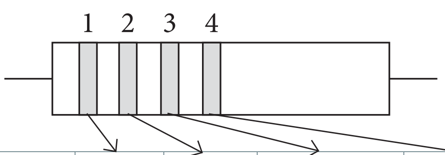
The fourth band from the end indicates the tolerance or percentage uncertainty in the resistance value (Table A6.2).

Table A6.2 Colour code for tolerances of resistors

Tolerance (%)	Colour
1	Brown
2	Red
5	Gold
10	Silver
20	No band D shown

If you hold the resistor so the stripes are on the left, you may find it easier to work out the resistance and the tolerance. If there are only three stripes, that is no tolerance band, the percentage uncertainty is 20%.

Figure A6.3 summarises the colour code details.



Colour	Value	Value	Multiply by	Tolerance
Black	0	0	1	Red = ±2%
Brown	1	1	10	Gold = ±5%
Red	2	2	100	Silver = ±10%
Orange	3	3	1000	No band = ±20%
Yellow	4	4	10 000	
Green	5	5	100 000	This gives the maximum error in the value of the resistor.
Blue	6	6	1 000 000	
Violet	7	7		
Grey	8	8	not	
White	9	9	used	

Figure A6.3 ▲ Summary of resistance colour code system

Examples:

Stripe (reading from the stripe nearest to the end):

- A-B-C: yellow (4), violet (7), orange (1000), gold (±5) ⇒ (47 000 ± 5%) Ω% or (47 ± 3) kΩ
- A-B-C-D: orange (3), white (9), brown (10), silver (±10%) ⇒ (390 ± 10%) Ω or (390 ± 39) Ω
- A-B-C: brown (1), green (5), black (1) and no fourth colour ⇒ (15 ± 20%) Ω or (15 ± 3) Ω

Resistance number–letter codes

In this system, which is often used on circuit diagrams, the numeral may have a letter in front of, behind or between the digits. The resistance is given to two significant figures. The letters R, K and M are used as multipliers: R for '×1', K for '$\times 10^3$' or M for '$\times 10^6$'. The letters R, K or M are used to show where the decimal point goes.

Tolerances are given letter codes at the end (Table A6.3).

Table A6.3 Letter code for tolerances of resistors

Tolerance (%)	Letter
1	F
2	G
5	J
10	K
20	M

Examples:

- 2R5J ⇒ (2.5 Ω(2.5 ± 5%) Ω or (2.5 ± 0.1) Ω
- 47KM ⇒ 47 k Ω(47 ± 20%) kΩ or (47 ± 10) Ω
- M22K ⇒ (0.22 ± 10%) MΩ or (220 ± 22) kΩ

Appendix 7: Periodic table of elements

Key

atomic number → 26
Fe
name of element → iron
standard atomic weight → 55.85

Symbol of element:
- gas at room temperature
- liquid at room temperature
- solid at room temperature
- synthetic (does not occur naturally)

- s block
- p block
- d block transition metals
- d block lanthanoids and actinoids

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 H hydrogen [1.007, 1.009]																	2 He helium 4.003
3 Li lithium [6.938, 6.997]	4 Be beryllium 9.012											5 B boron [10.80, 10.83]	6 C carbon [12.00, 12.02]	7 N nitrogen [14.00, 14.01]	8 O oxygen [15.99, 16.00]	9 F fluorine 19.00	10 Ne neon 20.18
11 Na sodium 22.99	12 Mg magnesium [24.30, 24.31]											13 Al aluminium 26.98	14 Si silicon [28.08, 28.09]	15 P phosphorus 30.97	16 S sulfur [32.05, 32.08]	17 Cl chlorine [35.44, 35.46]	18 Ar argon 39.95
19 K potassium 39.10	20 Ca calcium 40.08	21 Sc scandium 44.96	22 Ti titanium 47.87	23 V vanadium 50.94	24 Cr chromium 52.00	25 Mn manganese 54.94	26 Fe iron 55.85	27 Co cobalt 58.93	28 Ni nickel 58.69	29 Cu copper 63.55	30 Zn zinc 65.38(2)	31 Ga gallium 69.72	32 Ge germanium 72.63	33 As arsenic 74.92	34 Se selenium 78.96(3)	35 Br bromine [79.90, 79.91]	36 Kr krypton 83.80
37 Rb rubidium 85.47	38 Sr strontium 87.62	39 Y yttrium 88.91	40 Zr zirconium 91.22	41 Nb niobium 92.91	42 Mo molybdenum 95.96(2)	43 Tc technetium	44 Ru ruthenium 101.1	45 Rh rhodium 102.9	46 Pd palladium 106.4	47 Ag silver 107.9	48 Cd cadmium 112.4	49 In indium 114.8	50 Sn tin 118.7	51 Sb antimony 121.8	52 Te tellurium 127.6	53 I iodine 126.9	54 Xe xenon 131.3
55 Cs caesium 132.9	56 Ba barium 137.3	57–71 lanthanoids	72 Hf hafnium 178.5	73 Ta tantalum 180.9	74 W tungsten 183.8	75 Re rhenium 186.2	76 Os osmium 190.2	77 Ir iridium 192.2	78 Pt platinum 195.1	79 Au gold 197.0	80 Hg mercury 200.6	81 Tl thallium [204.3, 204.4]	82 Pb lead 207.2	83 Bi bismuth 209.0	84 Po polonium	85 At astatine	86 Rn radon
87 Fr francium	88 Ra radium	89–103 actinoids	104 Rf rutherfordium	105 Db dubnium	106 Sg seaborgium	107 Bh bohrium	108 Hs hassium	109 Mt meitnerium	110 Ds darmstadtium	111 Rg roentgenium	112 Cn copernicium		114 Fl flerovium		116 Lv livermorium		
		57 La lanthanum 138.9	58 Ce cerium 140.1	59 Pr praseodymium 140.9	60 Nd neodymium 144.2	61 Pm promethium	62 Sm samarium 150.4	63 Eu europium 152.0	64 Gd gadolinium 157.3	65 Tb terbium 158.9	66 Dy dysprosium 162.5	67 Ho holmium 164.9	68 Er erbium 167.3	69 Tm thulium 168.9	70 Yb ytterbium 173.1	71 Lu lutetium 175.0	
		89 Ac actinium	90 Th thorium 232.0	91 Pa protactinium 231.0	92 U uranium 238.0	93 Np neptunium	94 Pu plutonium	95 Am americium	96 Cm curium	97 Bk berkelium	98 Cf californium	99 Es einsteinium	100 Fm fermium	101 Md mendelevium	102 No nobelium	103 Lr lawrencium	

NUMERICAL ANSWERS

Chapter 1

Question set 1.2

2 0 K

3 Celsius

5 a $K = C + 273$

b −2°C

Worked example 1.1

1 2772 J kg^{-1} °C^{-1}, cooking oil

2 6.78×10^{-3} kg = 6.78 g

Worked example 1.2

1 b 366.6°C

Question set 1.3

1 450 kJ raises 1 kg of iron by 1 K

3 Units: [J] = [kg] [c] [K] ⇒ units of c are J kg^{-1} K^{-1}

4 30.5°C

5 0.3°C

6 3480 J kg^{-1} K^{-1}

Worked example 1.3

1 12.6 kJ

2 a 1500°C (actual melting point is 1538°C)

b 2800°C (actual boiling point is 2862°C)

Worked example 1.4

1 21 633 J

Worked example 1.5

1 21.3 g

Question set 1.4

8 3.36×10^7 J or 33.6 MJ

9 1376 s or 22 min 56 s

Chapter review questions

11 319.2 K

12 a 36 090 J

b 0.015 kg or 15 g (Heat is also provided by the condensed steam cooling from 100°C to 80°C.)

13 2940 J h^{-1}

17 a 138°C

18 9455 J s^{-1} or 9.5 kW

21 1.36×10^8 J

23 a 71 kg

Chapter 2

Worked example 2.1

121.5 J s^{-1} or 121.5 W

Question set 2.2

5 200 000 kJ or 200 MJ

Question set 2.3

9 1.26×10^9 J

Worked example 2.2

1 a 18 J

b 1.30×10^5 J

2 a 1.30×10^6 J

b 1.17×10^6 J

3 120 kg

4 a 12°C

Question set 2.4

5 10 h 45 min or 10.75 h

Question set 2.5

8 a 150 kJ

b 40%

9 a 4.7

b 560 W

Chapter review questions

3 $\Delta E = Q_{in} - Q_{out} - W$

ΔE is the change in energy to the system. Q_{in} and Q_{out} are energies entering and leaving the system respectively. W is the work done by the system.

7 North; at 62.5° to the horizontal (27.5° to the vertical)

Chapter 3

Question set 3.2

1 a alpha, beta, gamma

b/c α or ${}^4_2\alpha$ or ${}^4_2He$, + 2, β^- or e^-, −1, or β^+ or e^+, + 1, γ, 0.

Worked example 3.1

1 10.80

2 121.86

Question set 3.3

2 a Z (protons)
b 1 u = mass of a carbon-12 atom

5 $^{99m}Tc \rightarrow Tc + \gamma$. (There is no nuclide change.)

6 79

7 152.04

9 $^{98}_{42}Mo + ^{1}_{0}n \rightarrow ^{99}_{42}Mo$

Worked example 3.2

1 a $^{211}_{87}Fr \rightarrow ^{207}_{85}At + ^{4}_{2}He$
b Astatine-207

2 a $^{213}_{84}Po \rightarrow ^{209}_{82}Pb + ^{4}_{2}He$
b Lead-209

Worked example 3.3

1 a $^{207}_{81}Tl$ (Thallium-207)
b $^{82}_{36}Kr$ (Krypton-82)
c $^{190}_{78}Pt$ (Platinum-190)
d $^{145}_{62}Sm$ (Samarium-145)

2 a $^{200}_{80}Hg$ (Mercury-200)
b $^{209}_{83}Bi$ (Bismuth-209)
c $^{199}_{81}Tl$ (Thallium-199)
d $^{203}_{82}Pb$ (Lead-203)

Question set 3.4

2 a $^{A}_{Z}X \rightarrow ^{A-4}_{Z-2}Y + ^{4}_{2}He^{2+}$
b $^{A}_{Z}X \rightarrow ^{A}_{Z+1}Y + ^{0}_{-1}e + \overline{\nu}$
c $^{A}_{Z}X \rightarrow ^{A}_{Z-1}Y + ^{0}_{+1}e + \nu$
d $^{A}_{Z}X \rightarrow ^{A}_{Z}X + \gamma$

4 Neon-21

5 a $^{151}_{67}Ho \rightarrow ^{147}_{65}Tb + ^{4}_{2}He^{2+}$
b Terbium-147

6 a alpha particles
b $^{210}_{86}Rn \rightarrow ^{206}_{84}Po + ^{4}_{2}He^{2+}$

7 a $^{15}_{8}O \rightarrow ^{15}_{7}N + ^{0}_{+1}e + \nu$
b $^{18}_{9}F \rightarrow ^{18}_{8}O + ^{0}_{+1}e + \nu$ (for the fluorine-18)

8 $^{158}_{65}Tb \rightarrow ^{154m}_{63}Eu + ^{4}_{2}He^{2+}$
$^{154m}_{63}Eu \rightarrow ^{154}_{63}Eu + \gamma$

9 $^{198}_{79}Au \rightarrow ^{198}_{80}Hg + ^{0}_{-1}e + \overline{\nu}$ It emits an electron.

Question set 3.5

1 a γ rays
b α rays
c α rays

2 a α, β^{+}
b γ

3 α, as they have by far the highest mass

5 a less; $\frac{1}{9}$ of the count rate
b more; 4 times the count rate

6 a 10
b 20
c 100

7 a $80\,s^{-1}$
b $\pm 2\,s^{-1}$, $\pm 0.003\,s^{-1}$

Worked example 3.4

1 a $^{232}_{90}Th \rightarrow ^{228}_{88}Ra + ^{4}_{2}He^{2+}$
$^{228}_{88}Ra \rightarrow ^{228}_{89}Ac + ^{0}_{-1}e + \overline{\nu}$
b Radium-228, Actinium-228

2 a $^{235}_{92}U \rightarrow ^{231}_{90}Th + ^{4}_{2}He^{2+}$
$^{231}_{90}Th \rightarrow ^{231}_{91}Pa + ^{0}_{-1}e + \overline{\nu}$
b Thorium-228, Protactinium-228

Question set 3.6

1 $N = N_0\left(\frac{1}{2}\right)^n$

2 b U-238, U-235 Th-232, Np-237
c Pb-206, Pb-207, Pb-208, Tl-205

4 $\frac{15}{16}$

5 a $^{226}_{88}Ra \rightarrow ^{222}_{86}Rn + ^{4}_{2}He^{2+}$
b $^{210}_{82}Pb \rightarrow ^{210}_{83}Bi + ^{0}_{-1}e + \overline{\nu}$
$^{210}_{83}Bi \rightarrow ^{210}_{84}Po + ^{0}_{-1}e + \overline{\nu}$

6 a 1.5×10^{24}
b 1.5×10^{23}

7 a $^{11}_{5}B + ^{1}_{0}n \rightarrow ^{12}_{5}B^{*}$
$^{12}_{5}B \rightarrow ^{12}_{6}C + ^{0}_{-1}e + \overline{\nu}$
b $^{191}_{80}Hg + ^{1}_{0}n \rightarrow ^{192}_{80}Hg^{*}$
$^{192}_{80}Hg \rightarrow ^{192}_{81}Tl + ^{0}_{-1}e + \overline{\nu}$

8 4.7×10^{25}

Chapter review questions

4 a $^{A}_{Z}X \rightarrow ^{A-4}_{Z-2}Y + ^{4}_{2}He^{2+}$
b $^{A}_{Z}X \rightarrow ^{A}_{Z+1}Y + ^{0}_{-1}e + \overline{\nu}$
c $^{A}_{Z}X \rightarrow ^{A}_{Z-1}Y + ^{0}_{+1}e + \nu$
d $^{A}_{Z}X \rightarrow ^{A}_{Z}X + \gamma$

11 a 122
b 204.41

12 a $^{190}_{78}Pt \rightarrow ^{186}_{76}Os + ^{4}_{2}He^{2+}$
b Osmium-186

13 $^{119}_{53}\text{I} \rightarrow {}^{119}_{52}\text{Te} + {}^{0}_{+1}\text{e} + \nu$

$^{119}_{52}\text{Te} \rightarrow {}^{119}_{51}\text{Sb} + {}^{0}_{+1}\text{e} + \nu$

15 a $^{99}_{42}\text{Mo} \rightarrow {}^{99m}_{42}\text{Tc} + {}^{0}_{-1}\text{e} + \bar{\nu}$

b $^{98}_{42}\text{Mo} + {}^{1}_{0}\text{n} \rightarrow {}^{99m}_{43}\text{Tc} + {}^{0}_{-1}\text{e} + \bar{\nu}$

16 b $^{99m}_{42}\text{Tc} \rightarrow {}^{99}_{42}\text{Tc} + \gamma$

c 30 hours (5 half lives)

17 1.4 minutes or 84 s

19 a i Irid: 24.4 s^{-1}; Bism: 25.1 s^{-1}

ii Irid: ±2.4 s^{-1}; Bism: ±0.25 s^{-1}

20 1.1×10^{31}

Chapter 4

Worked example 4.1

1 a 1.508×10^{-10} J

b 941 MeV

Question set 4.1

2 $\Delta E = (\Delta m)c^2$ Energy, mass, speed of EMR

6 a Iron-56 with 26 protons and 30 neutrons

7 a 0.47 MeV or 470 keV

b 7.5×10^{-14} J

Worked example 4.2

1 a 4

b $^{233}_{92}\text{Mo} + {}^{1}_{0}\text{n} \rightarrow {}^{104}_{42}\text{Mo} + {}^{126}_{50}\text{Sn} + 4\,{}^{1}_{0}\text{n}$

c i 0.194 u

ii 3.22×10^{-28} kg

d 2.90×10^{-11} J

2 a $^{133}_{51}\text{Sb}$

b $^{239}_{94}\text{Pu} + {}^{1}_{0}\text{n} \rightarrow {}^{104}_{43}\text{Tc} + {}^{133}_{51}\text{Sb} + 3\,{}^{1}_{0}\text{n}$

c 132.93 u

d 2.84×10^{-11} J

Question set 4.2

8 $^{10}_{5}\text{B} + {}^{1}_{0}\text{n} \rightarrow {}^{11}_{5}\text{B}^*$

$^{11}_{5}\text{B}^* \rightarrow {}^{7}_{3}\text{Li} + {}^{4}_{2}\text{He}$

9 a 4 neutrons

b i 0.801 u

ii 1.33×10^{-27} J

c 1.20×10^{11} J

Worked example 4.3

1 a i 0.0059 u

ii 9.8353×10^{-29} kg

b 8.85×10^{-12} J

2 a i 0.0138 u

ii 2.300×10^{-29} kg

b 2.07×10^{-12} J

Question set 4.4

7 a $^{1}_{0}\text{n} + {}^{6}_{3}\text{Li} \rightarrow {}^{7}_{3}\text{Li}^*$

b $^{7}_{3}\text{Li}^* \rightarrow {}^{4}_{2}\text{He} + {}^{3}_{1}\text{H}$

8 D–T reaction produces neutrons.
$^{2}_{1}\text{H} + {}^{3}_{1}\text{H} \rightarrow {}^{4}_{2}\text{He} + {}^{1}_{0}\text{n}$. Neutron absorption in lithium releases energetic nuclides that transfer energy to the heat exchanger.

$^{1}_{0}\text{n} + {}^{6}_{3}\text{Li} \rightarrow {}^{4}_{2}\text{He} + {}^{3}_{1}\text{H}$

$^{1}_{0}\text{n} + {}^{7}_{3}\text{Li} \rightarrow {}^{4}_{2}\text{He} + {}^{3}_{1}\text{H} + {}^{1}_{0}\text{n}$

9 a 0.0021 u

b 3.150×10^{-13} J

Question set 4.5

2 $H = D \times W_R$

6 200 Gy

7 30 J

Chapter review questions

5 $H = \frac{E}{m} \times W_R$ or $H = D \times W_R$, H = equivalent dose, E = dose = total energy received, m = mass of irradiated body, W_R = radiation weighting factor, $D = \frac{E}{m}$ = absorbed dose

12 Carbon-12 is a neutron poison. Carbon-12 atoms absorb neutrons to become carbon-13, which beta decays to nitrogen-13.

$^{12}_{6}\text{C} + {}^{1}_{0}\text{n} \rightarrow {}^{13}_{6}\text{C}$

$^{13}_{6}\text{C} \rightarrow {}^{13}_{7}\text{N} + {}^{0}_{-1}\text{e} + \bar{\nu}$

14 0.0189 u

15 100 Gy

16 26 J

18 a 2

b i 0.1844 u

ii 3.07×10^{-28} J

c 2.77×10^{-11} J

Chapter 5

Question set 5.1

8 3.21×10^{18}

Question set 5.2

4 1.5 V

5 12 V

Worked example 5.1

1 a 7.49×10^{17}
  b 1.50×10^{-3} cm $= 1.50 \times 10^{-5}$ m
  c 1.50×10^{-5} m s^{-1}
  d 3.33×10^{5} s $= 92.6$ h
2 a 3.12×10^{17}
  b 2.4×10^{-4}
  c 3.48×10^{-5} m/s
  d 3.45×10^{4} s

Question set 5.3

5 630 mA
6 20 C
7 0.1 A or 100 mA
8 a 4.99×10^{17} electrons
  b 8.32×10^{-6} m
  c 8.32×10^{-6} m/s
  d 1.20×10^{5} s

Worked example 5.2

1 90 mA
2 2.25 A

Worked example 5.3

1 1.75 V
2 4 V
3 6 V

Question set 5.4

1 False
2 True
3 1.90 A
4 1.8 V
5 2.0 A
6 750 mV

Worked example 5.4

1 1.25×10^{-8} Ω m
2 0.009 Ω

Worked example 5.5

3 250 Ω

Question set 5.5

4 a 6.7 Ω
  b 20 V
5 432 V
6 300 Ω
7 a $\sqrt{18} : 1 \approx 4.24 : 1$
  b 1 A, 0.11 A

Worked example 5.6

1 a 4.3×10^{4} C
  b 120 V
  c 0.36 kW
  d 0.72 kWh

Question set 5.6

1 Rate of energy transformation 8.3×10^{-3} kWh
2 $P = \frac{E}{t}$; $P = VI$; $P = \frac{V^2}{R}$; $P = I^2R$
3 $E = VIt$
4 3.6×10^{5} J, 0.1 kWh
5 4.7×10^{-3} kWh
6 24 kJ
7 a 5.0 V
  b 5 A

Chapter review questions

11 7 V
12 75 mA
15 12 J
16 1.92×10^{-18} J
17 300 C
18 3.3 A
19 3.12×10^{18} electrons
20 12.0 V
21 0.88 V
22 12.0 V
24 a 2.0 C
  b 6.0 J
  c 40 s
  d 80 C

Chapter 6

Worked example 6.1

1 A c 0.75 A; d i 1.5 V; ii 4.5 V
  B c 1 mA; d i 6 mV; ii 6 mV
  C c 0.05 A; d i 0.5 V; ii 1 V

Worked example 6.2

1 A b 1.5 Ω; c 4 A; d i 3 A; ii 1 A
  B b 3 Ω; c 4 mA; d i 2 A; ii 2 A
  C b 0.1 Ω (0.0995 Ω); c 15.075 A; d i 75 mA; ii 15 A

Worked example 6.3

1 a 0.6 A
c 2.4 V
d 4.8 V
e 0.6 A

Question set 6.1

5 a 29.0 Ω
b 3.2 Ω
c 6.0 Ω
d 4.0 Ω
6 b 30.0 Ω
c 0.2 A
d 2.4 V, 2.4 V, 1.2 V
7 a 12 Ω
b 0.5 A
c 0.25 A, 0.25 A, 0.5 A
d 3 V
10 $I_T = 0.0823\ A, I_{100\,\Omega} = 0.0823\ A, V_{100\,\Omega} = 8.23\ V$
$V_{200\,\Omega} = 11.77\ V, I_{200\,\Omega} = 0.0589\ A,$
$V_{500\,\Omega} = 11.77\ V, I_{500\,\Omega} = 0.0235\ A$

Worked example 6.4

1 a 2 V
b 0.5 V

Question set 6.2

3 a 9 V
b 16 V
c 0.5 V
4 c $\frac{R_1}{R_2} = \frac{V_1}{V_2}$

Worked example 6.5

1 a 2.9 V
b 1.25 V

Worked example 6.6

a 7.5 μA
b $0.53 \times 10^5\ \Omega$
c 4 V

Question set 6.3

3 9 kΩ
6 a 950 Ω
b 1.9 V

Chapter review questions

3 $Series: R_T = \sum_1^n R_i = R_1 + R_2 + \ldots + R_n$
Current is constant.
$V_T = \sum_1^n V_i = V_1 + V_2 + \ldots + V_n$
Parallel: $\frac{1}{R_T} = \sum_1^n \frac{1}{R_i} = \frac{1}{R_1} + \frac{1}{R_2} + \ldots + \frac{1}{R_n}$
Potential difference is constant.
$I_T = \sum_1^n I_i = I_1 + I_2 + \ldots + I_n$
10 a 2.4 A
b 7.2 A
c 3
11 a 15 Ω
c i 0.8 A
ii 0.4 A
iii 0.4 A
iv 0.8 A
d i 4 V
ii 3.2 V
iii 3.2 V
iv 4.8 V
12 a 20 kΩ
b 0.4 A or 400 mA
13 a 4 kΩ
b 1.8 A
14 a $1.0 \times 10^4\ \Omega$
b 4.1 V
15 a 7.2 V
b 3.6 mA
16 c 3.15×10^{-5} A, 11.85 V

Chapter 7

Worked example 7.1

a i 250 m, direction: shops to Tia
ii 400 m, James to Tia
b i 900 m
ii 400 m, home to James

Worked example 7.2

a 50 m N53°W
b 90 km S55°W

Worked example 7.3

35 km

Worked example 7.4

73 $km\,h^{-1}$

Worked example 7.5

a i 32 m

ii 16 m

b 6 $m\,s^{-1}$

Question set 7.1

4 a $\frac{\text{rise}}{\text{run}} = \frac{\text{speed}}{\text{time}} = (LT^{-1}) \div T = LT^{-2}$

b Area = length × height = T × (LT^{-1}) = L

5 120 $km\,h^{-1}$

6 a 0 m

b 0 $m\,s^{-1}$

c 12.0 m

d 2.9 $m\,s^{-1}$

7 a Displacement: 130 km, N67°W;
Velocity: 24 $m\,s^{-1}$, N67°W

b Displacement: 604 m, N21°W;
Velocity: 69 $m\,s^{-1}$, N21°W

c Displacement: 1636 m, S64°W;
Velocity: 0.78 $m\,s^{-1}$, S64°W

8 a 43.3 $km\,h^{-1}$

9 150 s or $2\frac{1}{2}$ minutes

10 160 $km\,h^{-1}$

Worked example 7.6

a 8.0 $m\,s^{-2}$

b −1.9 $m\,s^{-2}$

Worked example 7.7

a 100 m

b 475 m

c 0.5 $m\,s^{-2}$

Worked example 7.8

Final speed occurs at 4.0 s

Worked example 7.9

a 64 m

b 548 m

c i 0.32 $m\,s^{-2}$

ii −0.67 m^{-2}

e 50 s

Question set 7.2

1 $v = \frac{\Delta d}{\Delta t}$

5 a i About 4 $m\,s^{-2}$

ii 0 $m\,s^{-2}$

iii About −1 $m\,s^{-2}$

b About 27 m

c At the start

d About 9.8 $m\,s^{-2}$ (just under 10 $m\,s^{-2}$)

6 a 3 $m\,s^{-2}$

b 54 m

8 b i 2.5 $m\,s^{-2}$

ii 12.5 $m\,s^{-1}$

iii 125 m

9 a 6.45 m

b 43.54 m

c 4.18 s

d 4.49 s

e 50 m

f −4.98 $m\,s^{-2}$

Worked example 7.10

1 b 4.5 s

c 99.2 m

Worked example 7.11

b 4.1 s

c 41 m

Worked example 7.12

a 1.0 s

b 9.8 $m\,s^{-1}$

Question set 7.3

1 a 50.1 $m\,s^{-1}$

b 5.1 s

c 19.6 $m\,s^{-1}$

d 3.6 s after being dropped

2 a 11.025 m

b 14.7 $m\,s^{-1}$

c After 0.51 s

d 9.8 $m\,s^{-2}$ downwards

3 a Object A: 51.7 $m\,s^{-1}$; Object B: 49.5 $m\,s^{-1}$

b 1.3 s

c 30 m

4 4.5 s

5 Pelican goes hungry; the fish has 0.24 s to move away.

Chapter review questions

9 a Area = length × height = T × (LT^{-1}) = L, the unit of displacement

10 a 9.43 $m\,s^{-1}$

11 5 paces take 6 s, so average speed is 6 ÷ 5 = 1.2 paces s^{-1}.

Displacement = +1 pace in 5 s, so average velocity = 1 ÷ 5 = 0.2 paces s^{-1}.

12 a Displacement: 260 km, S67°W; Velocity: 36 $m\,s^{-1}$, S67°W

b Displacement: 604 km, S69°W; Velocity: 412 $km\,h^{-1}$, S69°W

c Displacement: 164 m, S18°W; Velocity: 0.16 $m\,s^{-1}$, S18°W

14 a 34.3 $m\,s^{-1}$

b 60.0 m

15 b i 0.46 m

ii 10.46 m

iii 10.0 m, down

iv 1.77 s

16 b 12 s

17 a 2.26 s

b 22.1 $m\,s^{-1}$

20 6.53 m above cliff

21 a 2.7 s (40 km/h), 5.4 s (80 km/h)

b 4.11 $m\,s^{-2}$

22 a 447 m, S33°E

Chapter 8

Question set 8.1

9 c $\Sigma F(\text{on A}) = F(\text{by C on A}) - F(\text{by B on A})$

Worked example 8.3

1 467 N

2 6.7 s

Question set 8.3

1 b Either $\Sigma\vec{F}(\text{on object}) = m\vec{a}$ or

$$\vec{a} = \frac{\Sigma\vec{F}(\text{on object})}{\text{mass of object } (m)}0$$

3 0 N

5 500 N

Worked example 8.5

1 $9.4 \times 10^{-3}\ m\,s^{-2}$, N58°E

2 S50°W (220°)

Worked example 8.7

1 21.6 N, 7.9 N

2 $R\cos\theta$, $R\sin\theta$

Worked example 8.8

a 13.8 N

b 5.0 N

c 3.4 $m\,s^{-2}$

Question set 8.4

3 a 26 N

b 15 N

c 0.30 $m\,s^{-2}$, N30°E

4 b 3.8 $m\,s^{-2}$

c 2.4 $m\,s^{-2}$

5 98 N

8 a 2.8 N

b 3.4 N

c 7.5 $m\,s^{-2}$

9 b 1.45 N

10 a 6.67 $m\,s^{-2}$

b 20 N, left

c 10 N, right

Chapter review questions

8 640 N N51°W

9 640 N

10 200 or 200 : 1

12 a i 262 N

ii 291 N

b i 262 N parallel to the plane in the 'down' direction

ii 291 N perpendicular to the plane, away from the plane

13 a 634 N

b 170 N

c 2.5 $m\,s^{-2}$

17 b i 3.0×10^3 N away from the ship

ii -5.2×10^3 N (as it is downwards)

iii 2.6×10^3 N

iv 1.5×10^3 N

v -2.6×10^3 N

19 a 58.7 N

b 573.2 N (from his weight) or 514.5 N (net force including friction)

c 5.4 $m\,s^{-2}$

20 a $8.0\,m\,s^{-2}$

b i 120 N

ii 24 N

iii 120 N

iv 24 N

Chapter 9

Worked example 9.1

1 40 m

2 43.2 m

Question set 9.1

5 $W = \Delta E$

6 Kinetic energy: 1.39×10^5

Potential energy: 1.96×10^5

$E_p \approx 1.4 \times E_k$

8 a 1.6×10^5 J

b 3.7×10^5 J

Worked example 9.2

1 9 J

2 gradient = double (twice as steep); force for an extension of 60 cm = 60 N

Question set 9.2

1 $W = Fs\cos\theta$

6 150 J

7 a $400\,N\,m^{-1}$

b 40 N

c 0.5 J

d 1.5 J

8 5.2×10^4 J or 52 kJ

Worked example 9.3

a 1.28 MJ

b $160\,m\,s^{-1}$

Question set 9.3

2 $P = \dfrac{\Delta W}{\Delta t}$

7 1.15 km

8 a 1.4×10^5 J

b 1.3×10^4 W

9 a 0.81 N

b 19.2 N

c $16.0\,m\,s^{-2}$

Worked example 9.4

1 $50\,m\,s^{-1}$ to the left

2 $1155\,kg\,m\,s^{-1}$, 1.4 Kn

Worked example 9.5

0.076 J

Worked example 9.6

7.5×10^3 N

Worked example 9.7

a 1.2×10^3 N

b 15 m

Question set 9.4

6 a 2.0×10^3 N s

b $2.0 \times 10^3\,kg\,m\,s^{-1}$

7 a $6\,m\,s^{-1}$

b 1.0 s

8 a 3.0 N s to the left

b $3.0\,kg\,m\,s^{-1}$, left

c $21.7\,m\,s^{-1}$ to the left

Chapter review questions

2 Nm or J

3 Work

4 Δt

16 6.7×10^4 J

17 a $6.75\,m\,s^{-1}$

18 a $500\,Nm^{-1}$

b 0.625 J

19 $24.2\,m\,s^{-1}$

20 $0.39\,m\,s^{-1}$

23 a 7.2 kg

b 1.44×10^3 J

c 4.1×10^5 N

d 0.0035 s

Chapter 10

Worked example 10.1

a 5.6 cm

b $83\,m\,s^{-1}$ or 0.083 s

Question set 10.1

3 $f = \dfrac{1}{T}$

4 c

6 3.5 s

7 $332\ \text{m s}^{-1}$

8 a transverse

b transverse

c longitudinal

d longitudinal

Question set 10.2

3 A

4 1 kHz

5 1.62 m

6 a 4 m

b 10 cm

8 a 170 Hz

b *A*, *C*, *E* and *G*

Worked example 10.2

7.7 cm

Question set 10.3

8 36 m

Worked example 10.3

a 1.32 m

b 0.33 m

Question set 10.5

3 a $\frac{\lambda}{2}$

b $\frac{\lambda}{4}$

c $\frac{\lambda}{4}$

4 B and C

5 a 1.60 m

b 125 Hz

6 24 cm

7 60 Hz

Worked example 10.4

1 0.27 m

2 a 2.2 m

b 155 Hz

Worked example 10.5

a 0.46 m

b 1.3 kHz

Question set 10.6

2 d

3 $317\ \text{m s}^{-1}$

4 4

5 1.7 cm

Question set 10.7

4 b 41 dB

5 a About 3000 kHz

b About –8 dB

6 80 dB

Question set 10.8

6 a 6.6×10^{-5} s or 66 μs

Chapter review questions

6 a 1.0 m

b 0.5 m

7 33 cm

9 a 170 Hz, 510 Hz, 850 Hz

b 340 Hz, 680 Hz, 1020 Hz

10 a 1.20 m

b 300 Hz

11 a 1.06 m

b 53 cm

c 53 cm

12 a 2.27 m

b F

c 0.57 m

d 450, 750,1050, … General rule is $f_n = (2n - 1) \times 150$ Hz , where n is any positive integer.

15 a 142 Hz

b 283 Hz, 425 Hz; Note: $f_1 = 141.666\ldots$ (≈ 142) so $f_2 = 2 \times f_1 = 283.333\ldots \approx 283$

c B

17 Answers below are for an estimate of 1.0 m.

a $\lambda_1 = 4.0$ m

b $f_1 = 85$ Hz

19 a 1150 Hz

b 89 cm

20 a $333\ \text{m s}^{-1}$

b $\pm 53\ \text{m s}^{-1}$

c $333 \pm 53\ \text{m s}^{-1}$

d Yes, as it lies within $333 \pm 53\ \text{m s}^{-1}$.

21 a 0.11 Hz

22 a 1133 – 1700 Hz

Chapter 11

Worked example 11.2

a $25\ \text{W m}^{-2}$

b $625\ \text{W m}^{-2}$

Question set 11.2

8 10^{-14}; 10^{-7}; 10^{5}

9 a $8.0\ \text{W m}^{-2}$

b $2.0\ \text{W m}^{-2}$

c $0.5\ \text{W m}^{-2}$

d $0.3\ \text{W m}^{-2}$

Question set 11.3

6 1.65 m

Worked example 11.3

1 9.3°

2 23.6°

3 33.3°

Question set 11.4

2 $\frac{\sin i}{\sin R}$ is constant

7 a 20°

b Smaller (23°)

8 a i 3.06×10^{14} Hz

ii 3.06×10^{14} Hz

b $2.16 \times 10^{8}\ \text{m s}^{-1}$

c 17.7°

10 a 81.88°

b 8.12°

Worked example 11.4

1 a 10 cm behind the lens

b real, inverted

c 6.0 cm

d 1

2 a 15 cm behind the lens

b real, inverted

c 6 cm

d 1.5

Worked example 11.5

1 a 10 cm in front of the lens (−10 cm), which is the focus

b virtual, upright

c 2.0

d 4.0 cm

2 a 24 cm in front of the lens (−24 cm)

b virtual, upright

c 4.0

d 8.0 cm

Worked example 11.6

1 a 10 cm in front of the mirror (the same position as the object)

b real, inverted

c 6.0 cm

d 1.0

2 a 10 cm in front of the mirror

b real, upright

c $\frac{2}{3}$ (0.67)

d $2\frac{2}{3}$ cm (2.67 cm)

Question set 11.5

3 $\frac{1}{f} = \frac{1}{u} + \frac{1}{v}$ and $M = \frac{h_1}{h_0} = -\frac{v}{u}$

5 20 cm (same side)

6 20 cm the opposite side of the mirror, virtual, 10.0 cm tall

7 14 cm the opposite side of the lens, real, 6.0 cm tall

8 22.5 cm in front of the lens

Worked example 11.7

1 a 452 nm

2 72.7 nm

Question set 11.6

2 a Path difference $= n\lambda$, $n = 1, 2, 3, \ldots$

b Path difference $= (2n - 1)\frac{\lambda}{2}$ or $\left(n - \frac{1}{2}\right)\lambda$, $n = 1, 2, 3 \ldots$

5 500 nm

8 a 400 nm

b 20 mm

Chapter review questions

2 $I = \frac{1}{4\pi}\left(\frac{S}{r^2}\right)$

I is intensity at a distance of r m and S is the power of the source.

3 $\frac{\sin i}{\sin R}$ is constant

6 3.06×10^{14} Hz

7 b i 1.47

ii 90°

13 a 24.9°

b i 6.19×10^{14} Hz

ii 318 nm

iii 1.96×10^{8} m s^{-1}

14 a 8.0 cm

b 24 cm (in front of the mirror)

15 a 24 cm (behind the lens)

b real and inverted

c 8.0 cm

16 594 nm

19 a 1.4×10^{-2} W m^{-2}

b 1.2 W m^{-2}

21 b 3

22 48.8°

23 9.96°

25 a 16.1 mm

b 716 nm

GLOSSARY

absolute refractive index a measure of the refrangibility of a medium placed in a vacuum and subjected to an incident ray of light

absolute zero the theoretical lowest possible temperature −273.15°C on the Celsius scale or 0 K on the absolute or kelvin scale

absorbed dose energy per mass absorbed in a body (Gy)

acceleration time rate of change of speed (or velocity)

accurate the degree to which a measurement result approaches the 'true value'

active connection to the 240 V household supply

adiabatic process a process that occurs without the exchange of heat between a system and its environment.

alpha particle/alpha ray first known radioactive particle; it is a helium-4 nuclide with two positive charges

alternating current (AC) current that changes direction periodically

amplitude the maximum displacement of the particle of a wave from its mean position; also a measurement of the vertical range of a curve

analogue a device or scale that gives a continuous measurement; the scale is continuous and may show any value in a range

analyser material that allows or stops polarised electromagnetic radiation

angle of incidence at a boundary, the angle between the normal and the incoming ray

angle of reflection at a boundary, the angle between the normal and the outgoing ray

angle of refraction the angle that the refracted ray makes with the normal line

anthropogenic human derived; caused by human activity

antineutrino a weakly interacting particle involved in energy transformations, especially beta-minus emission; thought to have little or no mass and zero charge

antinode point along a standing wave at which the wave has maximum amplitude; it is the result of a crest overlapping a crest or a trough overlapping a trough

Aristotelian following in the tradition of Aristotle

atom particle; originally an indivisible particle; now known to comprise several smaller particles

atomic mass mass of a nuclide compared to a single carbon-12 nuclide

atomic mass number (*A*) total number of nucleons in a nuclide

atomic number (*Z*) number of protons in a nucleus

atomic weight weighted average of all naturally occurring nuclides of an element in a sample

average speed, v_{ave} speed which, if maintained for the entire journey, results in the same distance travelled in the same time

average temperature a measure of the average kinetic energy of the matter in a defined system

background radiation radiation that is naturally present at a location

balanced forces forces applied to an object that result in a vector sum of zero on that object

best estimate value chosen to represent the indication value or measurement result of a measurand

beta particle/beta ray electron (beta-minus) or positron (beta-plus)

Big Bang theory theory about the history of the physical universe

binding energy energy needed to disassemble a nucleus into its component nucleons; measure of stability of a nuclide

body waves seismic waves that pass through the body of Earth

boiling point the temperature at which a liquid changes state from liquid to gas

braking distance distance travelled while braking

Brewster angle angle at which light from a surface is polarised

centre of mass the position in an object at which we measure its position when we model the object as a point-like particle

chromatic dispersion refraction occurs differently for different colours in the same material; colours spread

circuit breaker electromechanical switch that trips when there is an overload; safety protection against overload

cladding glass surrounding the core of an optical fibre

classical electrodynamics unified understanding of the mutual interactions of electricity and magnetism

closed system a system that matter cannot enter or leave, but energy can

combination circuit elements connected in series and in parallel groups

component the projection of a vector quantity along an axis

component of force the portion of a force acting in a given direction

compression region of high pressure in a mechanical wave

condensation the phase change from a gas to a liquid

condenser a vessel that removes heat from steam, allowing it to condense back to water

conduction the process by which heat energy is transferred by the collision of particles

conduction band a range of electron energies in which the electrons are relatively free to move away from the atom

conductor a type of material that allows electrons to flow

constructive interference waves interacting and increasing in amplitude at the point of interaction

contact force a force resulting from two objects coming into contact

continuous variable a variable that is able to take any value, sometimes within a fixed range; for example, a rainbow is a continuous spectrum

continuous wave continuous wavefronts passing through a medium

control rod rod made from a neutron poison, used to absorb excess neutrons in a nuclear reactor

controlled chain reaction a chain of nuclear reactions that are controlled to limit the rate at which reactions occur. In steady state (reaction rate held constant), an average of one neutron from each reaction goes on to cause another reaction. This is the case for a nuclear power reactor running at constant power output

controlled variable the variable that is controlled by the experimenter, so that its values are chosen; also called the independent variable

convection the transfer of energy through the movement of the cooler more dense fluids displacing less dense warmer fluids

convection cell the condition that occurs when there are density differences within a body of liquid or gas; the density differences result in rising and/or falling currents

convection current fluid circulating as a result of heating at a point or localised region; movement of fluids due to convection

converging (convex) lens lens thicker in the middle than at the ends

coplanar in the same plane

core inner glass of optical fibre

core temperature the temperature of the internal organs in the chest cavity, abdominal region and head in animals

Coriolis effect a deflection of moving objects when they are viewed in a rotating reference frame

cosmic radiation radiation whose origins are in space

count rate number of counts per second

crest the positive peak of a wave

critical angle angle of incidence for which the angle of refraction is 90° (total internal reflection occurs); beyond the critical angle reflection, but not refraction, occurs

critical temperature the temperature at which the electrical resistance of a material becomes zero

critical test an experiment that, having been rigorously conducted, that shows that one or other of competing theories is false

cryogenics the study of low-temperature phenomena

current the rate of flow of charge

daughter nuclide nuclide resulting from radioactive decay

decay the decrease in amplitude when a vibrating force has been removed

decay series cascade of decays from a radioactive nuclide until a stable nuclide is reached

deceleration negative acceleration or acceleration in the negative direction when speed is in the positive direction

decibel scale a logarithmic scale of sound level

deflection difference between a straight path and the actual path

delocalised valence electrons the outer electrons of metal atoms that are free to move

demodulation process of separating the signal from the modulated wave

density the mass per unit volume of a substance under specified conditions of pressure and temperature

dependent variable the variable that changes as a result of changes to the independent or controlled variable

derived data data that is deduced from raw data by mathematical manipulation, such as graphs, algebraic equations and geometric constructions

design brief the document that specifies the requirements for a design, including performance of the final product

destructive interference waves interacting and decreasing in amplitude at the point of interaction

diffraction the bending of a wave around objects or the spreading of the wave after passing through a gap

diffuse reflection/scattering reflection from a rough surface; rays in a beam reflect in different directions

diffusion the spontaneous movement of substances from regions of high concentration to regions of low concentration

digital able to measure only a limited number of possible values, usually within a fixed range

diode a semiconductor device that allows current to flow in one direction only

direct current (DC) current that is always in one direction

discrete quanta particular value; values between this and another value are not permitted

discrete variable a variable that is able to take only specific values, not continuous; for example, a line spectrum is a discrete spectrum

disintegration nuclear process leading to a new nuclide and emission of alpha, beta and/or gamma-rays

displacement, $\vec{s}$ change in position

displacement interval, s change in displacement

distance, d length travelled by an object during its motion

distance interval, Δd or s change in the distance

diverging (concave) lens lens that is thicker at the edges than at the centre

Doppler effect the shift in frequency and wavelength of waves that results from the relative motion of source and receiver

dose energy received from a radiation source (J)

double insulation functional and protective layers of insulation that enhance safety devices

drag force frictional force acting on an object that is moving relative to a fluid (liquid or gas)

drift velocity the speed of electrons moving as a result of the electrical effect

earth connection back to substation

Earth's energy balance the balance occurring when the amount of solar energy absorbed by Earth is equal to the amount of energy radiated back into space

eclipse movement of Earth into the Moon's shadow (solar eclipse) or movement of the Moon into Earth's shadow (lunar eclipse)

efficient cause the external force that was required to make things move in a horizontal direction, according to Aristotle

Einstein's mass–energy equation $\Delta E = (\Delta m)c^2$

elastic collision a collision between two or more objects in which there is no loss of total kinetic energy

electrical circuit a complete loop through which charges can flow

electrical permittivity, ε property of a medium that affects the propagation of the electrical component of an electromagnetic wave

electromagnetic radiation energy that travels as waves and moves at the speed of light

electromagnetic spectrum spectrum of electromagnetic waves from gamma rays, through visible light to radio waves

electromagnetic wave model light acts like a transverse wave that has electric and magnetic components

electron-volt, eV small energy unit; 1.602×10^{-19} J

element a substance that only has nuclides with the same number of protons

emf electromotive force; source of potential energy per charge

energy efficiency, η the effectiveness by which one form of energy is transformed into the desired energy

energy security the ability to acquire and protect national energy resources

enrichment process of separating out U-235 from a sample and adding it to another sample, increasing the proportion of U-235 in natural uranium

epicentre point on the surface directly above the seismic focus

equivalent dose combination of absorbed dose and relative effect in biological systems of particular ionising radiations

evaporation the process in which some particles with high kinetic energy escape the surface of a liquid at a temperature below its boiling point

explanation generalised account of why a body of data occurs

extrapolation extension beyond the measured range of data to read or construct new data that has not been measured

falsifiability principle used to determine the experimental data that would disprove a model, law or theory; data from a critical test

falsifiable able to be disproved

fast breeder reactor nuclear reactor that uses neutrons with high energies (fast neutrons) to cause fission events

fast neutron neutron with kinetic energy of 100 keV or more

first harmonic the simplest mode of vibration and accounts for the fundamental tone

first law of thermodynamics in the universe energy can be neither created nor destroyed; however, energy can change forms and energy can flow from one place to another within the universe. The total energy of an isolated system remains constant

fissile able to undergo fission; U-235, U-238, U-233, Pu-239

fission fragment nucleus produced as a result of fission; fission product

fission product nucleus produced in a fission event; fission fragment

fission the splitting of a heavy nucleus ($Z > 56$) into fragments with lower atomic numbers and neutrons; energy stored in the nucleus becomes available

fluid a gas or a liquid

focal length distance from lens to focal point

focal point light parallel to the axis of a lens or curved mirror is concentrated at this point

forced vibration vibration occurring when an object is forced to vibrate by another vibrating object

forward biased connected to allow current flow

frame of reference settings from which measurements are taken

free body diagram a diagram that shows the magnitude and direction of all the forces acting on a body using arrows

free or natural vibrations vibrations occurring when an object is vibrating by itself

frequency the number of whole waves or oscillations generated in one second

friction a contact force applied by one surface on another; the parallel component of the contact force

friction force force by one surface on another that affects the relative motion of the surfaces

fuse temperature-dependent wire that melts if an overload occurs; safety protection against overload

fusion the coming together of two nuclei to form a new nucleus with greater atomic number ($Z < 56$)

gamma rays high-energy electromagnetic radiation

gravitational acceleration, g acceleration of objects due to the gravitational force applied

gravitational force force by one mass on another mass

gravitational potential energy the energy associated with the gravitational force acting on an object

gravity common name given to the force applied by the very large mass of Earth on relatively smaller masses

greenhouse gases any gas that traps heat in the atmosphere; the primary greenhouse gases are water vapour, carbon dioxide, methane, nitrous oxide and ozone

half-life time taken for half a sample to decay

heat (n.) energy transferred due to a difference in temperature

heat, to heat (v.) the process of transferring energy due to a difference in temperature

heat conductor a material that readily allows the transfer of heat

heat engine a system that converts heat energy into work

heat insulator a material that is a poor conductor of heat

heat pump a system that moves heat energy from a cooler to a warmer area; work must be done on a heat pump

heat sink an object or material that moderates the temperature of its surroundings due to its large specific heat capacity; also an electronic component used to transfer heat effectively from components to the surrounding air

heat-conversion system a system in which heat is transformed into another form of energy

heat-exchange system a system in which heat is exchanged between two or more bodies of matter

high-grade energy resource energy form that can be converted into other energy forms with a high rate of efficiency

Hooke's law the force applied by a spring is proportional to the extension of the spring

hypothesis a tentative prediction, usually based on an existing model or theory; also a tentative explanation of an observation based on an existing model or theory

ideal spring a spring that obeys Hooke's law

image picture of an object

impulse the action of a force over a time interval; the change in momentum

independent variable a variable upon which another variable depends; also called the controlled variable

indication value a single result of a measurement; the indication value gives a hint as to the 'true value'

inelastic collision a collision between two or more objects in which momentum is conserved but total kinetic energy is not

inertia the tendency of an object to resist a change in its motion

inertial mass the mass of an object determined by measuring the acceleration when a known force is applied

infrasound a sound with a frequency below the range of human hearing

input transducer device that takes energy from the environment and converts it for use in an electric circuit

insolation incoming solar radiation

instantaneous speed, v_{inst} speed of an object at a moment in time

insulator a type of material that does not allow electrons to flow

insulators materials that are poor conductors of heat

intensity a measure of the energy passing per unit time through a unit area taken at right angles to the direction of propagation of the wave; energy per unit area per unit time

interference wave overlap

internal energy the sum of the kinetic energy of the particles in the system and the potential energy stored in the system

internal processes processes that occur within a defined system

interpolation to read or construct a new data point that has not been measured but is within the range of measured data

ionising power ability to ionise materials; inversely proportional to penetrating power

ionising radiation energy in particle or electromagnetic form that can affect the number of electrons surrounding a nucleus

isobar nuclides with the same mass (nucleon) number

isolated system a system that neither energy nor matter can enter or leave

isomer nuclides with the same atomic mass and nucleon number, but different energy states

isotone nuclides with the same number of neutrons

isotope any nuclide of the same atomic number (from Greek meaning 'equal type'); nuclide of an element that differs only with respect to the number of neutrons

joule (J) SI unit of energy; $1\,J = 1\,kg\,m^2\,s^{-2}$

kinetic energy the energy a body possesses due to its motion; it can be in the form of translational, rotational or vibrational energy

kinetic friction the force that occurs when two surfaces slide relative to one another; it always opposes the direction of motion

kinetic particle model the model that explains the properties of the different states of matter. The particles in solids, liquids and gases have different amounts of energy, are arranged differently and move in different ways

latent heat the heat required to change the state of a substance at its melting or boiling point without a change in temperature; unit: $J\,kg^{-1}$

law of conservation of energy in the universe energy remains constant it cannot be created or destroyed

LED light-emitting diode

light nuclide lightweight nuclide with small atomic mass (small nucleon number)

light-dependent resistor (LDR) device in which resistance depends on illumination

limit of reading the minimum uncertainty in a measurement due to the precision with which the scale can be read

line of best fit the line that most accurately fits the data, usually calculated using linear regression

linearise to make linear; to convert into a form that can be described by a straight line

logbook the record of an experiment or investigation kept by the scientist performing the experiments; it is a legal record of the experiments and their results

longitudinal wave a wave whose particles oscillate about a mean position in the same line as the direction of travel of the wave

loudness a subjective quality of perception of the amount of sound energy arriving at a person's ear

lower limit of audibility the lowest frequency a human can hear (approximately 20 Hz)

low-grade energy resource energy form that is converted into other energy forms with a low rate of efficiency

luminous a source that produces light

magnetic permeability, μ property of medium that affects the propagation of the magnetic component of an electromagnetic wave

magnification ratio of image height to object height; $M = \frac{h_i}{h_o}$

magnitude size

mass defect mass difference between the constituents of a nucleus and the mass of the nucleus; measure of the energy needed to hold a nucleus together

mass the amount of matter in an object

measurand quantity being measured

measurement result best estimate of a 'true value'; numerical value based on judgements about one or more attempts to measure the 'true value'

mechanical wave a disturbance that requires a material medium (solid, liquid or gas) for its propagation

melting the phase change from a solid to a liquid

melting point the temperature at which a substance undergoes a phase change from solid to liquid (melts)

metabolic activity the set of internal chemical reactions that maintain an organism's life

metal lattice a regular arrangement of large numbers of metal atoms, ions or molecules

metastable nuclide a nuclide that persists in an energy state above ground state for more than 10^{-12} s

model a representation of a system or phenomenon that explains the system or phenomenon. A model may be mathematical equations, a computer simulation, a physical object, words or other form

moderator light atoms in a nuclear reactor that slow down fast neutrons to thermal speeds in order to increase the likelihood of further fission events

modulation process of combining the signal with the carrier wave

momentum a quantity related to the action of a force over a time interval; it is the product of the mass and velocity of an object

natural cause the source of natural motion

natural motion the striving of an object to return to its natural place

negative energy problem a problem whereby the energy output from a fusion reactor is less than the energy input required to produce fusion

net force the sum of all the forces acting on a single object: $\Sigma\vec{F}$(on A)

neutral zero potential in a household wiring system

neutrino a weakly interacting particle involved in energy transformations, especially positron emission; thought to have little or no mass and zero charge

neutron neutral nuclear particle with mass slightly greater than that of a proton

neutron poisons nuclei including fission fragments that absorb a neutron, thus making the neutron unable to cause further fission events

node point along a standing wave at which the wave has minimal amplitude; it is the result of the overlap of a crest with a trough

non-contact force force that acts over a distance, including through a vacuum

non-ionising radiation energy in electromagnetic form that does not affect atomic electrons

non-luminous a source that reflects light

non-ohmic resistance is not constant; $R \neq \frac{V}{I}$

non-zero net force forces acting on an object that, when added, result in a net force greater than zero on the object

normal at right angles to a surface

normal force a force applied by a surface and which prevents surfaces sinking into each other; the perpendicular component of the contact force

nuclear binding energy total energy needed to hold a nucleus together; the greater the binding energy, the greater the stability

nuclear transformation nuclear process leading to a new nuclide and emission of alpha, beta and/or gamma-rays

nucleons nuclear particles; protons and neutrons

nucleosynthesis period in the history of the physical universe when the majority of nuclides formed

nuclide species of atom

objective measurement measurement that has a numerical value; the result of the measurement does not depend on the person taking the measurement

ohmic device a component with constant resistance for different values of V and I

open system a system that both matter and energy can enter and le

optical centre centre of curvature of a lens or mirror

optical fibre transparent light guide making use of total internal reflection at a boundary between two materials of similar refractive index

origin point of reference for measurements of position

outlier a data point that does not fit the pattern shown by other measured data points

output transducer device that uses energy from an electric circuit to convert to energy to be sent into the environment

P wave longitudinal compression waves that pass through the body of Earth

parallel circuit circuit with multiple paths through which current can flow

paraxial assumptions for curved lenses and mirrors: (a) lenses and mirrors must be small and thin; (b) rays must be near and parallel to the principal axis and (c) a straight line approximates the curved surface

parent nuclide original nuclide before radioactive emission

partial eclipse when an object is not entirely obscured from an observer by another object

penetrating power ability of ionising radiation to move into or through materials

penumbra shadow region where some light penetrates

percentage error proportional error expressed as a percentage; not defined by BIPM

percentage uncertainty proportional uncertainty, expressed as a percentage

period (*T*) the time it takes before a wave repeats itself

phase change a change of state (e.g. solid to liquid)

phon a measure of loudness; it depends on both intensity and the frequency

photoconductive mode mode in which photodiodes exposed to light conduct current in reverse biased direction

photovoltaic mode mode in which photodiodes exposed to light generate potential differences

pitch a subjective sensory characteristic related to frequency

plagiarism presenting someone else's work, including their words or ideas, as your own

plasma a collection of free-moving electrons and ions that can be accelerated by magnetic and electric fields

point source single localised source from which light transmits equally in all directions

polarisation orientation in one direction of the electric part of electromagnetic waves

polariser material that selects the direction of polarisation

potential difference (*V*) potential energy per charge; $V = \frac{W}{q}$, sometimes referred to as voltage

potential energy energy that can be considered to be 'stored' within a body due to its position, composition or molecular arrangement

power (*P*) the rate at which work is being done on an object or the rate of energy transfer

precise the degree to which individual measurements cluster around the mean

precision the variation in repeated measurements, or the uncertainty of a measuring device

primary data data that you have measured or collected yourself

principal axis line through both focus and centre of a curved lens or mirror system and perpendicular to the axis of the lens

proportional error difference between a measurement result and an accepted value, expressed as a fraction of the accepted value; not defined by BIPM

proportional uncertainty relative uncertainty

proton positively charged subatomic particle

pulse a single wavefront passing through a medium

qualitative non-numerical data; descriptive information

qualitative measurements measurements with descriptive or non-numerical results

quantitative numerical data; specific amount

quantitative measurements measurements with numerical values

quantum physics physics based on discrete states of matter and energy at the smallest levels

radiation energy transfer across space; the process by which heat is transferred without the need for a medium; energy from radioactive atoms

radiation weighting factor relative effect of types of ionising radiation on biological tissue

radioactive decay particles or rays that come from energy re-arrangements in a nucleus

radioactivity emission of energy in particle or electromagnetic form from the nucleus of an atom

random error a variation that affects a measurement in a random way so that the measurement is as likely to change in any one direction as in any other

rarefaction region of lower pressure in a mechanical wave

ratemeter counter that records counts per second

raw data original data taken directly from a measurement system

ray a line drawn at right angles to the wavefront and in the direction of propagation

ray diagram a geometric construction using the law of reflection that is used to find the magnification and position of the image

reaction distance distance travelled from the time the driver notices a situation and the time the brakes are first applied

reaction force the force that is applied to object B by object A when object A applies a force to object B

reaction time the time taken for a driver to react and apply the brakes from the time a situation is noticed

real image rays pass through the site of the image

reference the source of a specific piece of information or quotation; to state the source of information

reflection (law of) the angle of reflection (r) equals the angle of incidence (i)

refraction bending of waves as they pass through different media

refractive index measure of refrangibility; measure of the relative amount of change of direction of waves or light rays when travelling from one medium into another

refrangibility ability of a material to refract light

regular/specular reflection predictable reflection from a very smooth surface; rays in a beam all reflect in the same direction

relative atomic mass atomic weight

relative refractive index comparative difference in refrangibility between two media with different absolute refractive indices

relative uncertainty ratio of uncertainty to value

reliable highly likely to be true; a trustworthy source of information or reproducible data

representation model of reality

reproducible giving the same result, within uncertainty, when repeated measurements are made

research question the specific question that a particular experiment or investigation is designed to answer.

residual current device (RCD) earth leakage protection device; safety protection against overload

resistance opposition to the flow of electrons

resistivity, ρ how much a material opposes the flow of charges

resolution the limit of reading of a measuring device

resonance oscillation induced in a physical system when it is affected by another system that is itself oscillating at the right natural frequency

resultant force sum of forces

reverberation the effect that occurs when too many sound wave reflections arrive at your ear for you to distinguish between the sounds

reverse biased connected to prevent current flow

S wave a transverse earthquake wave that shakes the ground back and forth perpendicular to the direction the wave is moving; also known as a shear wave

scalar quantities quantities with magnitude only

scaler counter that records total counts

scatter graph a graph or plot showing data points, without a line joining the points, and used to demonstrate or determine a mathematical relationship between variables. The axes are defined by the variables

scintillation spark of light

secondary data data or information that has been collected by someone else

seismic focus the underground point from which earthquake energy is released

seismic reflection a technique developed for mapping the rock layers underneath the ground

seismic waves mechanical waves of energy that travel through Earth's layers that result from earthquakes, explosions or volcanic activity and are propagated within Earth or along its surface

seismograph a device that records the amplitude and frequency of seismic waves and yields information about Earth and its subsurface structure

semiconductor a material that conducts electricity less than conductors but more than insulators

sensible heat heat energy causing change in temperature

series circuit circuit with only one path through which the charge can flow

shear waves one of two types of seismic waves; also known as S waves, they are transverse waves whose velocities vary with the density of the rock they pass through

short circuit connection between two points that allows current to flow with negligible resistance

significant figure digit reported in a measurement result; the number of significant figures is the number of meaningful digits in a measurement result

slow neutron neutron with kinetic energy around 0.1–20 keV; thermal neutron

solidification phase change from liquid to solid

sound wave movement of energy by longitudinal vibration through a medium

specific heat capacity the amount of energy required to increase the temperature of 1 kg of a substance by one degree Celsius (or kelvin). It can be thought of as the resistance of a material to an increase in temperature; unit: $J\,kg^{-1}\,°C^{-1}$ or $J\,kg^{-1}\,K^{-1}$

specific latent heat of fusion the heat required to change the state of 1 kg of a specific substance at its melting point without a change in temperature

specific latent heat of vaporisation the heat required to change the state of 1 kg of a specific substance at boiling point without a change in temperature

speed the distance covered in a given time interval

stability curve plot of nuclides showing stable isotopes

static friction the maximum force applied by a surface to a stationary object to keep it from moving

stationary wave or standing wave a wave that oscillates in place, without transmitting energy along its extent. Standing waves have stable points, called nodes, where there is no oscillation

stiffness a measure of how much force is required to extend or compress a spring

stopping distance the total distance covered from the recognition of the hazard to coming to a stop; it is the sum of the reaction distance and the braking distance

stopping time the total time taken from the recognition of the hazard to coming to a stop, it is the sum of the reaction time and the braking time

strong nuclear force force that overcomes the electrostatic repulsion force at nuclear distances (approx. 15 fm)

subjective indication an estimate by our senses that depends on the person making the measurement

superconductivity occurs when the electrical resistance of a material drops to zero at a temperature below its critical temperature

superfluidity the property a material has when it has zero viscosity

superheated steam steam under high pressure that has been heated to a temperature higher than the boiling point of water

superposition when two or more waves of the same nature travel past a point of the medium, the resultant displacement of the medium at that point is given by the sum of the individual displacements due to the waves

systematic error an error that acts to give a consistent offset in data; for example, a zero error

temperature a measure of the average kinetic energy of the particles in a sample of matter, expressed in terms of degrees designated on a standard scale

temperature gradient the gradual change in temperature within a medium

terrestrial radiation radiation from materials in Earth's crust and core

theory a collection of models and concepts that explain specific systems or phenomena. Scientific theories allow predictions to be made and hence are falsifiable

thermal conductivity a measurement of how efficiently heat can be conducted through a material

thermal equilibrium the condition under which two substances in physical contact with each other do not exchange heat energy; the two substances are at the same temperature

thermal mass a large mass of matter that can absorb and retain heat energy

thermal neutron neutron with kinetic energy around 0.1–20 keV; slow neutron

thermal reactor nuclear reactor that uses neutrons with thermal energies (slow neutrons) to cause fission events

thermals rising air columns caused by convection

thermistor temperature-dependent resistor; used to detect changes in temperature

thermodynamics the physics of heating and cooling phenomena and their explanation

thermohaline circulation the continuous circulation of sea water due to variations in temperature and salinity

Thévenin's theorem all circuits can be reduced to a single source and a single load

thin lens equation for thin lenses and small curved mirrors: $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$

timbre the combination of qualities of a sound that distinguishes it from other sounds of the same pitch and volume; it is a subjective sound quality

time interval, t or Δt change in time

time-base scale a scale based on time

tone colour/quality the aspect of sound that allows the listener to identify the sound source or combination of sound sources; also called timbre

total eclipse when an object is entirely obscured from an observer by another object; total solar eclipse is when the Sun is completely blocked by the Moon

total internal reflection the sound wave is contained within the tube and is coherently reflected back and forth along the tube

transmutation nuclear process leading to a new nuclide and emission of alpha, beta and/or gamma rays

transuranic elements element beyond uranium; artificially produced nuclide with more than 92 protons

transverse wave a wave whose particles oscillate about a mean position perpendicular to the line of travel of the wave

trough the negative peak of a wave

true value the exact value of a measurand; the 'true value' is an ideal that can never be known with certainty

ultrasound a sound with a frequency above the human hearing range

ultrasonography using ultrasound for medical purposes

umbra shadow region where no light penetrates

uncertainty estimate of the range of values within which the 'true value' of a measurement or derived quantity lies

uncertainty bars bars drawn above and below and/or to left and right of a data point on a graph to indicate the size of the uncertainty in that point

uncontrolled chain reaction a chain of nuclear reactions that is not controlled. Usually this means a reaction rate that increases rapidly. For this to occur the average number of neutrons from each reaction that go on to cause more reactions is greater than one

upper limit of audibility the highest frequency a human can hear (approximately 20000 Hz)

valence band a range of electron energies for which the electrons are still attached to the atom; energy levels involved in chemical reactions

valid results that are affected only by a single independent variable and hence are reproducible

vaporisation occurs when a liquid changes to a gas; there is no temperature change during vaporisation

variable something that can change or be changed, as distinct from a constant, which does not

vector a quantity that has both magnitude and direction

vector quantities quantities with magnitude and direction

vector subtraction this is the geometric addition of the negative; $\vec{A} - \vec{B} = \vec{A} + (-\vec{B})$

velocity, $\vec{v}$ speed with an associated direction

violent force an Aristotelian force that changes the natural motion of an object

virtual image image of an object where the rays do not pass through the image; the image cannot be projected onto a screen

voltage divider device used to vary voltage at the output depending on a control resistor; also called a potential divider

wavefront an imaginary surface joining all points in space that are reached at the same instant by a wave propagating through a medium

wavelength the distance between corresponding points on a continuous wave (e.g. two crests or two troughs)

weight force applied to a mass by another mass

work energy transferred due to the action of a force acting through a distance

zero error scale is not zero when measurements are taken; also called a calibration error

INDEX

NELSON PHYSICS UNITS 1 & 2

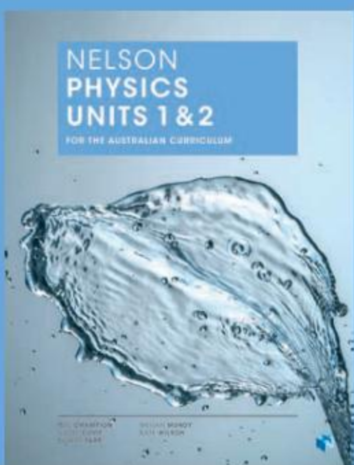
FOR THE AUSTRALIAN CURRICULUM

Nelson Physics for the Australian Curriculum Units 1 & 2 has been specifically designed to cover the three strands of the ACARA Australian Curriculum – Senior Physics.

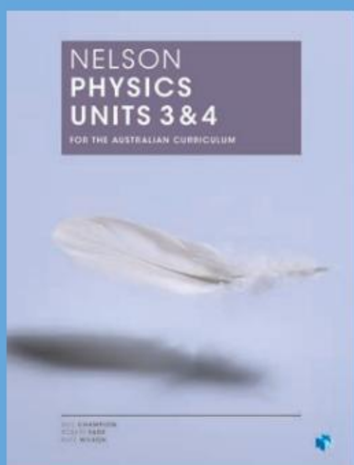
Written to the highest achievement standard, this engaging series brings Physics to life with clear language and relevant examples. Case studies and Scientific Literacy boxes help students connect the study of Physics to the real world.

nelsonnet

Register online at:
www.nelsonnet.com.au
to access digital resources to support improved learning & teaching*



NELSON PHYSICS UNITS 1 & 2
ISBN 9780170242103



NELSON PHYSICS UNITS 3 & 4
ISBN 9780170242110

NELSON CENGAGE Learning
For learning solutions, visit **cengage.com.au**

*Please note that complimentary access to NelsonNet and the NelsonNetBook is only available to teachers who use the accompanying student textbook as a core educational resource in their classroom. Contact your sales representative for information about access codes and conditions.

ISBN: 978-0170242103
9 780170 242103